

# Generative Image Sequence Modeling of Optical Imaging Data

Zur Erlangung des Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.)  
Genehmigte Dissertation von Kilian Leonard Heck aus Aschaffenburg  
Tag der Einreichung: 17.08.2023, Tag der Prüfung: 06.10.2023

1. Gutachten: Prof. Dr. Ralf Galuske  
2. Gutachten: Prof. Dr. Heinz Koeppel  
Darmstadt, Technische Universität Darmstadt



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Biology Department  
Systems Neurophysiology

Generative Image Sequence Modeling of Optical Imaging Data

Accepted doctoral thesis by Kilian Leonard Heck

Date of submission: 17.08.2023

Date of thesis defense: 06.10.2023

Darmstadt, Technische Universität Darmstadt

Bitte zitieren Sie dieses Dokument als:

URN: urn:nbn:de:tuda-tuprints-244138

URL: <http://tuprints.ulb.tu-darmstadt.de/24413>

Jahr der Veröffentlichung auf TUprints: 2023

Dieses Dokument wird bereitgestellt von tuprints,

E-Publishing-Service der TU Darmstadt

<http://tuprints.ulb.tu-darmstadt.de>

[tuprints@ulb.tu-darmstadt.de](mailto:tuprints@ulb.tu-darmstadt.de)

Die Veröffentlichung steht unter folgender Creative Commons Lizenz:

Namensnennung – Weitergabe unter gleichen Bedingungen 4.0 International

<https://creativecommons.org/licenses/by-sa/4.0/>

This work is licensed under a Creative Commons License:

Attribution–ShareAlike 4.0 International

<https://creativecommons.org/licenses/by-sa/4.0/>

---

## Abstract

---

This thesis focuses on the development of a data processing pipeline for inferring neural activity observed in cat's primary visual cortex. These activity patterns were measured in a grating stimulation paradigm using optical imaging based on fluorescent dyes, more specifically voltage-sensitive dye imaging. While offering a good compromise between spatial and temporal resolution, a low signal-to-noise ratio and dominant technical and biological noise components are inherent properties of the chosen data acquisition method. A high trial-to-trial variability of neural response activity poses additional challenges for data analysis. Further constraints on the chosen processing approach are presented in terms of computational efficiency as well as statistical robustness, which both are requirements for future closed-loop experimental designs. To tackle these aspects, the benefits of deep learning and probabilistic inference are taken advantage of by the utilization of a deep generative model framework, namely a variational autoencoder model architecture. Benchmarking and evaluating deep neural networks commonly requires training data with known ground truth information, which is not available for respective real data. For that purpose, an additional routine for generating synthetic image sequences resembling voltage-sensitive dye imaging recordings was developed. It incorporates knowledge about the data-generating process, including pre-defined spatio-temporal dynamics and typical signal- and artifact-related components. In six parameter studies on basis of both real and synthetic datasets, a wide range of model configurations was tested while considering different pre-processing steps. The thesis concludes with the implication that many of the tested model parametrizations offer a feasible trade-off between image reconstruction quality and model regularization, and can be adequately used for tracking signal- and noise-related features.

---

## Kurzfassung

---

In der vorliegenden Arbeit wird eine Datenverarbeitungs-Pipeline zur geeigneten Inferenz neuronaler Aktivitätsmuster vorgeschlagen. Diese Aktivität wurde über ein Stimulations-Paradigma mittels bewegter Balken-Muster im primären visuellen Kortex der Katze evoziert und unter Verwendung eines optischen Bildgebungsverfahrens auf Basis von Fluoreszenzfarbstoffen, dem Voltage-Sensitive Dye Imaging, aufgezeichnet. Zwar bietet diese Erhebungsmethode einen guten Kompromiss zwischen räumlicher und zeitlicher Auflösung, jedoch gehen damit auch ein niedriges Signal-Rausch-Verhältnis sowie dominante technische und biologische Störkomponenten einher. Eine hohe Trial-to-Trial-Variabilität der neuronalen Antwortaktivität stellt eine zusätzliche Herausforderung für anschließende Datenanalyse-Schritte dar. Weitere Einschränkungen für den gewählten Verarbeitungsansatz ergeben sich in Bezug auf Recheneffizienz und statistische Robustheit, welche wichtige Anforderungen für künftige Closed-Loop-Experimentaldesigns darstellen. Um mit diesen Aspekten umzugehen, werden die Vorteile des Deep Learnings sowie probabilistischer Inferenz durch Verwendung eines tiefen generativen Modells basierend auf der Modellarchitektur eines Variational Autoencoders genutzt. Für entsprechendes Benchmarking und Evaluation von Deep-Learning-Modellen sind üblicherweise Trainingsdaten mit bekannter Ground Truth erforderlich, welche für die gewählten Realdaten nicht verfügbar sind. Zu diesem Zweck wurde eine zusätzliche Routine zur Generierung synthetischer Bildsequenzen entwickelt, die Aufnahmen des Voltage-Sensitive Dye Imaging ähneln und Vorwissen über den Datengenerierungs-Prozess beinhalten. Dabei wurde die Zusammensetzung über vordefinierte Dynamiken und typische signal- und artefaktbezogene Komponenten auf raum-zeitlicher Ebene vorgenommen. In sechs Parameterstudien auf Basis realer und synthetischer Datensätze wurde unter Berücksichtigung verschiedener Vorprozessierungsschritte ein breites Spektrum an Modellkonfigurationen getestet. Die Arbeit schließt mit der Schlussfolgerung, dass viele der getesteten Modellparametrisierungen einen sinnvollen Kompromiss zwischen Bildrekonstruktions-Qualität und Modell-Regularisierung erzielen können und sich für das Tracking von signal- und rauschbezogenen Merkmalen eignen.



---

# Contents

<b>List of Figures</b>	<b>viii</b>
<b>List of Tables</b>	<b>ix</b>
<b>Acronyms</b>	<b>x</b>
<b>Notation</b>	<b>xii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Visual Processing . . . . .	2
1.1.1 Phototransduction . . . . .	2
1.1.2 The Retina . . . . .	4
1.1.3 Receptive Fields . . . . .	7
1.1.4 Organizational Principles of Visual Pathways . . . . .	9
1.1.5 Primary Visual Cortex . . . . .	12
1.2 Advanced Concepts of Cortical Information Processing . . . . .	15
1.2.1 Brain States . . . . .	15
1.2.2 Spontaneous / Ongoing Activity . . . . .	16
1.3 Motivation & Research Goals . . . . .	17
<b>2 Materials &amp; Methods</b>	<b>21</b>
2.1 Data Acquisition . . . . .	21
2.1.1 Voltage-Sensitive Dye Imaging . . . . .	21
2.1.2 Experimental Setup . . . . .	24
2.2 Data Pre-Processing . . . . .	27
2.2.1 Hardware-Based Approaches . . . . .	27
2.2.2 Software-Based Approaches . . . . .	28
2.3 Data Modeling . . . . .	32

---

2.3.1	Bayesian Inference . . . . .	32
2.3.2	Generative Modeling . . . . .	33
2.3.3	Latent Variable Modeling . . . . .	35
2.3.4	Approximation of Intractable Distributions . . . . .	37
2.3.5	Variational Autoencoder . . . . .	40
2.4	Synthetic Data Generation . . . . .	44
2.4.1	Motivation . . . . .	44
2.4.2	Synthetic Orientation Preference Maps . . . . .	45
2.4.3	Spatial Random Fields . . . . .	48
2.4.4	Conditioned Random Fields . . . . .	49
2.4.5	Temporal Noise Components . . . . .	52
2.4.6	Spatial Noise Components . . . . .	54
2.4.7	Random Noise . . . . .	57
2.4.8	Composition of Artificial VSDI Sequences . . . . .	57
<b>3</b>	<b>Results</b>	<b>59</b>
3.1	Approaches for Model Evaluation . . . . .	59
3.2	Parameter Studies . . . . .	61
3.2.1	Overview . . . . .	61
3.2.2	Setting A: CRF . . . . .	64
3.2.3	Setting B.1: Synthetic VSDI, raw . . . . .	72
3.2.4	Setting B.2: Synthetic VSDI incl. Pre-Processing (Blank Subtraction) . . . . .	75
3.2.5	Setting C.1: VSDI, raw . . . . .	78
3.2.6	Setting C.2: VSDI incl. Pre-Processing (Blank Subtraction) . . . . .	85
3.2.7	Setting C.3: VSDI incl. Pre-Processing (GLM) . . . . .	88
3.3	Computational Performance . . . . .	91
<b>4</b>	<b>Discussion</b>	<b>96</b>
4.1	Implications of Parameter Studies . . . . .	96
4.2	Limitations & Extensibility . . . . .	98
4.2.1	Stimulation Paradigm . . . . .	98
4.2.2	Data Acquisition . . . . .	99
4.2.3	Data Simulation . . . . .	102
4.2.4	Data Pre-Processing . . . . .	104
4.2.5	Data Analysis / Modeling . . . . .	105
4.2.6	Model Evaluation . . . . .	108
4.3	Related Work . . . . .	109
4.4	Conclusion & Outlook . . . . .	110

---



---

<b>Appendix A - Code Repository Overview</b>	<b>113</b>
<b>Bibliography</b>	<b>115</b>
<b>Curriculum Vitae</b>	<b>136</b>
<b>Acknowledgement</b>	<b>139</b>

---

---

## List of Figures

1	Structure of mammalian retina. . . . .	5
2	Distributed and parallel processing in the visual system . . . . .	11
3	Functional organization of a hypercolumn in primary visual cortex (V1) . .	13
4	Exemplary illustration of a hypothetical closed-loop experimental paradigm	18
5	Schematic overview of the applied optical imaging area. . . . .	26
6	Linear model decomposition of the VSDI signal by Reynaud, Takerkart, Masson and Chavane, 2011 . . . . .	32
7	Taxonomy of generative models based on maximum likelihood principle . .	35
8	General model architecture of the VAE . . . . .	41
9	Synthetic orientation preference map. . . . .	47
10	Realization of a 3-D spatial random field following a Matérn covariance model. . . . .	49
11	CRF key locations. . . . .	52
12	Temporal signal confounders of VSDI . . . . .	53
13	Artificial illumination signal component . . . . .	55
14	Intuition of space-colonization algorithm. . . . .	56
15	Synthetic blood vessel component . . . . .	57
16	Composition of of artificial VSDI greyscale image sequences. . . . .	58
17	Parameter study, settings A, B.1, B.2: CRF patterns. . . . .	63
18	Parameter study, settings A, B.1, B.2: Data basis. . . . .	65
19	Parameter study, setting A: MSE . . . . .	67
20	Parameter study, settings A, B.1, B.2: Input reconstructions. . . . .	68
21	Parameter study, settings A, B.1, B.2: Latent space walk. . . . .	70
22	Parameter study, settings A, B.1, B.2: Frame-wise encodings . . . . .	71
23	Parameter study, setting B.1: MSE . . . . .	74
24	Parameter study, setting B.2: MSE . . . . .	77
25	Parameter study, settings C.1, C.2, C.3: Data basis. . . . .	79

---

26	Parameter study, setting C.1: MSE . . . . .	80
27	Parameter study, settings C.1, C.2, C.3: Input reconstructions. . . . .	82
28	Parameter study, settings C.1, C.2, C.3: Latent space walk. . . . .	83
29	Parameter study, settings C.1, C.2, C.3: Frame-wise encodings . . . . .	84
30	Parameter study, setting C.2: MSE . . . . .	87
31	Parameter study, setting C.3: MSE . . . . .	90
32	Implemented VAE layer architecture. . . . .	93
33	Parameter study, settings C.1, C.2, C.3: ROI. . . . .	95
34	Orientation preference maps obtained via VSDI & ISI. . . . .	101

## List of Tables

1	Overview of parameter study settings. . . . .	62
2	Parameter study, setting A: Assessment of model fit. . . . .	66
3	Parameter study, setting B.1: Assessment of model fit. . . . .	72
4	Parameter study, setting B.2: Assessment of model fit. . . . .	76
5	Parameter study, setting C.2: Assessment of model fit. . . . .	85
6	Parameter study, setting C.3: Assessment of model fit. . . . .	88

---

## Acronyms

---

<b>2-DG</b>	2-Deoxyglucose Mapping
<b>cGMP</b>	Cyclic Guanosine 3'-5' Monophosphate
<b>CMOS</b>	Complementary Metal-Oxide Semiconductor
<b>CPU</b>	Central Processing Unit
<b>CRF</b>	Conditioned Random Field
<b>EEG</b>	Electroencephalogram
<b>ECG</b>	Electrocardiogram
<b>ELBO</b>	Evidence Lower Bound
<b>FLOBS</b>	fMRI's Linear Optimal Bias Set
<b>FFT</b>	Fast Fourier Transform
<b>fMRI</b>	Functional Magnetic Resonance Imaging
<b>GC</b>	Guanylate Cyclase
<b>GLM</b>	General Linear Model
<b>GPU</b>	Graphics Processing Unit
<b>GTP</b>	Guanosine Triphosphate
<b>ICA</b>	Independent Component Analysis
<b>iCDF</b>	Inverse Cumulative Distribution Function
<b>KDE</b>	Kernel Density Estimation

---

<b>KL</b>	Kullback-Leibler
<b>LGN</b>	Lateral Geniculate Nucleus
<b>MCMC</b>	Markov Chain Monte Carlo
<b>MEG</b>	Magnetoencephalography
<b>MNIST</b>	Modified National Institute of Standards and Technology
<b>MSE</b>	Mean Squared Error
<b>MT</b>	Middle Temporal Area
<b>NIRS</b>	Near Infrared Spectroscopy
<b>OIIS</b>	Optical Imaging of Intrinsic Signals
<b>PCA</b>	Principal Component Analysis
<b>PDE</b>	Phosphodiesterase
<b>PET</b>	Positron Emmission Tomography
<b>PMLS</b>	Posteromedial Lateral Suprasylvian Area
<b>RAM</b>	Random-Access Memory
<b>ROI</b>	Region of Interest
<b>SNR</b>	Signal-to-Noise Ratio
<b>SRF</b>	Spatial Random Field
<b>SSD</b>	Solid State Drive
<b>TTL</b>	Transistor-Transistor Logic
<b>V1</b>	Primary Visual Cortex
<b>VAE</b>	Variational Autoencoder
<b>VSDI</b>	Voltage-Sensitive Dye Imaging

---

## Notation

---

Symbol	Description
$a$	A vector $a$ .
$A$	A matrix, or physical quantity $A$ .
$A_+$	The Moore-Penrose pseudo inverse of matrix $A$ .
$\alpha$	A parameter $\alpha$ .
$\hat{\alpha}$	An estimator of parameter $\alpha$ .
$\mathcal{D}$	A family of approximate densities.
$\mathbb{E}[\cdot]$	The expectation operator.
$f(x)$	A function $f$ of variable $x$ .
$\mathcal{F}(q)$	The evidence lower bound (ELBO) $\mathcal{F}$ dependent on the distribution $q$ .
$\gamma(\cdot)$	The semi-variogram $\gamma$ .
$\Gamma(\cdot)$	The Gamma function.
$I$	The identity matrix $I$ .
$\text{KL}(q(z)  p(x))$	The Kullback–Leibler divergence between distributions $q(z)$ and $p(x)$ .
$\mathcal{L}$	The log-likelihood.
$\text{mod}(a, b)$	The modulo operation of two positive numbers $a$ and $b$ .
$\mathcal{N}(\mu, \sigma)$	Probability density function of the Normal (or Gaussian) distribution with mean $\mu$ and variance $\sigma^2$ .
$\mathcal{N}(0, 1)$	Probability density function of the standard Normal (or Gaussian) distribution.
$p(x)$	A probability density function $p$ of variable $x$ .
$p(x z)$	A conditional probability density function $p$ of variable $x$ given variable $z$ .
$p(x, z)$	A joint probability density function $p$ of variables $x$ and $z$ .
$\rho(\cdot)$	The correlation function $\rho(\cdot)$ .

---



---

# 1 Introduction

---

Over the course of evolution, the visual system has developed to meet the ecological needs of mammals. Here, learning rules about structures of our environment have been established. From these rules, some have shaped neural circuitries directly, and others in turn are guiding the brain to interpret the current visual scenery from past experience. In these scenes, visual images offer a rich source of information about our environment, from which the brain is able to rapidly infer distinct object features (e.g. identity, size, shape, distance, velocity), while not being dependent on light conditions or occlusion. Segmenting a scene into foreground and background elements is part of this object recognition and is based on both geometric as well as cognitive principles. Latter comprise states of attention and expectation, which in turn can facilitate perceptual tasks in form of additional information like priming stimulations or internal representations. Besides mere object recognition, the brain also guides body movement such as hand movement on basis of visual information (Kandel, Koester, Mack and Siegelbaum, 2021).

The following introduction covers principles of information processing in mammalian visual system, starting from phototransduction of light as physical signal into a neural signal (Sect. 1.1.1), latter being converted and pre-processed by different retinal cell types (Sect. 1.1.2). Through receptive fields (Sect. 1.1.3), informations are further distributed along parallel but interacting pathways (Sect. 1.1.4) formed by reciprocal connections between several brain areas. Here, the primary visual cortex (V1) (Sect. 1.1.5) processes low-level features such as object orientation and direction, which are vital information for survival. All visual processing is conditioned by advanced concepts of cortical information processing, such as brain states (Sect. 1.2.1) as well as spontaneous (or ongoing) activity (Sect. 1.2.2). From these general information, the motivation and research goals of this thesis are derived (Sect. 1.3).

---

## 1.1 Visual Processing

### 1.1.1 Phototransduction

Fundamentally, light is electromagnetic radiation with a defined bandwidth of wave lengths; here, visible light comes in a narrow value range of 400 – 700 nm. The energy of a light photon can be calculated via the Planck relation (or Planck-Einstein relation)  $E = h * c / \lambda$ , where  $h$  is Planck's constant of  $6,626 * 10^{-34}$  Js,  $c$  is the speed of light with  $c = 3 * 10^8$  ms and  $\lambda$  denotes the wavelength of light in nm. This has the following important implication for the relation between wave length and energy of light: the higher the wave length, the lower the energy of light will become (Einstein, 1905; Planck, 1901). Light makes a peculiar physical signal, as its intensity as well as its corresponding wave length is highly variable for a) different media (e.g. air or water) and b) the time of day, which is both influencing the spectral composition of light. Shorter wave lengths are more reflectable from objects in directional manner. Thus, given that the right detectors are accordingly available in the subject, it is possible to instantaneously depict an object with high spatial resolution. Therewith, mammals can visualize their environment by properties of light (Warrant and Johnsen, 2013).

The necessary detectors for decoding reflected light photons are photoreceptors. Those translate the light signal into an electrical signal (phototransduction), which in turn can be further processed on neural basis (Sect. 1.1.2). Photoreceptors exist in different resolving systems such as the lenticular eye (as in vertebrates and evertebrates) and the compound eye (as in insects). Depend on species, there exist diffuse light sensing organs, which in turn can lead to behavioural adaptations like phototaxis and photophobia (Jékely, 2009). Light activation in the retina's photoreceptors results in a graded change in membrane potential and respective alterations in the rate of transmitter release onto postsynaptic neurons. Graded potentials play a significant role in the retina's processing, mostly because action potentials are not necessary at the relatively small distances at issue. This contrasts with the majority of other sensory systems, where the activation of receptors leads to a depolarization of the cell membrane, which in turn stimulates an action potential and leads to the release of transmitters onto connected neurons (Purves et al., 2001).

The balance of membrane conductances to  $\text{Na}^+$  and  $\text{K}^+$  ions controls the membrane potential of a photoreceptor. In darkness, the receptor is in a depolarized state which is accompanied by a membrane potential of approx.  $-40$  mV. This comes from levels of the cyclic guanosine 3'~5' monophosphate (cGMP) which control the influx of  $\text{Na}^+$  into the photoreceptor by opening nonselective cation channels. From guanosine triphosphate

---

(GTP), a guanylate cyclase (GC) is generating cGMP, latter being hydrolyzed by a phosphodiesterase (PDE). In the absence of light, the activity of PDE is low, cGMP levels are high and cGMP-gated ion channels are open. Rhodopsin constitutes the light-sensitive visual pigment in rod photoreceptor cells and consists of two coupled portions: (i) the protein component opsin, and (ii) the light-absorbing chromophore retinal. As each cone photoreceptor type in human retina produces a variant of the opsin protein, three different cone pigments can be differentiated by their absorption spectrum, which denotes wavelength of the efficiency of light absorption. The comparison of signals from these cone types enables the brain for color vision (Alberts, Wilson and Hunt, 2008; Kandel et al., 2021; O'Brien, 1982).

When a photon is absorbed by rhodopsin molecules in the outer-segment discs, a biochemical cascade is initiated with the isomerization of the 11-cis chromophore attached to the rhodopsin molecule to all-trans retinal. This switch leads to a conformational change in the protein (opsin) to an activated state called metarhodopsin II, which in turn causes the activation of an intracellular messenger transducin. This G protein is then activating a PDE that hydrolyzes cGMP at the disk membrane. This lowers the concentration of cGMP bound to the plasma membrane cation channels, thus closing these cGMP-gated channels and bringing the cell towards the  $K^+$  equilibrium potential. Due to the closure of ion channels, the membrane is hyperpolarized up to a level of  $-65$  mV. Also, the rate of neurotransmitter release (glutamate) at the synapse is reduced as it depends on voltage-sensitive  $Ca^{2+}$  channels. Thereby, a neural signal is initiated (Alberts et al., 2008; Kandel et al., 2021; O'Brien, 1982; Purves et al., 2001).

To become responsive for another photon again, the photoreceptor has to return to its dark state. Therefore, the duration of the amplifying G protein cascade has to be limited. In rods, this is accomplished via independent mechanisms which shut off single elements of the cascade. For example, metarhodopsin II is rendered inactive through phosphorylation by a particular rhodopsin kinase and binding of the soluble protein arrestin, which in turn prevents the interaction with transducin. By negative feedback mechanisms which are mediated by concentration changes of  $Ca^{2+}$  in the cell, large responses are terminated more quickly. As the cell is responding to light, the cGMP-gated channels close and the  $Ca^{2+}$  level quickly decreases. The lower level of  $Ca^{2+}$  is influencing at least three components of the cascade, namely rhodopsin, GC and cGMP-gated channels. Therefore, a decrease in  $Ca^{2+}$  antagonizes the excitation caused by light (Alberts et al., 2008; Kandel et al., 2021).

The biochemical cascade has two important implications. On the one hand, G protein-coupled receptors activate a variety of intracellular pathways that rely on relay chains of intracellular proteins and mediators, which are amplifying the response signal to ex-

---

tracellular signals. On the other hand, the signal amplification depends on the level of illumination, which is referred to as light adaptation. Photoreceptors are most sensitive at low levels of illumination, though sensitivity diminishes as illumination increases. This prevents saturation of receptors and therefore broadening the range of light intensities on which they operate. In this regard,  $\text{Ca}^{2+}$  appears to be important in the outer segment for light-induced regulation of photoreceptor sensitivity due to its regulatory effects (Alberts et al., 2008; Kandel et al., 2021; Purves et al., 2001).

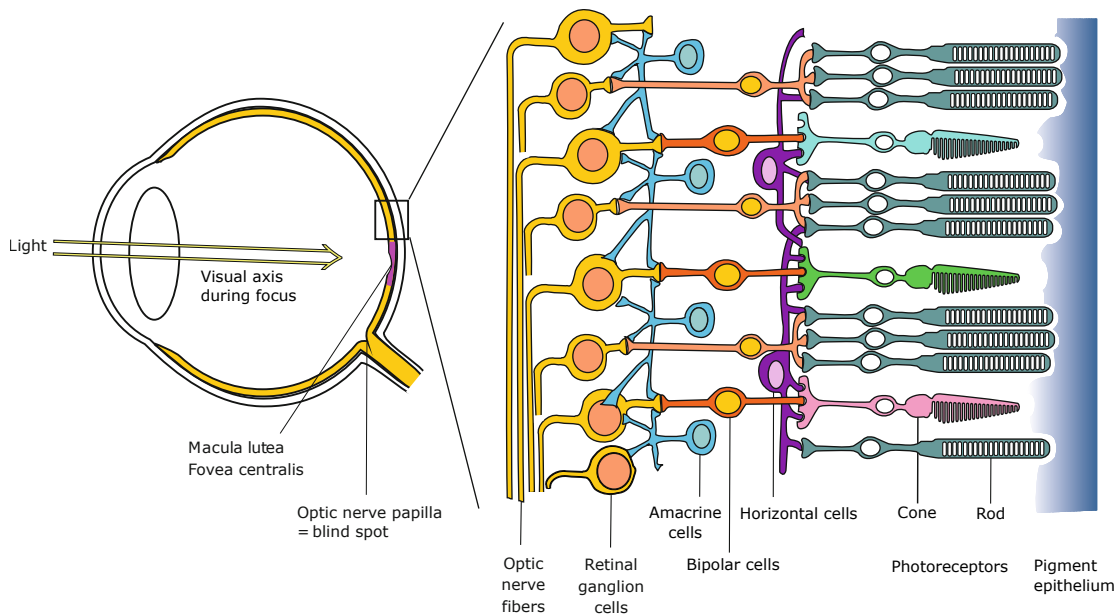
### 1.1.2 The Retina

Due to the inside-out design of the mammalian retina, light has to pass its first couple of neuron layers until it reaches the photoreceptors. Latter are converting light into a neural signal, as previously described in Sect. 1.1.1. In humans, two types of photoreceptor cells are differentiated: rods and cones. Due to comparably longer outer segments holding the light-absorbing photopigment rhodopsin, rods are more sensitive to light intensity. They are capable to detect the absorption of a single photon, therefore enabling vision under weaker light conditions. While there exist only one type of rods in humans, three classes of cones can be differentiated due to their respective type of photopigment, each having specific absorption maxima. Therefore, each cone type is selectively responding for certain wavelengths of light. As the brain is comparing signals between these three cone types, this in turn leads to selectivity for color informations. Furthermore, cones respond faster than rods (Kandel et al., 2021).

The distribution of rods and cones across the retina is varying considerably. On the one hand, the number of rods increase towards the peripheral retina, enabling scotopic vision as visual acuity decreases but higher light sensitivity under low-light levels. On the other hand, the amount of cones increases towards the retina's center, enabling photopic vision with higher visual acuity in case of higher levels of light, e.g. during daylight conditions. The overall density of photoreceptors decreases towards the peripheral retina (Steinberg, Reid and Lacy, 1973).

Informations of photoreceptors are subsequently collected and integrated by retinal circuits illustrated in Fig. 1 which are established from different cell types:

- vertical connections via bipolar cells up to magno-, parvo-, and koniocellular retinal ganglion cells;
- lateral connections by horizontal and amacrine cells, which allow for converging or diverging signal transmission through inhibitory or excitatory connections.



**Figure 1:** Structure of mammalian retina. Light has to pass several cellular layers and their respective processes before reaching the photoreceptors (rods and cones). Information is then passed (i) vertically from photoreceptors to bipolar cells to retinal ganglion cells, and (ii) horizontally through horizontal and amacrine cells. Fibers of retinal ganglion cells form the optic nerves, the output of the retina. Modified from W. A. Müller, Frings and Möhrlen (2019).

Photoreceptors connect to bipolar and horizontal cells via ribbon synapses and are using glutamate as neurotransmitter. Latter is released in dependence on illumination. Under low light conditions, glutamate is released continuously, while illumination hyperpolarizes the receptor. In turn, less calcium enters the synaptic terminal and less glutamate is released (Kandel et al., 2021).

Horizontal cells connect laterally to several photoreceptors and modulate the synaptic gain between photoreceptors and bipolar cells. While photoreceptors are sending glutamatergic output to horizontal cells, rods and cones are receiving inhibitory feedback (Mangel, 1991; Thoreson, Babai and Bartoletti, 2008; Werblin, 1974). These feedback signals are spreading to presynaptic terminals of adjacent photoreceptors. This effect, also known as lateral inhibition, enhances differences between stimulated photoreceptors and

---

their unstimulated neighbourhood. Along with direct feedforward input from horizontal to bipolar cells, lateral inhibition leads to the receptive field organization of bipolar cells in terms of a center-surround antagonism in response to light stimulation, as well as color-opponent interactions (Baylor, Fuortes and O'Bryan, 1971; Burkhardt, 1993).

Bipolar cells are responding to glutamate with graded postsynaptic potentials. Depending on corresponding glutamate receptors, this response comes in two different ways: on the one hand, OFF bipolar cells are characterized by excitatory ionotropic glutamate receptors and depolarize by glutamate released in the dark. On the other hand, inhibitory metabotropic glutamate receptors of ON bipolar cells lead to hyperpolarization of the cell when activated by glutamate in darkness. Both bipolar cell types can be further differentiated by their shape, axon terminations and morphology of dendrites. Also, their post-synaptic targets vary, as they connect with specific amacrine and retinal ganglion cells (Kandel et al., 2021).

Aside of horizontal cells, amacrine cells form another class of interneurons in the retina. As they vary in size, numbers and stratification, around 30 types of amacrine cells are distinguished. This diversity also reflects their diverse functions (Masland, 2012). Amacrine cells interconnect between bipolar cells and retinal ganglion cells. While some subtypes sending direct feedback to bipolar cells, other types form an inhibitory feedback network via electrical coupling. Through this network, a bipolar cell can be inhibited by a distant bipolar cell in similar fashion as discussed for horizontal cells (Kandel et al., 2021).

Retinal ganglion cells are differentiated physiologically and morphologically. In vertebrates, M-cells (lat. magnus: large) and P-cells (lat. parvus: small) are building the two major classes of cells. On the one hand, M-cells have larger cell bodies and receptive fields, also they respond quicker to a given stimulus. P-cells on the other hand have smaller bodies and receptive fields, and are responding slowly to stimulation. Another differentiation can be made in terms of color sensitivity, which comes from their respective way of integrating information of different cone types: M-cells sum up the inputs of different cone type, therefore effectively being color-blind. P-cells are sensitive for wavelengths due to the subtractive integration of different cone inputs (Livingstone and Hubel, 1988).

Specifically for cat's retina, a differentiation into three types of retinal ganglion cells is made in terms of morphology as well as response behaviour. Type Y ganglion cells feature phasic responses, broad receptive fields, and hence low spatial resolution. While not being selective for color information, they are sensitively responding to movement. X-type ganglion cells on the other hand exhibit exceptional spatial resolution due to small receptive fields. They respond tonically to slow or static stimuli, therefore assessing colors, patterns,

---

and finer details. The third category consists of W-type cells, which respond to bright and dark light spots both phasically and tonically. Their conduction velocity is substantially lower compared to other ganglion cell types (Boycott and Wässle, 1974; Cleland, Levick and Wässle, 1975).

Retinal ganglion cells form the output of the retina. Their axons converge at the optical disk and continue through the retina to exit at the rear of the eye. From there, they are building the optical nerves of both eyes, which are merging and crossing in the optic chiasm. Ganglion cell axons corresponding to (i) the temporal half of the ipsilateral retina, as well as (ii) the nasal half of the contralateral retina form the optic tract for each eye. Therefore, the right hemisphere receives input from the left visual hemifield, and vice versa. The optic tract then terminates in four nuclei, each covering different functions:

- the lateral geniculate nucleus (LGN) of the thalamus (visual perception);
- the pretectum of the midbrain (pupillary light reflex);
- the superior colliculus of the midbrain (eye movement);
- the suprachiasmatic nucleus of the hypothalamus (circadian rhythm)

(Kandel et al., 2021).

### 1.1.3 Receptive Fields

Receptive fields exist on successive relays along the visual hierarchy, at which their size and complexity increases due to the considerable convergence and divergence of synaptic connections along the visual pathways. The size of a receptive field on the retina depends on the field's eccentricity - its relative position to the fovea. Receptive fields towards the fovea indicating smaller sizes due to inputs from only a few photoreceptors, and larger receptive fields are recognized towards the periphery as input from many receptors are received (Cleland, Harding and Tulunay-Keeseey, 1979; Goodchild, Ghosh and Martin, 1996). At successive relay stations, their optimal stimulus is getting more complex as well. This underlines the integrative function of receptive fields: to form a unified percept from multiple components from large areas of the visual field (Kandel et al., 2021).

A bipolar cell's receptive field is characterized physically by the position and distribution of receptor cells with which it establishes synaptic contact. It is antagonistically structured due to opposing influences of (i) hyperpolarized photoreceptor cells in case of illumination, and (ii) depolarized photoreceptor cells in case of illumination of its surroundings due to

---

synaptical contacts of horizontal cells (Kandel et al., 2021).

Lateral inhibition in the inner retina, which is mediated by sustained activity of GABAergic amacrine cells, is substantially contributing to a concentric center-surround property of retinal ganglion cells' receptive field (Cook and McReynolds, 1998). This means that one portion of the receptive field is excitatory and the other inhibitory. This leads to the categorization of ON-center and OFF-center receptive fields, having distinct discharge patterns depending on the configuration of illumination. For example, when light is stimulating the central region of the ON-center type, this will lead to an increase in the cell's firing rate. This type also has an inhibitory surrounding area, decreasing the frequency of action potentials when stimulated. When light falls on both center and surrounding portions, this will lead to little to no response. Another discharge pattern concerns light-dark boundaries across the receptive field, in turn leading to brisk responses (Kuffler, 1953). This indicates the preference of retinal ganglion cells for borders, leading to enhancements of spatial contrast information such as an edge between two areas with inhomogeneous illumination (Enroth-Cugell and Robson, 1966). Also, they selectively process temporal changes in light intensities while rejecting features which are constant in space or time. Retinal ganglion cells therefore serve for deducing shapes and identities as well as sudden movements or changes of objects (Kandel et al., 2021). Similar receptive field properties in retinal ganglion cells such as the aforementioned center-surround antagonism were also found for LGN neurons, which in turn are topologically mapping to locations on the retina (Hubel and Wiesel, 1961).

Along the visual pathways, properties of the receptive fields are changing between corresponding brain areas. While receptive fields in the cortex are sensitive for contrast such as retinal ganglion cells and LGN neurons, they furthermore are analyzing contours. This was first observed in 1958 by Hubel & Wiesel in V1 of anesthetized cats. Here, a cell discharged in response to a moving line shadow formed by the edge of a slide when placed into a ophthalmoscope. The cell would only fire when the line was angled within a specific range. Additional measurements indicated that many more cells were likewise sensitive to boundary orientation. Thus, this concluded into the finding of orientation tuning in the early visual pathway (Hubel and Wiesel, 1998).



---

## 1.1.4 Organizational Principles of Visual Pathways

Within the cerebral cortex, several visual areas are differentiated on basis of multiple criteria. From traditional anatomical perspective, the cell size, shape and packing in the cortical layers as well as myelin thickness and density were used in classification of Brodmann areas (Brodmann, 1903, 1909).

Aside from the neuron's anatomical properties, their function is also used for area demarcation. At early levels of visual pathways, corresponding receptive fields are still small which allows the derivation of precise visuotopic maps. These maps in turn can offer insights about the functional distinction of areas. As the size of receptive fields increases along the hierarchy of visual areas, the precision of visuotopic maps decreases, therefore they become a more unreliable indicator of area boundaries (Kandel et al., 2021).

The brain forms a unified percept by an important organizational principles of vision: distributed and parallel processing. Multiple brain areas are part of parallel but interacting neural pathways, which selectively analyze features of a visual scene on different hierarchical levels:

- lower level: orientation, contrast, disparity, color, movement direction;
- intermediate level: contour integration, surface properties, shape discrimination, surface depth & segmentation, distinction between foreground and background, object motion;
- higher level: object identification.

The visual system is extensively using parallel processing, which is already established in the retina. Low-level attributes are extracted by distinct retinal circuits from photoreceptors, latter covering small sections of the visual field. This is also accompanied by adaptations of the retina's sensitivity to continuing changes in illumination (Kandel et al., 2021).

The anatomical segregation by eye and object feature continues along the geniculostriate pathway shown in Fig. 2, which spans from retinal channels past the LGN along the optic radiation up to V1. The LGN poses a thalamic relay station. It is primarily structured in six layers, which receive input from different ganglion cell types:

- midget ganglion cells (red-green information) connect with four parvocellular layers; these make up about 80 % of all retinal ganglion cell projections to LGN;

- 
- parasol ganglion cells (achromatic contrast) project to two magnocellular layers; they comprise approx. 10 % of all retinal ganglion cell projections to LGN;
  - bistratified ganglion cells (blue-yellow information) connect to koniocellular layers, latter pairing each magno- and parvocellular layer (Callaway, 2005; Leventhal, Rodieck and Dreher, 1981; Rodieck and Watanabe, 1993).

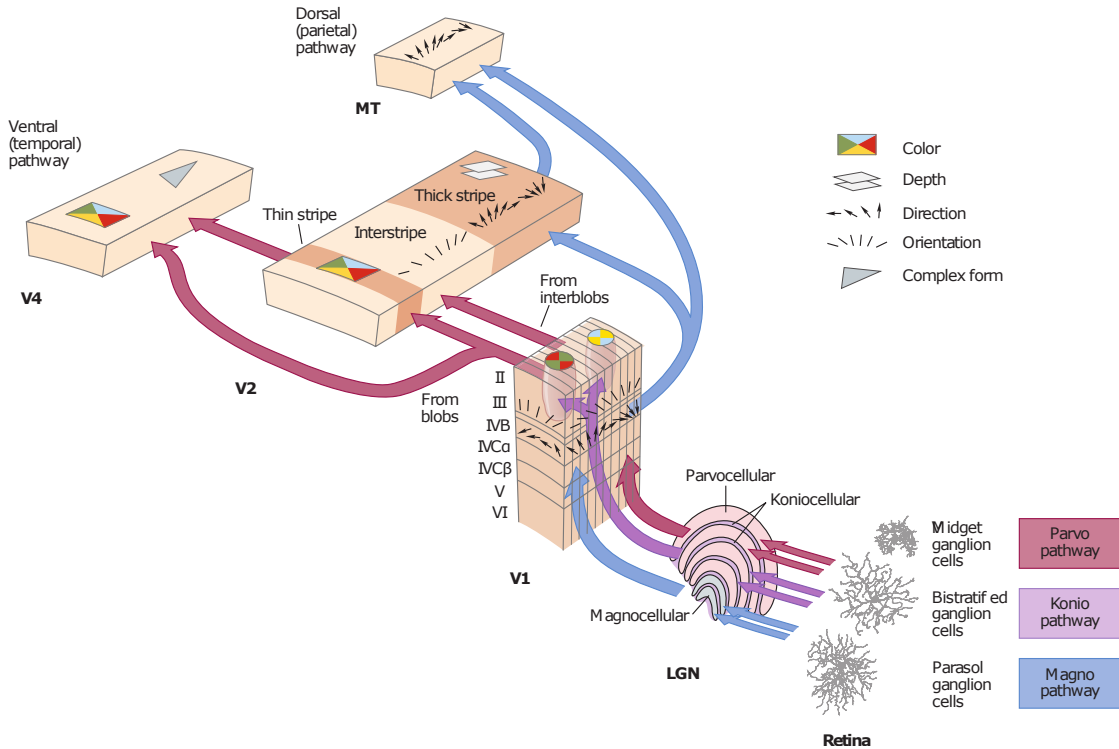
V1 represents the first cortical area of visual processing. It is divided into six different layers, which receive input from the different retinal channels, mainly entering layers IV and VI. Here, a recombination of visual input into new feature sets takes place, such as tuning for orientation and motion, as well as object depth. As V1 in each hemisphere is covering more than half of the visual field, an overlap of their representations is formed at the vertical meridian and unified by the corpus callosum connecting both hemispheres (Kandel et al., 2021).

From V1, parallel processing of visual information continues on two distinct pathways, which are illustrated in Fig. 2:

- the ventral pathway into the temporal cortex, which is selective for form and objects;
- the dorsal pathway through parietal cortex, which is selective for motion tracking, attention as well as visuomotor integration.

Though running in parallel, both pathways are interconnected and can share information. Object recognition in the ventral pathway can be facilitated by kinematic cues which stem from areas of the dorsal pathway. After an object has been identified, it is further linked with already formed memories of shapes and associations of their meaning (Kandel et al., 2021).

All connections between visual areas are reciprocal, which means that information is shared not only via feedforward connections from lower to higher (bottom-up), but also from higher to lower (top-down) areas through feedback connections (Felleman and Van Essen, 1991; Rockland and Pandya, 1979). This is additionally complicating the aforementioned functional distinction of areas. Feedback connections exist either directly between areas or indirectly via the thalamus, in particular the pulvinar. In context of the visual system, an important aspect of the LGN is concerning its modulatory effects on retinal information. These modulations are caused by mechanisms related to attention, expectation and the perceptual task, which are expressed in the form of inhibition of the LGN as well as feedback connections from visual cortex (Gilbert and Li, 2013). Feedback connections are critical for distinguishing a figure from the background. Deactivation



**Figure 2:** Distributed and parallel processing in the visual system. Depending on the type of retinal ganglion cells, three parallel pathways are projecting into parvo-, magno- and koniocellular layers of LGN. From there, these pathways terminate in different layers of V1: parvocellular fibers in layer  $IVC\beta$  and magnocellular fibers in layer  $IVC\alpha$ . From V1, a separation of functions is carried out by the dorsal and ventral pathways. Information about form and color are further processed on the ventral pathway through V2 and V4, while the dorsal pathway passing V2 and MT is specialized for movement directions. Note that pathways are running in parallel while allowing for interconnections in several areas. Modified from Kandel, Koester, Mack and Siegelbaum (2021).

---

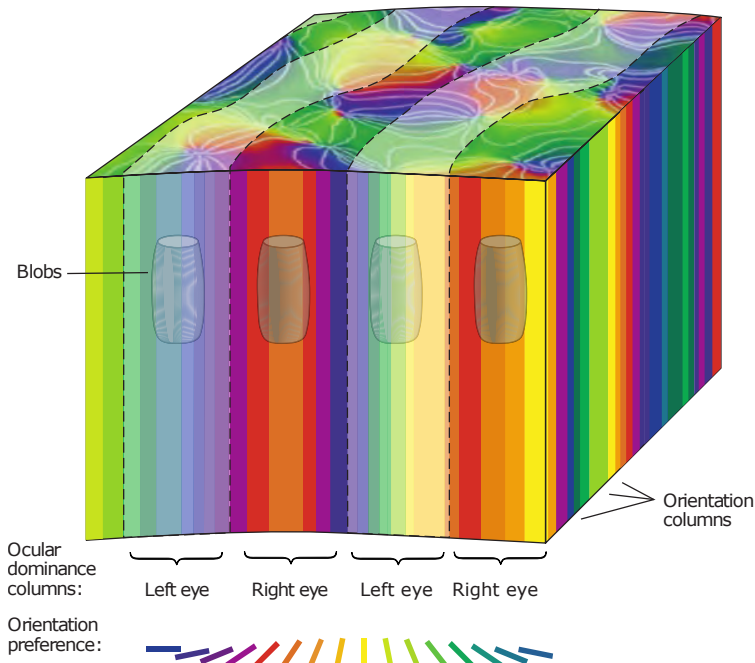
experiments by cooling middle temporal area (MT/V5) in primates showed that neurons in area V3 exhibited decreased center response and center-surround interactions, especially when dealing with low intensity stimuli. Here, feedback signal amplification for stimuli with low visibility can be expressed in terms of gain enhancements (Hupé et al., 1998).

In cat's visual cortex, the posteromedial lateral suprasylvian sulcus (PMLS) represents an area of higher order within the dorsal pathway. It corresponds to primates' MT/V5 and holds direction-selective neurons for processing movements within the extrastriate cortex (Dreher, Wang and Burke, 1996; Lomber, Payne, Cornwell and Long, 1996). By extensive feedback connections from visuoparietal cortex back to V1, it modulates activity in V1, more specifically in area 18 (Payne and Lomber, 2003; Symonds and Rosenquist, 1984). Experiments using reversible cooling revealed that deactivating PMLS leads to disruptions in direction selectivity of neurons in cat's area 18 while leaving orientation selectivity largely intact (Galuske, Schmidt, Goebel, Lomber and Payne, 2002).

### **1.1.5 Primary Visual Cortex**

In the case of cat's visual system, V1 is defined by two adjacent visual areas 17 (striate cortex) and 18 (parastriate cortex), as they are reciprocally connected through equally important association systems (Tretter, Cynader and Singer, 1975). Fibers of motion-sensitive retinal ganglion cells (Y type) are mainly projecting via parvocellular layers of LGN into area 18, further connecting to PMLS as gateway for the dorsal visual pathway. Retinal ganglion cells selectively encoding color and shape (X type) are innervating magnocellular layers of LGN and continuing via area 17 to area 21a, latter representing as entrance for the form-processing ventral pathway (Burke, Dreher and Wang, 1998; Dreher, Wang, Turlejski, Djavadian and Burke, 1996; Hickey and Gullery, 1974). As the topography of neighboring regions in the visual field is preserved by corresponding adjacent field representations along visual pathways from the retina to the visual cortex, the visual field is mapped retinotopically on the surface of area 17 (Tusa, Palmer and Rosenquist, 1978) and area 18 (Tusa, Rosenquist and Palmer, 1979).

Aside of this visuotopic organization, neural processing in V1 is underlined by its functional organization illustrated in Fig. 3. Neurons that share similar fundamental properties, namely orientation and color preference as well as ocular-dominance, are located close together. They form columnary units which stretch vertically from the pia to the white matter and whose patterns form locally smooth maps over the cortical surface (Hubel and Wiesel, 1962, 1968; Mountcastle, 1957). The functional organization of V1 is offering advantages through its efficient connectivity, as it minimizes (i) the distance between



**Figure 3:** Functional organization of a hypercolumn in primary visual cortex (V1). It is formed from a complete set of orientations, as well as pairs of ocular-dominance columns corresponding to both eyes. Interspersed are cytochrome oxidase blobs and interblobs selectively responding for color information. Modified from Kandel, Koester, Mack and Siegelbaum (2021).

neurons with similar preferences, and (ii) the number of neurons required to analyze different properties of a stimulus. Both aspects lead to economic usage of brain volume and high-speed information processing (Kandel et al., 2021).

Some of these columns are consisting of neurons with similar selectivity for the orientation of a visual stimulus. When viewed from the cortex' surface, this leads to repeating clockwise and anti-clockwise cyclings of orientation preferences across the cortex. The cycles in turn can form pinwheel-shaped patterns with sudden changes of orientations at their center (Blasdel and Salama, 1986; Bonhoeffer and Grinvald, 1991). The preference for a certain orientation originates from the shape of receptive fields: whereas the shape of retinal and LGN receptive fields is circular, multiple LGN cells are projecting to one simple cell in the primary visual cortex, which in turn leads to combined receptive fields

---

with elongated shape (Hubel and Wiesel, 1962).

Ocular-dominance concerns the integration of information from both eyes. As previously indicated in Sect. 1.1.4, retinal ganglion cells of either the ipsilateral or contralateral eye form segregated connections to the LGN. This segregation is also maintained in thalamocortical projections from separate layers of LGN to V1. This produces striped patterns (from surface view of the cortex) of ocular-dominance corresponding to inputs from the left or right eye (Hubel and Wiesel, 1972; Wiesel, Hubel and Lam, 1974).

Within the structure of orientation preference and ocular-dominance columns, populations of color-sensitive neurons with weak orientation selectivity are forming cytochrome oxidase blobs and inter-blobs. Due to their selectivity profile, they respond preferentially to surfaces rather than edges (Murphy, Jones and Van Sluyters, 1995; Shoham, Hübener, Schulze, Grinvald and Bonhoeffer, 1997).

A cortical patch which contains a full cycle of orientation preferences, or alternatively a left-plus-right ocular-dominance set, is commonly referred to as hypercolumn. This computational module is covering about 1 mm<sup>2</sup> of cortical surface and is repeated several hundred times, each representing a small subset from the visual field (Hubel and Wiesel, 1974).

To integrate information from different parts of the visual field, long-range horizontal connections are established within each layer of V1. These horizontal connections link columns having similar response characteristics, such as preferred orientation. This is achieved via axon collaterals of pyramidal cells, which connect to other pyramidal cells and inhibitory interneurons (Bosking, Zhang, Schofield and Fitzpatrick, 1997; Gilbert, Das, Ito, Kapadia and Westheimer, 1996; Gilbert and Wiesel, 1989).

Along the ventral and dorsal visual pathways, higher visual areas receive input from V1. Corresponding areas are not only receiving feedforward, but also passing informations via feedback connections to lower areas, latter also including the LGN. Due to the convergence and divergence at synaptic relays of the afferent visual pathway, receptive field size as well as complexity is increasing. This leads to the implication that feedback connections are holding an integrative function (Hupé et al., 1998). Still, interactions between feedforward and feedback visual processing are largely unknown and therefore the subject of ongoing research (Kandel et al., 2021).

---

## 1.2 Advanced Concepts of Cortical Information Processing

### 1.2.1 Brain States

Brain states have been traditionally linked with either behavioral states or a prevalent type of cortical states defined by correlated activity of neural populations, whereby both descriptions of brain states frequently overlap (Pace-Schott and Hobson, 2002; Sabri and Arabzadeh, 2018).

Functional brain states influence how sensory inputs are processed and control neuronal excitability at different spatial scales. On a whole-brain level, there is the classical view of two cortical states related to the sleep cycle: a sleep-like state characterized by synchrony and low-frequency fluctuations with high amplitude, and an awake-like persistent state operating more desynchronized and characterized by low-amplitude high-frequency fluctuations (Sanchez-Vives and McCormick, 2000; Steriade, Timofeev and Grenier, 2001; Stroh et al., 2013). Both synchronized and desynchronized state are likely two poles of a continuous spectrum of states. Furthermore, stages of arousal and anesthesia can shape these global network states as well (Harris and Thiele, 2011). Global network states can substantially influence cortical information processing by impacting synchrony of membrane potential dynamics (Poulet and Petersen, 2008) as well as interactions between sensory evoked responses and ongoing activity (Castro-Alamancos, 2004; Curto, Sakata, Marguet, Itskov and Harris, 2009).

On a more local level, states can be defined based on circuit excitability which could correspond to the alertness to certain sensory inputs (Galuske, Munk and Singer, 2019; Galuske et al., 2002; K. E. Schmidt, Lomber, Payne and Galuske, 2011).

On the level of a neuronal assembly, states are shaped by spontaneous and sensory-evoked activity and could reflect low-dimensional representations of specific sensory inputs and their functional processing. Especially in cortical areas such as the visual cortex, states may change on rapid timescales, directly impacting the processing of sensory input (Fries, Neuenschwander, Engel, Goebel and Singer, 2001; Waschke, Tune and Obleser, 2019).

On the level of single cells, the membrane potential is oscillating slowly between depolarization and hyperpolarization, which are associated with alternating periods of asynchronous irregular firing (Up state) and silence (Down state) (Steriade, Nunez and Amzica, 1993).

Despite their significance, relatively little is known about the underlying neural processes that generate brain states, as well as their exact influence on neuronal processing and behavior. It remains to be established to which extent such states emerge from

---

cellular activity across spatial scales, spanning from ensemble activity in a given cortical microcircuit to cortex-wide activity measures. Also, it has to be examined, how different brain states are related to each other and how they form internal low-dimensional representations of sensory information.

### **1.2.2 Spontaneous / Ongoing Activity**

Even in the absence of any external sensory input or motor activity, the brain is constantly active. This spontaneous (also known as ongoing) activity was initially reported by Caton (1875) in rabbits' and monkeys' brain and later by Berger (1929) in humans using electroencephalography (EEG) recordings. While no external stimulus is presented, single neurons as well as neuron populations exhibit activity patterns with high similarity to those of evoked responses (Arieli, Sterkin, Grinvald and Aertsen, 1996; Grinvald et al., 1999; Kenet, Bibitchkov, Tsodyks, Grinvald and Arieli, 2003). In this context, the structure of the underlying brain circuitry is reflected in the spontaneous activity, e.g. in terms of orientation columns (Tsodyks, Kenet, Grinvald and Arieli, 1999).

Experimental techniques to determine functional characteristics of various stimuli or tasks typically rely on averaging many trials to suppress background activity as well as noise components, while aiming solely at repeatable activity across trials. In the visual cortex of mammals, it has been shown that even under identical repetitive stimulation conditions, columnar activity shows variability in structural response. A growing body of research therefore emphasizes that dynamics of spontaneous, system-level activity modulates response activity while being also accountable for large parts of the observed trial-to-trial variability (Arieli et al., 1996; Kenet et al., 2003; Luczak, Barthó and Harris, 2009).

Through top-down effects that represent expectations, predictions, as well as attentional processes, spontaneous activity might contribute to internal processing of visual inputs (Galuske et al., 2002; Gilbert and Sigman, 2007; K. E. Schmidt et al., 2011) and may serve as an internal model for response activity patterns to sensory stimuli (Ringach, 2009). Thereby spontaneous activity manifests itself in modulations of the network's functional connectivity (Nauhaus, Busse, Carandini and Ringach, 2009).

When interpreted in a probabilistic framework, spontaneous activity acts as Bayesian prior for stimulus-evoked activity by expressing expectations about the sensory environment (Fiser, Berkes, Orbán and Lengyel, 2010; Luczak et al., 2009), which in turn offers shorter reaction times (Neal, 2001) and therefore poses an alternative explanation of the human visual system solving image classification tasks in such a time-efficient manner (Thorpe,



---

Fize and Marlot, 1996).

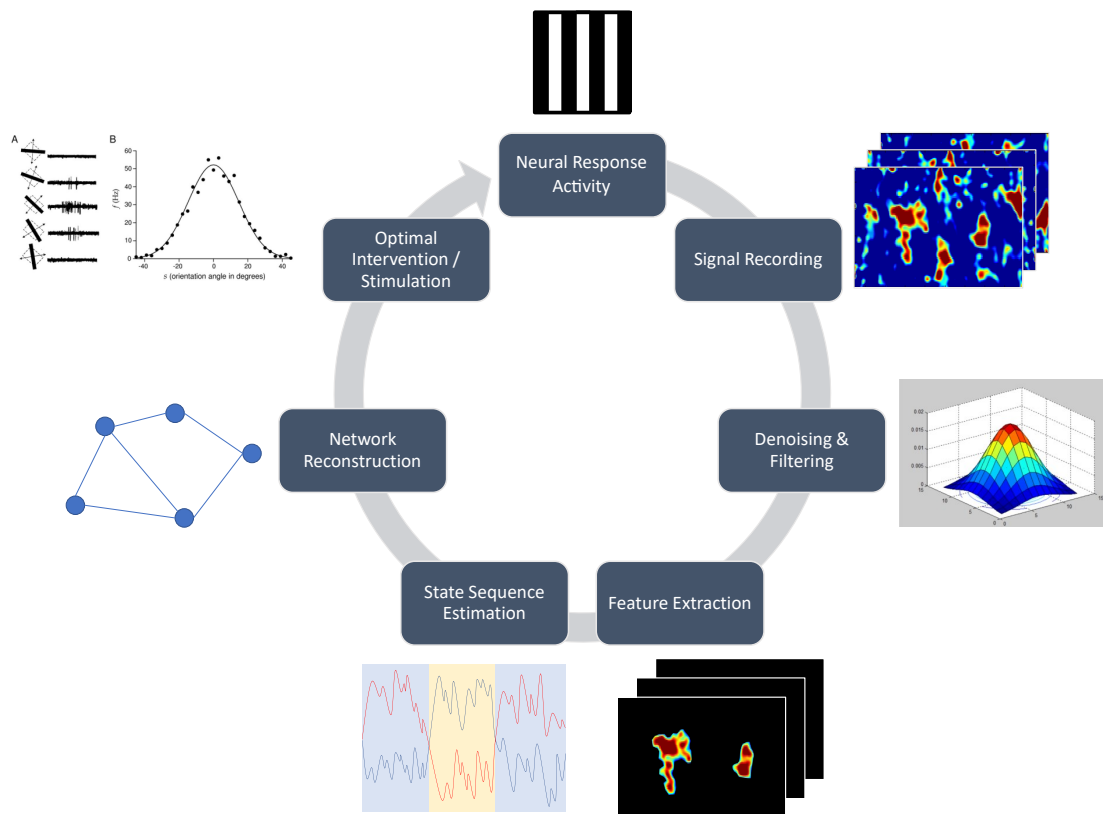
Further important findings regarding spontaneous activity in cortical populations concern the dimensionality of correlated neural activity, which spans large distances over millimeters of cortical surface (Arieli, Shoham, Hildesheim and Grinvald, 1995; Ts'o, Gilbert and Wiesel, 1986) and on different time scales (Kohn and Smith, 2005; Smith and Kohn, 2008). Ongoing activity patterns in the neighborhood of an individual neuron can impact its firing (Nauhaus et al., 2009; Tsodyks et al., 1999). Moreover, the amplitude of oscillations observed in spontaneous population activity is comparable to that of the mean reaction to a high-contrast stimulus (Arieli et al., 1995; Y. Chen, Geisler and Seidemann, 2006).

Consequently, the consensus has grown that spontaneous activity may have a functional role in perceptual processes which is connected to internal states of cell assemblies. This suggests the rejection of the hypothesis that spontaneous activity is merely representing stochastic noise, particularly as it is entailed with high energy costs (Attwell and Laughlin, 2001).

### **1.3 Motivation & Research Goals**

Functional brain states are possibly confined to local circuits or established as distributed patterns, and may crucially shape neural processes from perception to action (Bathellier, Ushakova and Rumpel, 2012; Pascual-Leone and Walsh, 2001). To better understand the interaction of brain states on different levels, it is necessary to analyze the relationship of activities in interconnected cortical areas. Particularly the variability of cortical responses to visual stimuli results from the interplay between global brain state, local states, spontaneous activity and sensory inputs (W. Chen, Park, Pan, Koretsky and Du, 2020).

To disentangle this complex interplay, one requires high-resolution data acquisition techniques together with causal manipulations of neural circuits during recording, and classifying the respective macro- and microstates instantaneously. In this context, rapid advances in experimental techniques can be witnessed in systems neuroscience. While the number of neurons that can simultaneously be recorded is doubling every seven years on average (Stevenson and Kording, 2011), new techniques allowing for controlled interventions on the single-cell level are quickly emerging (Fenno, Yizhar and Deisseroth, 2011). Integrating both advances together with data analysis opens up the opportunity for closed-loop experimental paradigms (as exemplified in Fig. 4). Here, data-driven inter-



**Figure 4:** Exemplary illustration of a hypothetical closed loop experimental paradigm. Starting from arbitrary stimulation or intervention (e.g., a grating pattern as visual orientation stimulus), the corresponding response activity is recorded via optical imaging. After pre-processing and denoising the raw image sequences, relevant features related to neural signal sources have to be extracted. Those features are then further used to classify different brain states and to estimate their respective state-switching sequence. Subsequently, these state sequences can then be employed for network reconstructions. Depending on the individual research focus, these networks can comprise different spatial scales such as intra- or inter-area domains. Eventually, the emitted activity patterns can be used to determine an optimal response configuration of the network, suggesting variations of the stimulus or intervention configuration in an iterative fashion. In case of the orientation stimulus, this would result in an orientation tuning curve.

---

rogations of neural assemblies are performed through direction intervention or external simulations in real-time or within small delays. Hypothesis-driven and model-driven manipulations are key to move from plain covariation to true causality relations (Pearl, 2009).


However, processing and analysis of high-dimensional neural datasets on its own is a challenging task. Closing the real-time loop makes this challenge even more pronounced, as these data analysis methods are constrained to be both

- *computationally efficient* to enable the fast application of subsequent stimulus or intervention adaptations with time delays that are sufficiently small relative to the time-scale of the investigated neural process;
- *robust* with respect to measurement artifacts and model mismatch to avoid erroneous interpretations and to prevent error accumulation through the closed-loop system.

Both aspects raise a trade-off between computational complexity and robustness, leading to difficult algorithmic challenges.

This thesis is therefore dedicated to the development of tailored computational methods for contributing to the understanding of functional brain states and functional connectivity on multiple scales. Using optical imaging of voltage-sensitive dyes (Grinvald, Lieke, Frostig and Hildesheim, 1994), it is possible to record activity dynamics of large neural populations with high spatio-temporal resolution. Resulting image sequences exhibit high dimensionality, complex neural activity patterns as well as highly variable confounding artifacts. For the goal of fast and robust extraction of relevant features, dimensionality reduction techniques can prove useful. By choosing a deep generative model architecture, particularly the framework of a variational autoencoder (Kingma and Ba, 2014), advantages from both deep learning and probabilistic modeling are combined for constructing latent representations (Kingma and Welling, 2019).

Deep learning usually requires large amounts of annotated data for adequate model training. Through annotations, explicit knowledge about the relationship between a data-point and corresponding ground truth information is reflected. On pixel-level, this would express e.g. the membership of a pixel to an object within an image. However, for real datasets such as optical imaging data, ground truth information is not available. Also, as the data collection process is expensive and time-consuming, the number of recordings is substantially limited. The use of synthetic data can alleviate both problems, as an infinite amount of datasets can be artificially generated with pixel-perfect labeling due to a fully known data-generating process. This enables benchmarking and accuracy improvement of



---

the implemented deep learning model (Nikolenko, 2021). For these purposes, a pipeline for generating synthetic image sequences corresponding to voltage-sensitive dye imaging (VSDI) is implemented based on previous insights about fundamental spatial and temporal signal components (Chemla et al., 2017; Reynaud, Takerkart, Masson and Chavane, 2011).

---

## 2 Materials & Methods

---

### 2.1 Data Acquisition

#### 2.1.1 Voltage-Sensitive Dye Imaging

Neural response activity related to visual processing arises from complex, dynamic interactions in vast cortical networks. For understanding the function of an involved cortical region, it is necessary to track these dynamics of neural populations with high spatial and temporal resolution. Despite the fact that this trade-off between both domains has long been recognized, the sole focus on either the spatial or temporal elements of cortical functions has been predominant due to limitations in conventional recording techniques (Shoham et al., 1999). Metabolic changes generated by neural activity are often evaluated with high spatial resolution but are time-limited, as it is the case for e.g. 2-deoxyglucose mapping (2-DG), positron emission tomography (PET), optical imaging of intrinsic signals (OIS), near infrared spectroscopy (NIRS), and functional magnetic resonance imaging (fMRI). Corresponding metabolic indicators - whose relationship to neural processes is not always straightforward - are much slower when compared to neural dynamics, therefore leading to imaging approaches with a temporal resolution of a few hundred milliseconds. In contrast, electrophysiological recordings are often well resolved in time but spatially limited, leading to mere point measurements in terms of intracellular, extracellular single or multiunit recordings, or local field potentials (Shoham et al., 1999). Determining the signal sources becomes especially intricate for electroencephalography (EEG) and magnetoencephalography (MEG), as inference of the position of the current sources from electrode potentials is not uniquely identified. Latter aspect is also known as the inverse problem (Grech et al., 2008).

Optical imaging has been proposed as one potential alternative, as this recording technique is easily expandable in terms of finding a compromise between temporal and spatial resolution. The term *optical imaging* refers to imaging methods initially used for observing intrinsic signals in terms of the light-dependent absorption or emission effects on active neurons (Hill and Keynes, 1949). Since then, intrinsic signals such as haemodynamic or

---

light-scattering (Grinvald, Lieke, Frostig, Gilbert and Wiesel, 1986) were recorded, providing high-resolution mapping of functional cortical organization, such as the architecture of the visual cortex in macaques (Blasdel and Salama, 1986) and cats (Frostig, Lieke, Ts'o and Grinvald, 1990). On the other hand, intrinsic signals are unrelated to electrical activity and therefore are limited in terms of their temporal resolution on the order of seconds (Grinvald and Hildesheim, 2004).

An important extension of optical imaging is the usage of specially adapted and synthesized voltage-sensitive fluorescent dyes, leading to the methodology of VSDI. First reports on squid giant axons by Tasaki, Watanabe, Sandlin and Carnay (1968) and Cohen et al. (1974) certified that these dyes are sensitive to changes in membrane potentials and can therefore be used to indirectly monitor neural activity. When exposed cortical tissue is stained with an appropriate dye, its molecules adhere to the cell membrane's surface. The chemical properties of the dye enable molecular transduction, converting changes in membrane voltage into an optical signal which occurs due to changes in absorption or emitted fluorescence. When illuminated with light-imaging devices at the dye's peak excitation spectra, a fluorescent light is emitted and image sequences of the fluorescing cortex can be recorded with high-speed cameras. This allows for measurements on mesoscopic scale, i.e., a network of neurons from a cortical column to a whole area with high resolution in both temporal (1-10 ms) and spatial ( $< 50 \mu\text{m}$ ) dimensions (Grinvald and Hildesheim, 2004; Grinvald et al., 1999; Shoham et al., 1999). In this context, the spatial resolution is mainly limited by light scattering of the emitted fluorescent signal (Orbach and Cohen, 1983).

However, resolving the contribution of different neural units to the VSDI signal remains difficult. For *in vivo* measurements, it was shown that the voltage-sensitive dye signal accurately conveys membrane voltage changes. This was verified by pairing direct intracellular recordings with VSDI, first measured in anaesthetized cat (Grinvald et al., 1999; Sterkin, Lampl, Ferster, Grinvald and Arieli, 1998) and later for rat barrel cortex (Petersen, Grinvald and Sakmann, 2003; Petersen, Hahn, Mehta, Grinvald and Sakmann, 2003). Previous limitations due to pharmacological side effects and phototoxicity have been resolved by the development of newer dyes (Shoham et al., 1999). The fluorescent VSDI signal is linearly related to the membrane area stained with the dye. Here, the recorded dynamics of each measuring pixel is reflecting multiple neural compartments, such as dendrites, axons and somata of cell populations. As dendrites and non-myelinated axons take up an area which is orders of magnitude larger than for somata, the *in vivo* VSDI signal mostly reflects dendritic activity (Grinvald and Hildesheim, 2004). Also, different cell types (both excitatory and inhibitory), are affected equally by the dye staining procedure.

---

For glial cells, it was shown in frog optical nerve that dye molecules can bind to those cells which are in turn weakly contributing to the transmitted light intensity with slow depolarizing afterpotentials (Konnerth and Orkand, 1986). This leads to the conclusion that the dye signal in a cortical region does not inevitably indicate for action potentials of cortical neurons at that location (Grinvald and Hildesheim, 2004).

The overall VSDI signal strength is impacted by the staining quality of cortical tissue, which is especially the case for in vivo experimental procedures (Takagaki, Lippert, Dann, Wanger and Ohl, 2008). Therefore, the dye concentration has to be chosen carefully to be both hydrophilic enough to pass through the outer layer of neural tissue, yet hydrophobic enough to stick to the cell membrane (Grinvald, Anglister, Freeman, Hildesheim and Manker, 1984).

As commonly seen for other fluorescence-based recording techniques, dye photobleaching is also affecting VSDI. In combination with the baseline fluorescence level, which can highly vary between recordings, this makes the strongest artifact of this recording technique. After excitation through illumination, the dye fluorophores are degrading or reacting with other molecules, and hence are stopped from releasing light. This leads to degradation of the overall fluorescent signal over time (Grinvald and Hildesheim, 2004; Grinvald, Hildesheim, Farber and Anglister, 1982). It is typically observed in functional form of a slowly decaying single-exponential (Bathellier, Van De Ville, Blu, Unser and Carleton, 2007) or double-exponential (Gavrilyuk et al., 2007).

For in vivo experiments, signal components related to the physiology of the experimental animal are dominant sources of noise commonly observed as periodic signals. These artifacts are related on the one hand to the animals' heartbeat resulting from blood pumping, on the other hand from respiration (Inagaki, 2003; Shoham et al., 1999). As the cortical surface will consequently pulsate, this can lead to changes in the focal plane of the camera and therefore affect the global light conditions. Furthermore, the VSDI signal can still be overlaid by spectral components related to the excitation of oxygenated and deoxygenated blood flow phases in the peripheral cortical blood vessels (Hofmann, 2020), although this effect should be reduced due to the RH-1691 dye's emitting range between 645 nm to 665 nm and therefore being slightly higher compared to hemoglobin (Ratzlaff and Grinvald, 1991).

Camera-related technical noise, especially in form of shot noise, is a limiting factor of VSDI. In case of exposure, light photons do not run vertically, but are deflected through their natural scattering property. This causes them to hit neighboring camera pixels, contaminating the recorded image in a stochastic manner. To further reduce illuminatory artefacts, the light source should be incoherent and filtered to the desired excitation spectrum, e.g. a filtered filament halogen bulb light. This removes speckle noise known

---

from coherent light sources (e.g. lasers) reflected by uneven surfaces, which introduces undesired structural spatial fluctuations through coherent interference patterns (Kompanets and Zalyapin, 2020). The shot noise is proportional to the square root of the used light intensity, therefore this source of interference can be reduced by increasing illumination (Grinvald et al., 1999). However, when using high intensities of light, the high number of photons accumulates into electrons. If the amount of light exceeds the possible number of electrons per pixel (well depth), over-saturation will occur in respective pixel. This can be solved either by cameras with larger sensor arrays or by applying pixel binning. Latter is indicating the spatial combination of several neighboring pixels into one macropixel (Grinvald et al., 1999).

### **2.1.2 Experimental Setup**

The following informations concern the data acquisition and stimulation methodology of optical imaging recordings which were analyzed in this thesis. All real datasets assessed for the present study were recorded in area 18 (area occipitalis), which together with area 17 has been defined as cat's primary visual cortex (Payne and Peters, 2002). For the present study, pre-recorded datasets from an adult cat (23 months old) were analyzed. This experimental animal came from the breeding of the Max Planck Institute for Brain Research (Frankfurt a. M.). Investigations and required surgical procedures were carried out in 2013 within the framework of an approved animal experimentation procedure and in accordance with the German Animal Welfare Act.

It should be noted that the author did not perform optical imaging experiments but was provided with corresponding raw datasets. Experiments were carried out by Dr. Daniel Hofmann, Fabian Hoffmann, Dr. Mathias Peter, Dr. William Barnes and Prof. Dr. Ralf Galuske. The same datasets have been analyzed by Peter (2019) and Hofmann (2020) with distinct research foci. These publications contain in-depth information, particularly about medical and surgical procedures, histological analysis as well as the camera setup. Therefore, only a broad outline will be given here.

#### **Anesthesia**

Initial anesthesia was provided by intramuscular administration of ketamine hydrochloride (ketamine 10%, 10 mg per kg body weight; Bela-Pharm GmbH & Co. KG, Vechta) and xylazine hydrochloride (1 mg per kg body weight; Rompun 2, BayerVital, Leverkusen). Atropine sulfate (50% atropine sulfat in NaCl, 0.2 ml per kg body weight; Fresenius Kabi, Bad Homburg) was intramuscularly injected for stabilizing circulation. Animals were



---

artificially ventilated with a mixture of N<sub>2</sub>O (70%), O<sub>2</sub> (29%), and halothane (1%) to maintain inhalation anesthesia via an vaporizer (Halothan Vapor 19.3, Drägerwerk AG, Lübeck). Throughout the experiment, concentrations of O<sub>2</sub>, CO<sub>2</sub> and halothane were strictly monitored in inhalation and exhalation (Peter, 2019).

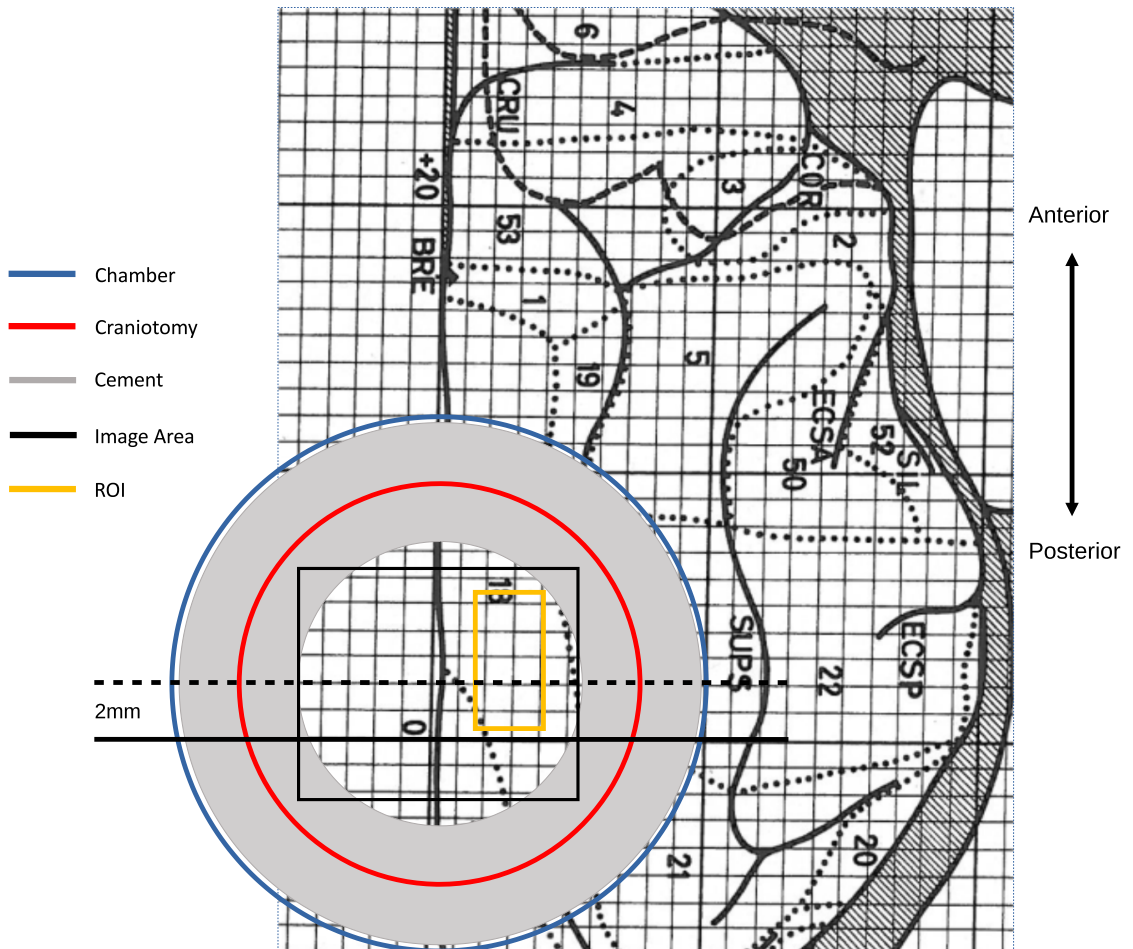
### **Camera Setup**

A high-speed CMOS-camera (Photonfocus, MV1 D1312-160-CL-12) was used as optical sensor system, offering a maximal resolution of 1312 x 1082 pixels. Each pixel captures an area of 8 x 8  $\mu\text{m}$  of the cortex. Because signals from higher neocortical layers are of major relevance, the focus was fixed 600  $\mu\text{m}$  below the cortical surface to compensate for blood vessel-related image artifacts. Illumination was conducted with an excitation wavelength of 630 nm using a halogen cold light source (150W Phillips, Type 6550, Al/234, 15V, G6.35) and subsequent infrared filter and bandpass filter of 630  $\pm$  10 nm (Schott Glas, Mainz, Germany). The light was directed through a macroscope perpendicularly onto the cortical surface using a dichroic mirror. Subsequently, a high-pass filter (Schott Glas, Mainz, Germany) ensured that only photons of wavelengths  $\geq$  665 nm reaching the camera chip's pixels and transmitted energy to the semiconductor electrons, which were then released (Hofmann, 2020; Peter, 2019). As the camera's readout rate is depending on spatial resolution and exposure time, an optimization of both parameters is required to minimize technical noise components typically observed for CMOS systems (e.g. read noise, dark noise). Therefore, the setup was configured to a frequency of 150 Hz and 1024 x 990 pixels with an exposure time of 0.01 ms, which leads to a temporal resolution of 6.7 ms per image (Hofmann, 2020).

Area 18 of cat's V1 was defined as the target of the optical imaging procedure. To determine the ideal position of the imaging chamber, the center of the interaural line was established as the zero point of a three-dimensional coordinate system. From this point, the location of the chamber can be exactly aligned to the position of area 18, which is illustrated in Fig. 5. This method is based on a combination of Horsley-Clarke stereotactic coordinates (Horsley and Clarke, 1908) and the topographic atlas of the cat brain according to Reinoso-Suárez (1961). The chamber was centered in the midline +2 mm on the anterior-posterior axis (Peter, 2019).

### **Voltage-Sensitive Dye**

The voltage-sensitive dye RH-1691 (Optical Imaging Ltd., Israel), was utilized as reporter in the conducted optical imaging experiments (Hofmann, 2020). It is a "water-soluble aromatic anionic oxonol compound based on a vinylogous carboxylate conjugate system"



**Figure 5:** Schematic overview of the applied optical imaging area. The dorsal view of cat's right hemisphere is illustrated via the coordinate system from the Reinoso-Suárez (1961) brain atlas, also showing the numeration and boundaries of cortical areas (dotted lines). Axes show dimensions in mm with a zero point on the anterior-posterior axis. The implanted imaging chamber (blue circle) positioned over area 18 and the craniotomy area (red circle) are overlaid. For stabilizing the chamber construction, a ring of dental cement (grey ring) was used. The image area (black rectangle) is containing the region of interest (ROI, yellow rectangle). Modified from Reinoso-Suárez (1961).

---

(Lippert, Takagaki, Xu, Huang and Wu, 2007). These properties enable its external integration mostly into the lipid bilayer of plasma membranes. Through conformation changes it responds to brief changes in membrane voltage in terms of fluorescence, which is linearly related to the membrane voltage. The fluorescence signal is mainly originating from cortical dendrites and non-myelinated axons, as respective membrane areas are orders of magnitudes larger in comparison to cell somata (Grinvald and Hildesheim, 2004; Lippert et al., 2007; Shoham et al., 1999). The dye absorbs light with an excitation wavelength of 630 nm and in turn emits the fluorescent signal with wavelengths of  $> 650$  nm. Excitation wavelengths of the dye RH-1691 have little overlap with the absorption spectrum of hemoglobin. Furthermore, this distance between the spectra of blue dyes and hemoglobin is reportedly minimizing movement artifacts due to the pulsation of blood vessels and brain tissue. These movements are caused by a pulse wave which can be attributed to the contraction of the heart's left ventricle during systole (Grandy, Greenfield and Devonshire, 2012; Lippert et al., 2007; Shoham et al., 1999).

## 2.2 Data Pre-Processing

### 2.2.1 Hardware-Based Approaches

For enabling analyses of optical imaging sequences on basis of single trials, ensuring high data quality is of utter importance. Already at the level of the data acquisition process, several biological and technical phenomena can be accounted for via hardware-related approaches. Artifacts due to cardiovascular and respiratory activity of the experimental animal can be significantly reduced by synchronization of the optical imaging recording setup. Latter consisted of the following hardware components:

- optical imaging system Imager3001 (Optical Imaging Inc., Rehovot, Israel);
- data acquisition computer with imaging software VDAQ2.5 (Optical Imaging Inc., Rehovot, Israel);
- high-speed CMOS camera (MV1 D1312-160-CL-12; Photonfocus, Lachen, Switzerland)
- stimulation unit containing software StimulPL (Prof. Dr. Rainer Goebel, Maastricht University, Netherlands) connected to a 21-inch CRT monitor (Accuvue, HM4921D);
- electrophysiological recording system;
- respiratory pump (Ugo Basile 6025, Gemonio, Italy)

---

(Hofmann, 2020; Peter, 2019).

All components of the setup were synchronized via the Imager3001 as central interface by using transistor-transistor logic (TTL). The Imager3001 first primes the stimulation unit, which enters a waiting state, for the upcoming stimulation condition and sends a stop signal to the respiratory pump, latter stopping after the next cycle for 1.7 – 3 s and sending a signals back to the Imager3001. The interface waits for the next imidiate QRS peak of the heartbeat and instantly transmits a simultaneous TTL start trigger to the electrophysiological system, the camera/shutter and stimulation unit. Consequently, each recording started at the same phase of the animal's heartbeat (Hofmann, 2020; Peter, 2019).

Aside of synchronizing the optical imaging hardware components, motion artifacts can be further mitigated by physically stabilizing the anesthetized experimental animal. This was achieved by fixation of the animal's body within a stereotaxic frame. By using a head holder attached with dental cement and bone screws to the skull, head movements were consequently minimized. Furthermore, by filling the imaging chamber with incompressible silicone oil (DS Fluid, Boss Products, Elizabethtown, Kentucky, USA) and sealing it with a silicone ring and a glass plate, movement artifacts due to heartbeat and respiration as well as swelling of the brain tissue were further decreased (Peter, 2019).

### **2.2.2 Software-Based Approaches**

Even after incorporating hardware-based approaches to increase the SNR, the overall VSDI signal is confounded by several noise components. Their respective impact on the signal is depending on multiple factors related to data acquisition indicated in the following (non-exhaustive) list:

- hardware-related aspects: camera, dye composition
- recording-related aspects: recording duration
- organism-related aspects: animal species, awakesness vs. anesthesia
- activity-related aspects: evoked vs. ongoing activity

(Raguet et al., 2016).

It is therefore necessary to separate neural activity dynamics from the confounded VSDI signal. This makes software-based pre-processing a crucial step for adequate data analysis. In the following section, a set of methods commonly used for VSDI data are further described.

---

## Blank Subtraction

For data pre-processing, blank subtraction (Blasdel, 1992; Shoham et al., 1999) poses a traditional approach for establishing a reference signal level and for removing sources of noise such as uneven illumination (Grinvald et al., 1999). Its initial step consists of dividing all frames of a recording by its respective baseline level, latter being calculated as the average of multiple frames prior to stimulus onset (commonly known as zero-frame or z-frame). Subsequently, the blank signal is computed as the average over all available blank recordings within an imaging session, for which no stimulation was carried out. This average is then subtracted from all stimulus-evoked sequences in pixel-wise manner (Arieli et al., 1995; Grinvald et al., 1994; Shoham et al., 1999). However, this method is substantially noise-sensitive as pixel variability of the evoked and blank recordings is summed up (Bathellier et al., 2007). Also, the initial frame division usually poses an inaccurate normalization procedure which was reported to lead to dynamical biases in amplitude quantification of neural activity (Takagaki et al., 2008). By applying baseline subtraction, several problematic assumptions are implicitly made: artifacts are additive noise components, which all are proportional to the baseline fluorescence level and whose dynamics are identical for stimulus-evoked and blank regimes (Raguet et al., 2016). A common variation of blank subtraction is the usage of a *cocktail blank*. Instead of generating an image of the unstimulated cortex, the goal is to obtain an image of the uniformly activated cortex as reference. Latter is estimated as the sum of response activities to a set of all available stimulus configurations (e.g. different orientations of a visual stimulus), which is then used to normalize activity maps (Grinvald et al., 1999). By this approach, non-stimulus related responses can be eliminated, yet it is accompanied by an information loss regarding the overall neural dynamics (Raguet et al., 2016). As both blank approaches require multiple trials for computing averages, they are unsuitable for single trial analysis (Reyraud et al., 2011).

## Principal Component Analysis & Independent Component Analysis

Principal component analysis (PCA) (Hotelling, 1933; Pearson, 1901) is a widely known approach for dimensionality reduction, data compression and feature extraction (Jolliffe, 2002). It decomposes the input signal into additive, uncorrelated components. The input data is projected orthogonally on a reduced linear subspace via eigenvectors of the covariance matrix, the *principal components*. The optimization criterion can be defined by minimizing the mean squared distance between data points and projections (Pearson, 1901) or by maximizing the variance of projected data (Hotelling, 1933). Principal components are derived by solving eigenvalue decomposition of the covariance matrix,

---

or alternatively from singular value decomposition of the centered data matrix. As those principal components are defined just from the respective input data, components do not have to be specified a priori. Thus, this technique is very adaptable to many different data types in various disciplines (Jolliffe and Cadima, 2016).

Another blind source separation technique known as independent component analysis (ICA) (Hyvärinen and Oja, 2000) targets for separating latent sources from an observed mixture signal by linear representations. This is based on the assumptions of statistical independence and non-Gaussianity of components, as well as an unknown linear mixing system (Hyvärinen and Oja, 2000). For pursuing its corresponding goals of i) maximization of non-Gaussianity, ii) minimization of mutual information and iii) maximum likelihood estimation, several algorithms like FastICA (Hyvärinen, 1999), projection pursuit and Infomax (Hyvärinen and Oja, 2000) have been developed. Regarding VSDI, application of ICA has been focused either on the temporal data domain by treating individual pixels as observations, e.g. for the task of extracting neural dynamics from heartbeat- and respiration-contaminated trials recorded in primary auditory cortex of guinea pig (Inagaki, 2003; Maeda, Inagaki, Kawaguchi and Song, 2001), or on the spatial domain by specifying each frame as observation, e.g. for extracting functional cortical maps in olfactory bulb and the somatosensory cortex of mice as well as the visual cortex of monkeys (Reidl, Starke, Omer, Grinvald and Spors, 2007). For taking into account both temporal and spatial dimensions, ICA has been combined with complementary approaches. For instance, combinations of temporal ICA with local similarity minimization on the spatial domain have been applied to remove spatial biological artefacts from trials recorded in primary cortices of awake monkeys and anesthetized cats (Fekete, Omer, Naaman and Grinvald, 2009).

PCA and ICA are offering convenience due to their possibility to separate sources directly from data acquisition. Nevertheless, components can only be classified a posteriori. Therefore, an identified component can comprise both signal- and noise-related features (Chemla et al., 2017).

### **Linear Modeling**

An alternative approach for denoising and feature extraction of neural activity is multiple linear regression, which was initially applied to fMRI data (Friston et al., 1994). With respect to VSDI recordings acquired in awake monkey, Reynaud et al. (2011) use a General Linear Model (GLM). It is specified as finite weighted-sum of pre-defined regressors  $X$  to approximate the measured signal  $y$  for each trial, so that

$$y = X\beta + r \tag{1}$$

---

, with 2-D matrix  $X$  (containing all regressors on separate columns), vector of weights  $\beta$  and vector of residuals  $r$  (Reynaud et al., 2011).

The best linear unbiased estimator  $\hat{\beta}$  of weights is obtained via the Moore-Penrose pseudo inverse  $X_+$  of  $X$ , assuming that the residuals  $r$  are following statistical white noise. The estimator is constructed as

$$\hat{\beta} = (X'_+ X_+)^{-1} X'_+ y \quad (2)$$

(Reynaud et al., 2011).

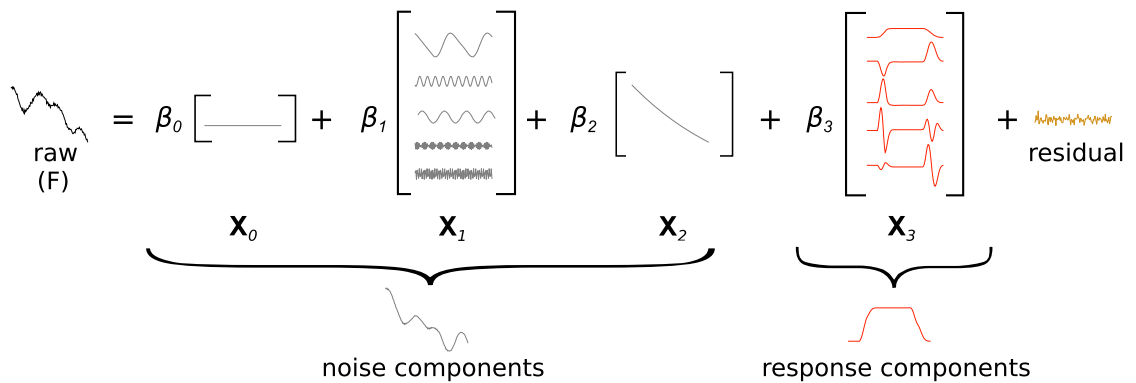
The main assumptions of this model comprise linear independence of all regressors as well as their linear additive contribution to the signal. In this context, Reynaud et al. (2011) were building matrix  $X$  from the following regressors illustrated in Fig. 6:

- the baseline fluorescence level  $X_0$ , considered as constant value
- technical and physiological artifacts  $X_1$ , such as fundamental heartbeat frequency or respiration, which are interpreted as periodic oscillations and, therefore, modeled as Fourier series allowing for phase change
- the dye bleaching  $X_2$  modeled as decaying exponential function

(Reynaud et al., 2011).

When subtracting the weighted linear combination of regressors from the measured signal, the residuals contain neural response activity as well as statistical white noise, but also non-deterministic spontaneous neural activity. In case that spontaneous activity is central to the research question, it would be necessary to explicitly account for evoked neural activity. For this purpose, template shapes for the expected response  $X_3$  can be included additionally in the linear model. This approach was initially developed for fMRI by Hossein-Zadeh, Ardekani and Soltanian-Zadeh (2003) and Woolrich, Behrens and Smith (2004) as fMRI's linear optimal bias sets (FLOBS). For VSDI, this was adapted by including a set of temporal regressors, which are reflecting the first eigenvectors obtained from singular value decomposition of a large number of artificial response changes (Chemla et al., 2017). This dimensionality reduction of neural responses marks a critical step of the GLM framework, as it depends on the complexity of stimuli which is affecting pixel-wise response dynamics, for example in terms of delays as well as rising and decreasing times. Because regressors have to be specified a priori, this method allows for single-trial analysis. However, as recorded time-series are processed independently for each pixel, spatial dependencies between pixels within the observed images are ignored (Chemla et al., 2017; Raguet et al., 2016).





**Figure 6:** Linear model decomposition of the VSDI signal by Reynaud, Takerkart, Masson and Chavane, 2011. The raw pixel dynamics (F) is modeled via multiple regressors related to neural response dynamics as well as noise signal sources. Latter comprise the baseline fluorescence level ( $X_0$ ), oscillatory signals such as heartbeat or respiration ( $X_1$ ) and bleaching behavior of the dye ( $X_2$ ). Different response templates are represented by  $X_3$ . The weighted sum of all regressors is normalized together with the residuals using the baseline activity ( $X_0$ ). Modified from Chemla et al., 2017.

## 2.3 Data Modeling

### 2.3.1 Bayesian Inference

To illustrate concepts like generative or latent variable modeling, the following basic terms of Bayesian statistics will be used:

- the observed data vector  $x$
- the unobserved latent variables  $z$
- the prior distribution  $p(z)$
- the likelihood  $p(x|z)$
- the joint distribution  $p(x, z) = p(x|z)p(z)$
- the posterior distribution  $p(z|x)$ .



---

To infer the unknown but observable  $x$  in latent variable modeling, the *marginal* or *prior predictive distribution* is defined as

$$p(x) = \int p(x, z)dz = \int p(x|z)p(z)dz \quad (3)$$

, which can be used for predicting an unknown observable, or – as it will be relevant from the perspective of generative modeling – is providing the information of the probability of generating a data point (Gelman, Carlin, Stern and Rubin, 2003).

### 2.3.2 Generative Modeling

Pattern recognition tasks such as image segmentation can be broadly categorised in *supervised*, and *unsupervised* learning problems. Supervised learning is possible if the training data comprises both an input data vector  $x$  and a corresponding target vector  $y$ , latter being usually denoted as data labels. Here, a function is learned to map  $x \rightarrow y$ . Typical applications are classification and regression tasks such as assigning images of handwritten digits to their corresponding digit label, commonly demonstrated via the Modified National Institute of Standards and Technology (MNIST) database. In unsupervised learning, such data labels  $y$  are not available, therefore the goal is to learn some hidden structure only from the data  $x$ . This includes tasks like clustering, density estimation or visualization (Bishop, 2006).

Another categorization of machine learning models can be done by differentiating between *discriminative* and *generative* models. Discriminative models are set out to model the posterior conditional probability  $p(y|x)$  directly, or model a function  $f(x)$  for learning a direct mapping of  $x \rightarrow y$ . On the other hand, generative models represent a branch of unsupervised learning and are learning a model of the joint distribution  $p(x, y)$ , which makes it possible to predict  $p(y|x)$  via Bayes rule (Ng and Jordan, 2001).

Let  $x$  again represent the data having the probability distribution  $p(x)$ . From the perspective of generative modeling, the joint distribution  $p_{\theta}(x, z)$  describes the generative process for observing the data  $x$  given the unobserved latent variable  $z$ :

$$p_{\theta}(x, z) = p(x|z)p(z) \quad (4)$$

$$z \sim p(z) \quad (5)$$

$$x \sim p(x|z) \quad (6)$$

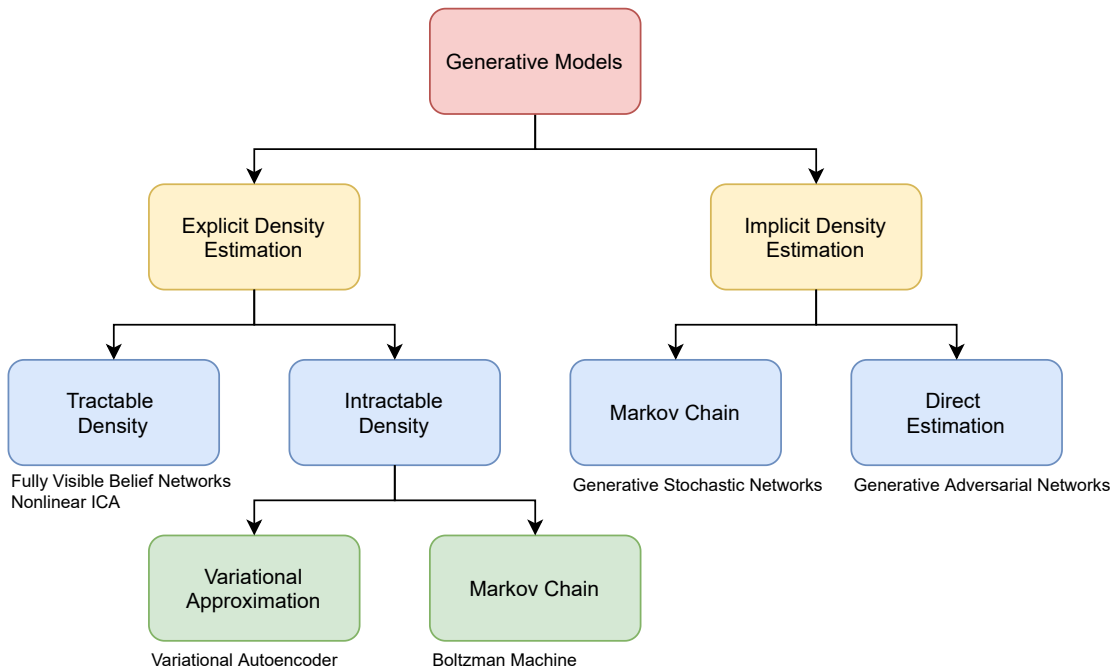
---

By knowing the data distribution  $p(x)$ , several goals can be pursued:

- generating a new data point  $x^{(i)}$  by first sampling  $z^{(i)} \sim p_\theta(z)$  from the prior distribution  $p_\theta(z)$  and subsequently sampling  $x^{(i)} \sim p_\theta(x|z^{(i)})$  from the conditional distribution  $p_\theta(x|z^{(i)})$ , latter known as the likelihood
- evaluating  $p(x)$  for a test data point
- extracting the latent representation  $z$  from the data  $x$ .

However, depending on the complexity of the underlying model class, these goals are either impossible or only achievable via approximations (Fraccaro, 2018).

Generative models can then be further classified into *implicit* and *explicit* density models, as illustrated in Fig. 7. On the one hand, models defining an explicit density function  $p(x; \theta)$  can be used for estimating the likelihood for a single data point  $x$ . More formally stated by Goodfellow (2016), these models maximize the likelihood by “simply [plug] the model’s definition of the density function into the expression for the likelihood, and follow the gradient uphill” (Goodfellow, 2016). As illustrated in Sect. 2.3.4, these explicit densities may be computationally tractable or intractable, latter requiring approximation techniques such as variational or Monte Carlo approximations. Examples for tractable explicit density models are fully visible belief networks (Frey, 1998; Frey, Hinton and Dayan, 1995) and nonlinear ICA (Burel, 1992; Deco and Brauer, 1995), while variational autoencoders (VAE) (Kingma and Welling, 2013) can be classified as intractable explicit density models. On the other hand, implicit density models do not represent a probability distribution  $p(x)$  over the data space. Instead, they are based on sampling from the underlying distribution after model training. This is done either by defining a Markov chain transition operator and running it several times to get a sample, as it is the case for generative stochastic networks (Bengio, Laufer, Alain and Yosinski, 2014); or these samples can be drawn directly in a single step, as in generative adversarial networks (Goodfellow et al., 2014).



**Figure 7:** Taxonomy of generative models based on maximum likelihood principle. These models can be broadly classified depending on their method of representing or approximating the maximum likelihood. Adapted from Goodfellow (2016).

### 2.3.3 Latent Variable Modeling

Latent low-dimensional representations are based on the assumption that high-dimensional observations  $x$  are generated by a (usually substantially) smaller number of latent variables  $z$  that are not directly observable. Traditionally, the observations are assumed to be related to the latent variables through an unknown linear or nonlinear transformation (Bengio, Courville and Vincent, 2013; Calabrese, Schumacher, Schneider, Paninski and Woolley, 2011; Parsons, Haque and Liu, 2004). The objective of dimensionality reduction techniques is to identify the lower-dimensional subspace in which the latent variables reside. In this context, the goal is to find compact descriptions of the high-dimensional data for the purpose of increased interpretability and reduction of computational complexity in subsequent signal processing steps (Cunningham and Yu, 2014; Nonnenmacher, Turaga and Macke, 2017).

---

To give a more formal definition, let the complicated data distribution  $p(x)$  be the modeling target. Here, each observation  $x$  is depending on the unobserved latent variables  $z$ . Their joint distribution can be stated as  $p(x, z) = p(x|z)p(z)$ , where  $p(x|z)$  is the likelihood of the observations  $x$  given the latent variables  $z$ , and  $p(z)$  is the prior distribution of the latent variables  $z$ . Both  $p(x|z)$  and  $p(z)$  are much simpler to define than  $p(x)$ . It can easily be shown that by marginalizing over the latent variables  $z$ , the data distribution  $p(x)$  can be obtained again:

$$p(x) = \int p(x, z)dz = \int p(x|z)p(z)d(z). \quad (7)$$

To infer the unobserved latent variables  $z$  given the observed data  $x$ , the posterior distribution  $p(z|x)$  can be computed as

$$p(z|x) = \frac{p(x, z)}{p(x)} = \frac{p(x|z)p(z)}{p(x)}. \quad (8)$$

which results from the application of Bayes' theorem. The posterior distribution  $p(z|x)$  reflects the prior belief about the latent variables  $z$ , which is confronted by the observed data  $x$  and updated accordingly.

Much can be gained from capturing the low-dimensional structure of neural activity and neural representations. Extracting important dimensions or meaningful latent variables can elucidate important structural and dynamical properties of the brain. This includes the manifold in activity space covered by multiple activity patterns and the mapping of the stimulus space onto a given neural activity pattern from whole brain to local circuitry. Computational methods and insights gained in this part are also useful for detecting transitions between functional brain states in an online manner, for designing optimal model-based interventions and for statistically reconstructing functional architecture. In the past, latent variable models were successfully applied in neuroscientific data analysis tasks, namely dimensionality reduction and visualization (Cunningham and Yu, 2014), signal deconvolution (Vogelstein et al., 2010), denoising (Wu, Nagarajan and Chen, 2016) and decoding (Z. Chen, Gomperts, Yamamoto and Wilson, 2014), explorative data analysis (Latimer, Yates, Meister, Huk and Pillow, 2015), as well as the assessment of variability (Whiteway and Butts, 2017).

Recently, nonlinear and sparsity-based dimensionality reduction techniques such as dictionary learning and sparse low-rank matrix factorization approaches as well as their robust Bayesian versions and probabilistic extensions gained increasing popularity due to their flexibility in modelling the data and by incorporating additional problem-specific features, e.g., non-negativity or group sparsity (Cunningham and Yu, 2014; Ganguli and

---

Sompolinsky, 2012; Vogelstein et al., 2010).

Exemplary challenges in this domain along with computational speed and robustness constraints are:

- (i) Recovered latent state representations are unstable with respect to model mismatch, in particular to assumptions on the observation noise; this is especially pronounced for large noise amplitudes, e.g. in neurophysiological recordings.
- (ii) The dimension and structure of the recovered latent representation is sensitive to the presence of unobserved confounders that often lead to more latent dimensions caused by the introduced correlation from the confounders.
- (iii) New computationally efficient and convergent optimization algorithms for nonlinear low-rank matrix and tensor approximations for fast online implementations are yet in their infancies (Yang, Pesavento, Chatzinotas and Ottersten, 2018).

### 2.3.4 Approximation of Intractable Distributions

Because the integral of the marginal likelihood in Eq. 7 is intractable for higher dimensions, it is impossible to evaluate or differentiate the marginal likelihood (Kingma and Welling, 2013); instead, the posterior distribution  $p(z|x)$  can be approximated, either by

- (a) sampling-based approaches, or
- (b) variational inference.

In the past decades, the Markov Chain Monte Carlo (MCMC) paradigm became the cornerstone for sampling-based approaches. Here, an ergodic Markov chain on  $z$  is constructed, whose stationary distribution is  $p(z|x)$ . To collect samples of the stationary distribution, samples are instead taken from the chain. Subsequently, an empirical estimate either about the distribution itself or its punctual statistics is made from all or only subsets of the samples. This estimate is then taken as an approximation of the posterior (Hastings, 1970). Given infinite computational resources, this approach guarantees to generate exact samples for the target density, which makes this technique very appealing. However, MCMC algorithms tend to be computationally costly and particularly slow for large datasets or very complex models (Robert and Casella, 2004).

As an alternative to sampling-based approaches, variational inference rather relies on optimization. First, a family of approximate densities  $\mathcal{D}$  over the latent variables is posited.

---

It is then tried to find a member  $q^*$  of that family which minimizes the distance between the variational distribution  $q(z)$  and the exact posterior distribution  $p(z|x)$ , here measured via Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951), so that

$$q^*(z) = \operatorname{argmin}_{q(z) \in \mathcal{D}} \operatorname{KL}(q(z)||p(z|x)). \quad (9)$$

The closer  $q(z)$  is to  $p(z|x)$ , the smaller the KL divergence will be. Hence, the choice of distribution family for  $\mathcal{D}$  plays an important role, as it is required to be flexible enough for close approximation of  $p(z|x)$ , while at the same time being simple enough for efficient optimization (Blei, Kucukelbir and McAuliffe, 2017).

The KL divergence term in Eq. 9 is defined as

$$\operatorname{KL}(q(z)||p(z|x)) = \mathbb{E} [\log q(z)] - \mathbb{E} [\log p(z|x)] \quad (10)$$

$$= \mathbb{E} [\log q(z)] - \mathbb{E} [\log p(z, x)] + \log p(x) \quad (11)$$

, and is having several important properties, such as

- *Non-negativity*, as  $\operatorname{KL}(q(z)||p(z|x)) \geq 0$  for all  $q, p$
- *Asymmetry*, as  $\operatorname{KL}(q(z)||p(z|x)) \neq \operatorname{KL}(p(z|x)||q(z))$

(Blei et al., 2017).

From the expanded conditional in Eq. 10, the dependency on  $\log p(x)$  becomes obvious. This makes a direct computation impossible, because the marginal likelihood is intractable, as previously stated. Alternatively, the KL can be re-written to obtain a lower bound which can then be optimized instead. This is called the *evidence lower bound* (ELBO)  $\mathcal{F}(q)$ , also known as *variational lower bound* (Jordan, Ghahramani, Jaakkola and Saul, 1999).

The ELBO can be derived either (i) by using Jensen's inequality, or (ii) directly from the KL definition. For (i), a lower bound of the marginal  $\log p_\theta(x)$  following a family of distributions with unknown parameter  $\theta$  can be derived from the log-likelihood  $\mathcal{L}_i(\theta)$  with

datapoints  $i = 1, \dots, N$  as

$$\mathcal{L}_i(\theta) = \log \int p_\theta(x, z) dz \quad (12)$$

$$= \log \int \frac{p_\theta(x, z)}{q(z)} q(z) dz \quad (13)$$

$$= \log \mathbb{E}_{q(z)} \left[ \frac{p_\theta(x, z)}{q(z)} \right] \quad (14)$$

$$\geq \mathbb{E}_{q(z)} \left[ \log \frac{p_\theta(x, z)}{q(z)} \right] = \mathcal{F}(q) \quad (15)$$

where Jensen's inequality can be applied in Eq. 14 due to concavity of the logarithm (Fraccaro, 2018).

For (ii), the ELBO  $\mathcal{F}(q)$  can also be formulated as negative KL divergence plus an additional constant  $\log p(x)$  by using Bayes' rule in Eq. 10, so that

$$\text{KL}(q(z)||p(z|x)) = -\mathbb{E} \left[ \log \frac{p(x, z)}{q(z)} - \log p(x) \right] \quad (16)$$

$$= -\mathbb{E} \left[ \underbrace{\log \frac{p(x, z)}{q(z)}}_{\mathcal{F}(q)} \right] + \log p(x) \quad (17)$$

$$(18)$$

(Blei et al., 2017).

Note that the log-evidence  $\log p(x)$  can be re-formulated as combination of the ELBO and KL, so that

$$\log p(x) = \mathcal{F}(q) + \text{KL}(q(z)||p(z|x)) \quad (19)$$

It follows from the non-negativity property of the KL divergence, that minimizing the KL divergence is equivalent to maximizing the ELBO. Therefore it is possible to circumvent the intractable KL computation between the approximate and exact posteriors by maximizing the ELBO instead, which in turn is computationally tractable, e.g. by using gradient-based methods (Bishop, 2006).

---

## 2.3.5 Variational Autoencoder

### Basic Model Framework

The variational autoencoder (VAE) (Kingma and Welling, 2013) is a deep latent variable model for unsupervised and semi-supervised learning of meaningful latent representations  $z$  from a dataset  $x$ . Instead of learning a deterministic mapping as pursued in traditional autoencoders, the VAE instead models distributions of the latent variables. To give a more formal model description of the VAE, it is necessary to describe i) its *inference network* as variational approximation, ii) its *generative network* as latent variable model, and iii) how parameters are learned.

The posterior inference problem previously described in Sect. 2.3.4 can be approached by introducing a parametric inference model  $q_\phi(z|x)$ , also known as the *encoder* or *recognition model*. The model parameters  $\phi$  are optimized, so that the true posterior  $p_\theta(z|x)$  with corresponding parameters  $\theta$  is approximated by

$$q_\phi(z|x) \approx p_\theta(z|x). \quad (20)$$

Using a deep neural network NN,  $q_\phi(z|x)$  can be parametrized as  $\phi$  will include the network's weights and biases. By constructing this encoder model as neural network architecture as illustrated in Fig. 8 (left), it is possible to compute the parameters of the posterior approximation  $q_\phi(z|x)$  given the data point  $x$ , therefore learning a mapping from  $x$  to  $z$ . For example, in terms of a Gaussian this would result in

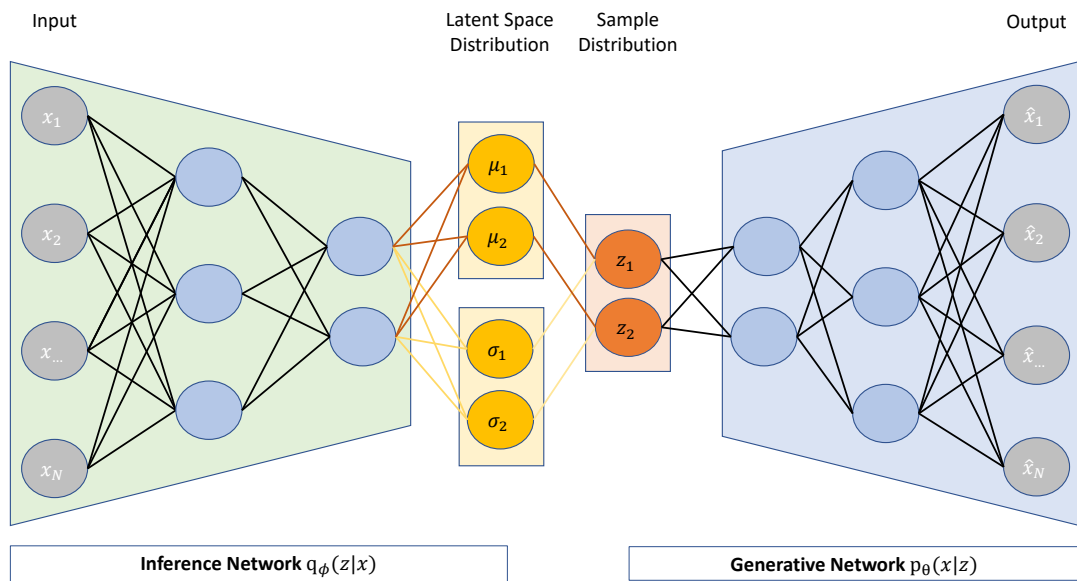
$$(\mu, \log \sigma^2) = \text{NN}_\phi(x) \quad (21)$$

$$q_\phi(z|x) = \mathcal{N}(z; \mu, \text{diag}(\sigma^2)) \quad (22)$$

(Kingma and Welling, 2013, 2019).

Instead of having different sets of parameters  $\phi_i$  to learn for each data point  $x_i$ , an alternative approach is known as *amortized* variational inference, where the variational parameters are shared across all observations (Gershman and Goodman, 2014). The term “amortized” indicates that the cost of learning  $\phi$  amortizes across all data points, which makes computations substantially efficient, as per-datapoint optimization loops can be circumvented. Also, when observing a new data point, it is possible to immediately compute its corresponding variational approximation without re-running an optimization step of the ELBO (Kingma and Welling, 2019). However, because the parameters of the inference network are shared across all data points, the posterior approximation produced with





**Figure 8:** General model architecture of the VAE. For given model input  $x$ , the inference network (or *encoder*) produces a mapping to the latent posterior approximation  $q_\phi(z|x)$ . The latent probability distribution is following a multivariate Normal distribution. For illustration purpose, latter is parametrized by a vector of two means  $[\mu_1, \mu_2]$  and covariance matrix  $\begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}$ . These parameters are the output of the encoder after updating the weights and biases of the network. After drawing samples  $z$  from the latent distributions, the generative network (or *decoder*)  $p_\theta(x|z)$  projects those latent representations back to the data distribution.

amortized inference will always be poorer than the one determined with conventional techniques. As discussed by Cremer, Li and Duvenaud (2018), the approximations introduced by amortized inference are the main cause of inference sub-optimality in VAEs, rather than the impact of restricting the distribution families of the variational approximation.

Aside of the inference network, the second important compartment of the VAE is the *generative network* specified by the joint distribution  $p_\theta(x, z) = p_\theta(x|z)p_\theta(z)$ . Its purpose is to learn a mapping from the latent space representation  $z$  back to the observations  $x$ , so that newly generated data conditioned on  $z$  is closely resembling the input data. This is achieved by the likelihood term  $p_\theta(x|z)$ , also known as *decoder*. Typically, in the

continuous case this has the form of a centered isotropic multivariate Gaussian, so that  $p_\theta(x|z) = \mathcal{N}(x; \mu, \sigma^2)$ . Corresponding parameters of the Gaussian are again parametrized by two deep neural networks  $\text{NN}_1$  and  $\text{NN}_2$  as indicated in Fig. 8 (right), so that

$$\mu = \text{NN}_1(z) \quad (23)$$

$$\log \sigma^2 = \text{NN}_2(z) \quad (24)$$

(Kingma and Welling, 2013, 2019).

In case of the VAE, the objective is to maximize the log-likelihood  $\log p_\theta(x)$ . As previously shown, the log-likelihood is lower-bounded by the ELBO term  $\mathcal{F}_{\theta, \phi}$  with

$$\mathcal{F}_{\theta, \phi} = \mathbb{E}_{q_\phi(z|x)} \left[ \log \frac{p_\theta(x, z)}{q_\phi(z|x)} \right] \quad (25)$$

$$= \underbrace{\mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x|z)]}_{\text{Reconstruction term}} - \underbrace{\text{KL} [q_\phi(z|x) || \log p_\theta(z)]}_{\text{Regularization term}} \quad (26)$$

The ELBO combines two terms with different purpose:

- a *reconstruction* term, minimizing the reconstruction error and in turn maximizing the marginal log likelihood  $\log p_\theta(x)$  for improving the generative model
- a *regularization* term, minimizing  $\text{KL}(q_\phi(z|x) || p_\theta(z|x))$  for encouraging the learned distribution  $q_\phi(z|x)$  to be similar to the true posterior distribution.

It is possible to differentiate the ELBO and jointly optimize it for both parameters  $\theta$  and  $\phi$ , e.g. via stochastic gradient descent. However, to compute the gradients it would be necessary to backpropagate from  $x$  and  $\phi$  through  $z$ . This becomes impossible, as  $z$  is a random variable and therefore a stochastic node in the graph. Here, a change of variables can be applied, commonly known as *reparametrization trick* (Kingma and Welling, 2013; Rezende, Mohamed and Wierstra, 2014). This step includes a differentiable and invertible transformation function  $g$  of  $z \sim q_\phi(z|x)$  with another random variable  $\epsilon$ , given  $z$  and  $\phi$ , so that

$$z = g(\epsilon, \phi, x) \quad (27)$$

where  $\epsilon$  is following a simple distribution  $p(\epsilon)$  and is independent of  $x$  and  $\phi$ . To illustrate this step, consider  $z \sim q_{\mu, \sigma} = \mathcal{N}(\mu, \sigma)$ . Instead of sampling from  $q$ , the reparametrization trick is applied by introducing  $\epsilon \sim \mathcal{N}(0, 1)$ , into which all the randomness in  $z$  is externalized, as

$$z = g_{\mu, \sigma}(\epsilon) = \mu + \sigma \odot \epsilon \quad (28)$$

with  $\odot$  denoting the Hadamard (or element-wise) product (Kingma and Welling, 2013; Rezende et al., 2014).

---

## Model Extensions

As described in Sect. 2.3.5, a trade-off between the two compartments of the VAE objective becomes evident: on the one hand, the reconstruction term  $\mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)]$  is promoting the latent code in  $z$  to provide essential information for reconstructing the observations in  $x$ ; on the other hand, the regularization term  $\text{KL}[q_\phi(z|x)||\log p_\theta(z)]$  penalizes the posterior approximation  $q_\phi(z|x)$  for straying too far from the prior  $p_\theta(z)$ . In practice, both compartments of the ELBO can raise difficulties in terms of an undesired local optimum.  $\theta, \phi$  are randomly initialized at the start of VAE training, leading to poor latent codes in  $z$ . The posterior approximation is then substantially pushed by the regularization term towards the prior, so that  $q_\phi(z|x) \approx p_\theta(z)$  for all  $x$ . Consequently,  $z$  remains uninformative over the training. This phenomenon is denoted as posterior (or KL) collapse (Bowman et al., 2016; Fu et al., 2019; Long, Cao and Cheung, 2019; Lucas, Tucker, Grosse and Norouzi, 2019). Simultaneously, to ensure interpretability of latent representations, it has to be aimed for disentanglement of inferred latent variables  $q_\phi(z|x)$ . When a single latent variable is sensitive to changes in a single generative factor of variation in the data but generally invariant to changes in other variables, this representation is said to be disentangled (Bengio et al., 2013).

Balancing the ELBO's reconstruction and regularization terms has been shown to be beneficial for the disentanglement of latent representations (Bowman et al., 2016; Fu et al., 2019; Higgins et al., 2016; Higgins et al., 2017). For this purpose, Higgins et al. (2017) proposed  $\beta$ -VAE by augmenting the original VAE objective. Here, the ELBO is extended by  $\beta$  as rescaling weight of the regularization term, so that

$$\mathcal{F}_{\theta,\phi} = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] - \beta \text{KL}[q_\phi(z|x)||\log p_\theta(z)]. \quad (29)$$

As learning constraints applied to the model are modulated by this weight, the capacity of the latent information channel is limited and statistically independent latent factors are emphasized by encouraging their factorization (Higgins et al., 2017). In context of  $\beta$ -VAE, the standard VAE formulation by Kingma and Ba (2014) is corresponding to  $\beta = 1$ . Choosing  $\beta < 1$  results in more accurate reconstructions but with the downside of a less regularized and more entangled latent space. Values of  $\beta > 1$  in turn can result in the aforementioned posterior collapse, in which all reconstructions are reduced to the average input and  $\text{KL}[q_\phi(z|x)||\log p_\theta(z)] \rightarrow 0$  (Rydhmer and Selvan, 2021).

To alleviate the problem of posterior collapse, Bowman et al. (2016) suggested a simple annealing schedule for the regularization weight  $\beta$  to monotonically increase over the VAE training (usually from  $\beta = 0$  to  $\beta = 1$ ). Despite being widely adapted, especially in the

field of natural language processing, monotonical annealing comes with the tendency to under-weight the prior regularization. In this case, the VAE substantially degrades into a standard autoencoder with point estimates of the learned  $q_\phi(z|x)$  and poor decoder learning (Fu et al., 2019).

An alternative approach proposed by Fu et al. (2019) concerns a cyclical annealing schedule, starting with  $\beta = 0$ , quickly increasing and keeping it at certain level (originally at  $\beta = 1$ ) for several training iterations. This process is subsequently repeated for several cycles, so that

$$\beta_t = \begin{cases} f(\tau), & \tau \leq R \\ 1, & \tau > R \end{cases} \text{ with} \quad (30)$$

$$\tau = \frac{\text{mod}(t - 1, [T/M])}{T/M} \quad (31)$$

with  $t = 1, \dots, T$  training iterations, a monotonically increasing function  $f$  (e.g. linear, sigmoid, or cosine), the number of cycles  $M$ , and  $R$  denoting the increase proportion of  $\beta$  within a cycle. Within the first cycle, the model is supported to converge to the ELBO and infer its first raw full latent distribution. When restarting the annealing procedure in the consecutive cycle, the ELBO is perturbed and pushed away from previous convergence. Training is continued on basis of the full distribution  $z \sim q_\phi(z|x)$  learned in the previous cycle as a warm restart. The annealing process is then repeated multiple times to achieve better convergence. For more disentangled representations, the authors advise to set the upper limit of the regularization weight to  $\beta > 1$  for putting stronger capacity constraints on  $z$  as in the standard VAE (Fu et al., 2019).

## 2.4 Synthetic Data Generation

### 2.4.1 Motivation

For evaluating the implemented deep latent variable model (Sect. 2.3.5) in terms of its ability to extract response- and/or noise-related features from VSDI sequences in a fast and robust way, it is necessary to provide a suitable data basis. However, ground truth about the contribution of signal components as well as their composition is unknown for real datasets.

In order to circumvent this issue, artificial image sequences are generated using methods from geostatistics and signal processing, which are described in the following sections. The

---

goal is set to create spatio-temporal dynamics with similar complexity and dimensionality as in VSDI data obtained from grating experiments, while having complete knowledge about the signal composition and data-generating process.

First, an artificial orientation preference map is created (Sect. 2.4.2). From corresponding activated regions, spatial locations of hypothetical orientation columns are extracted. These locations are then employed in the generation of conditioned random fields (CRF) (Sect. 2.4.3 & 2.4.4) in terms of spatial conditioning sites. Additionally, temporal dynamics at these locations are conditioned on regime-switching timings of a typical grating stimulation paradigm, here defined by a switch between baseline and stimulation phase within a single record. Both the spatial and temporal dynamics of generated CRF can be precisely controlled, as calculations are built on the idea of kriging (also known as Gaussian process regression), which is a spatial interpolation method originally developed in geostatistics (Krige, 1951). By this, random fluctuations in proximity of the specified spatial and temporal conditioning points can be accounted for by the choice of an appropriate covariance model, e.g. Matérn. This allows for approximating spatiotemporal phenomena commonly observed in cat's primary visual cortex such as trial-to-trial variability of evoked response patterns (Carandini, 2004; Heggelund and Albus, 1978) as well as spontaneous (or ongoing) activity (Arieli et al., 1995; Kenet et al., 2003).

To further incorporate knowledge about the composition of the VSDI signal, temporal noise components (Sect. 2.4.5) related to technical and physiological sources are modeled stochastically by following routines of Reynaud et al. (2011). Additionally, the generated sequences are convolved with artificial spatial image artifacts (Sect. 2.4.6) related to illumination and blood vessel structures. Ultimately, all synthetic signal components are included in a weighted linear model of the overall signal (Sect. 2.4.8).

## 2.4.2 Synthetic Orientation Preference Maps

Functional orientation preference maps represent a traditional visualization technique of the location and extent of neural populations responding to certain properties of orientation stimuli, e.g. a certain orientation degree. These maps are usually built by temporally vector-averaging several trials of optical imaging (Grinvald, Frostig, Siegel and Bartfeld, 1991; Hubel and Wiesel, 1968).

To simulate optical imaging data based on a hypothetical orientation stimulus paradigm, the procedure of generating orientation preference maps is basically reversed. First, an artificial orientation preference map is built, storing the pixel-wise preference to a certain

---

orientation. This map then shows pixel patches and clusters, indicating neural populations with this certain response preference. Informations about placement and extent of these patches can then be used as prior spatial information for subsequently generating spatio-temporal neural response dynamics. This procedure offers a theoretically plausible approximation of the number and spatial positioning of orientation columns within the 2-D imaging area.

To generate synthetic orientation preference maps, an approach by Macke, Gerwin, White, Kaschube and Bethge (2009, 2011) is followed. Two instances of white noise are convolved with Mexican hat filters, which in turn are generated using Difference-of-Gaussians. The resulting matrices are then used as real and imaginary image parts of a synthetic orientation preference map (Fig. 9a). Hereby, common properties of smoothness and semi-periodic structure of real orientation preference maps are taken into account. For every specified orientation, an activation map (Fig. 9b) can then be returned via functions of the corresponding *gp-maps-python* repository in Python. Each activation map stores information about each contour in the image domain, in turn representing hypothetical orientation columns (Macke et al., 2011).

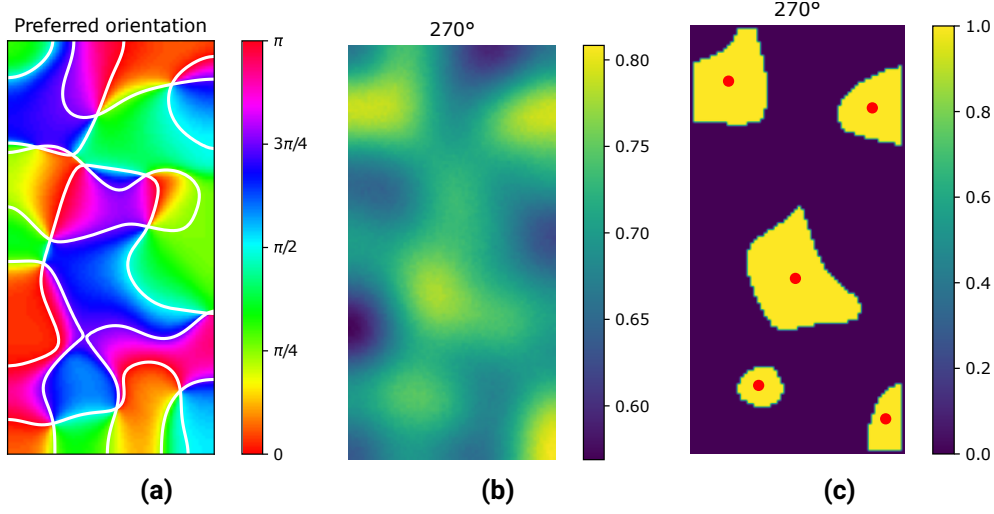
Building upon this approach, the centerpoint of each contour (Fig. 9c) is estimated by calculating its center of mass in the following steps:

1. Convert the current activation map to a grayscale image.
2. Binarize the image.
3. Calculate the image moments via Green’s theorem (Cauchy, 1846).
4. Find the center of mass  $C_x, C_y$  by using first-moment ratios.

To further elaborate on this, image moments are holding information about basic statistical properties of each contour, e.g. its area, centroid and orientation. More precise, image moments are the weighted average of pixel intensities  $I(x, y)$  of a contour within its boundary  $i = 1, \dots, n$ . For that purpose, functions of the computer vision toolbox *OpenCV* (Bradski and Kaehler, 2008) in Python are used to calculate the spatial moments  $M_{p,q}$ :

$$M_{p,q} = \sum_i^n (I(x, y)x^p y^q). \quad (32)$$

with  $p$  representing the  $x$ -order and  $q$  the  $y$ -order. The term “order” is here defined as the power to which the corresponding component is taken (Bradski and Kaehler, 2008).



**Figure 9:** (a) Synthetic orientation preference map. Here, the hypothetical orientation responses to four different orientations (in rad) are color-coded. (b) Activation map for a single orientation, for which selective pixels in (a) were color-coded in blue. Values are normalized to range  $[0, 1]$ . (c) Binarized contours of corresponding activation map. Respective contour centroids (indicated as red dots) are taken as prior information for the location of orientation columns. These locations can then be used as conditional positions in the subsequent generation of random fields (see 2.4.4).

As the centroid  $C$  of a contour containing a set of  $k$  contour points in  $\mathbb{R}^k$  is minimizing the sum of squared Euclidean distances between itself and each point in the set, the x-y-coordinates of the contour's center of mass  $C_x, C_y$  can be calculated as:

$$C_x = \frac{M_{10}}{M_{00}} \quad (33)$$

$$C_y = \frac{M_{01}}{M_{00}}. \quad (34)$$

For detailed proofs and derivations of image moments and how to find the centroid of a continuous distribution of mass, please see Swokowski (1979) and Simmons (1996).

The extracted centerpoints are used as prior information for the subsequent genera-

tion of conditioned random fields described in Sect. 2.4.4 in terms of restricting the 3-D kriging operation through the centerpoint positions.

### 2.4.3 Spatial Random Fields

For generating artificial image sequence with complex spatio-temporal fluctuations, spatial random fields (SRF) were generated using the randomization method described by Heße, Prykhodko, Schlüter and Attinger (2014) and was implemented in Python via functions of the GeoStatTools package (S. Müller and Schüler, 2020). The SRF is represented by a stochastic Fourier integral and its discretized modes are evaluated at random frequencies. For this application, a 3-D field was generated with a Matérn covariance model given by the correlation function

$$\rho(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \cdot \left(\sqrt{\nu} \cdot \frac{r}{\ell}\right)^\nu \cdot K_\nu\left(\sqrt{\nu} \cdot \frac{r}{\ell}\right) \quad (35)$$

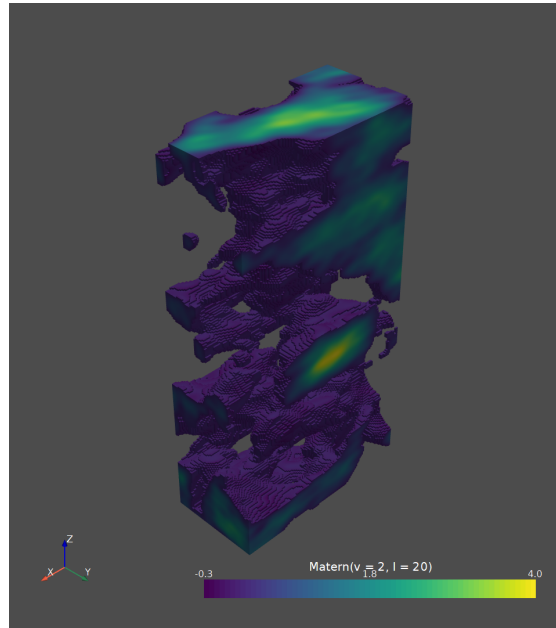
(Rasmussen and Williams, 2006), where  $\Gamma$  is the Gamma function,  $K_\nu$  denotes the modified Bessel function of the second kind, and  $\ell$  represents the length scale.  $\nu$  is a shape parameter and should be chosen so that  $\nu \geq 0.2$ . In case that  $\nu > 20$ , a Gaussian model is used, since it is the limit case:

$$\rho(r) = \exp\left(-\frac{1}{4} \cdot \left(\frac{r}{\ell}\right)^2\right) \quad (36)$$

To emulate spatial patterns with size proportions close to columnar response signals from V1, the centroids of spatial activity have to reach diameters within a plausible range for a cortical column of approx. 300 – 600  $\mu\text{m}$  (Mountcastle, 1997). In terms of the VSDI camera setup, this corresponds to around 19 – 38 pixel while considering a single pixel covering eight  $\mu\text{m}^2$  of recorded cortical surface area with a pixel binning factor of two. Similar size proportions of the spatial activity patterns were achieved by feasible parametrization of the Matérn correlation function, namely the shape parameter  $\nu = 2$  and length scale  $\ell = 20$ . A realization of an equally specified SRF is illustrated in Fig. 10.

Each generated 3-D SRF is then traversed through its z-axis – which is here determined for representing the temporal dimension – in a given step size. By this approach, spatio-temporal activity patterns following temporal dependency structures similar to an autoregressive process of order  $p$  are emulated, where  $p$  can be interpreted as the step size of the traversal. In this case,  $p$  was defined as  $p = 1$  to match a frame-wise succession.





**Figure 10:** Realization of a 3-D spatial random field following a Matérn covariance model. This field of shape  $(x = 128, y = 64, z = 255)$  was generated by specifying shape parameter  $\nu = 2$  and length scale  $\ell = 20$ . Exclusively for this illustration, values were thresholded to range  $[-0.3, 4.0]$  for highlighting structural dependencies within the field.

#### 2.4.4 Conditioned Random Fields

When considering cortical activity induced by a similar stimulation paradigm used for recording optical imaging (Sect. 2.1.2), synthetically generated image sequences have to incorporate several assumptions of the spatio-temporal response dynamics of orientation columns. First, the center of an activated region representing an orientation column has to be spatially fixed at a certain location. Here, centroid coordinates which were previously extracted from synthetic orientation preference maps (Sect. 2.4.2) are used to ensure a plausible distributions of artificial columns across the image area. Secondly, corresponding dynamics have to reflect the response behaviour of cortical response to a drifting grating stimulus over time, which is assumed to be expressed by regime switching between baseline and stimulation phases. For this kind of paradigm using a single stimulation phase, response shapes recorded with VSDI are typically characterized by rising, plateau and

---

decaying phases (Reynaud et al., 2011).

The aforementioned SRFs (Sect. 2.4.3) are characterized by complex dependency structures in 3-D with plausible spatial size ratios. To generate random fields which are bounded at pre-defined spatial and temporal points, but still contain random variability according to a specified covariance model, a conditioned random field (CRF) can be used alternatively. Here, field realizations are achieved by combining the random field generation with kriging. Respective CRF generation routines are also available in the *GSTools* package in Python. The following descriptions are accordingly taken from the corresponding documentation provided by S. Müller and Schöler (2020).

A CRF is generated in the following steps:

1. a field is generated by a specified kriging approach
2. a random field with mean of 0 and variance of 1 is generated
3. the random field is multiplied with the kriging standard deviation.

Based on a given covariance model, a value on a field  $z$  at some point  $x_0$  can be derived by fixed conditioning values  $z(x_1), \dots, z(x_n)$  at target location points  $x_i$ . The value  $z_0$  is calculated as the weighted mean

$$z_0 = \sum_{i=1}^n w_i \cdot z_i \quad (37)$$

where the weights  $w_i$  are depending on the pre-defined covariance model as well as the target location. The covariance model is used to characterize the following semi-variogram  $\gamma$  of the random field:

$$\gamma(r) = \sigma^2 \cdot (1 - \rho(r)) + n \quad (38)$$

with the lag distance  $r$ , main correlation length  $\ell$ , variance  $\sigma^2$  and nugget (or subscale variance)  $n$ . The correlation function  $\rho(r)$  is given by

$$\rho(r) = \text{cor}\left(s \cdot \frac{r}{\ell}\right) \quad (39)$$

with re-scaling factor  $s$ . This is resulting in the covariance function

$$C(r) = \sigma^2 \cdot \rho(r) \quad (40)$$

---

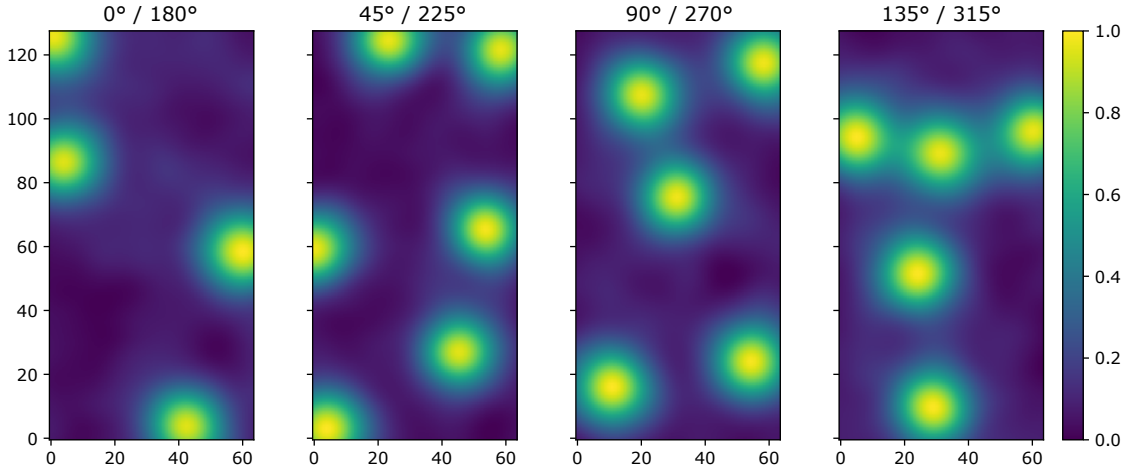
(S. Müller and Schüler, 2020).

In some cases, the mean of the random field is known or can be assumed, which can improve the estimates by “simple kriging” (Webster and Oliver, 2007). In terms of the corresponding functions of the GSTools package, the mean is assumed to be zero for simplicity. The resulting equation system for weight matrix  $W = (w_1, \dots, w_n)$  then becomes

$$W = \begin{pmatrix} c(x_1, x_1) & \dots & c(x_1, x_n) \\ \vdots & \ddots & \vdots \\ c(x_n, x_1) & \dots & c(x_n, x_n) \end{pmatrix}^{-1} \begin{pmatrix} c(x_1, x_0) \\ \vdots \\ c(x_n, x_0) \end{pmatrix} \quad (41)$$

(S. Müller and Schüler, 2020).

Accordingly, field dynamics are restricted via phase-specific key values at given key positions (Fig. 11) on the field’s axes. The fields  $xy$ -plane is interpreted as spatial dimension representing the 2-D image domain, and the  $z$ -axis as temporal dimension separating frames of the 3-D image sequence. A key value of zero is set for all corresponding baseline frames, which is incorporating the assumption of more unstructured spatio-temporal fluctuations during the absence of a stimulus. Starting with the frame index of the hypothetical stimulus onset, an increase in key values is introduced up until reaching a maximum value. By this, a reaction phase of columnary response to the stimulation onset is emulated. The maximum key value is held for several frames, reflecting the assumption of a relatively stable plateau phase of response. Finally, an adaptation phase to the stimulus is considered by decreasing the key values in several steps. As CRFs will pose the data basis for the subsequent model training and evaluation, corresponding field realizations will be illustrated in greater detail in Sect. 3.2.1.



**Figure 11:** CRF key locations. Several centroids are defined for the kriging operation in CRF generation. Each conditioned field is then interpreted as spatio-temporal columnary response in V1 to a hypothetical orientation stimulus.

## 2.4.5 Temporal Noise Components

### Dye Bleaching

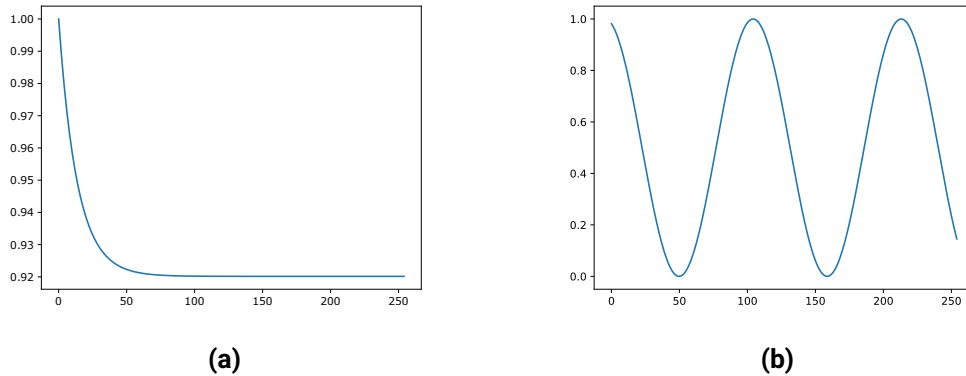
A photo-dynamic bleaching phenomenon as previously described in Sect. 2.1.1 is commonly reported as prominent component of VSDI as well as  $\text{Ca}^{2+}$  imaging (Bathellier et al., 2007; Chemla et al., 2017; Grinvald et al., 1982; Reynaud et al., 2011; Stetter, Greve, Galizia and Obermayer, 2001).

Accordingly, a synthetic dye bleaching artefact is incorporated in the generation of artificial VSDI sequences in terms of a global spatial signal component, which means that this component is affecting the time course of every available pixel identically. By this, a simplified assumption of an even dye staining procedure of cortical surface is resembled.

For generating a synthetic bleaching curve  $b$  in the domain of a single pixel at coordinate  $(x, y)$ , a combination of a double negative exponential and a linear term parameterized by  $\tau_i$  is used, so that

$$b(t, x, y) = \tau_0 + \tau_1(e^{-\frac{t}{\tau_2}} - 1) + \tau_3(e^{-\frac{t}{\tau_4}} - 1) + \tau_5 t \quad (42)$$

where the decay spans over a single sequence with  $T = 255$  frames indexed by  $t = 1, \dots, T$ . This approach is in accordance with Hofmann (2020) and accounts for variability in



**Figure 12:** (a) Realization of artificial dye bleaching dynamics on pixel domain. Generated by combining a double negative exponential and a linear term. (b) Realization of artificial heartbeat dynamics. For both components, the simplified assumption was made that every pixel is identically impacted. Curve specifications are made in accordance with Hofmann (2020).

bleaching behavior through different acceleration stages, especially from early light exposure, depending on the quality of the staining procedure as concentration of actual bound dye molecules in the lipid layer, which cannot be directly controlled.

As reported by Chemla et al. (2017), a substantial trial-to-trial variability of bleaching dynamics is recognizable. To take this variability into account,  $\tau_i$  is re-drawn per generated sequence from  $\tau_i \sim N(\mu_{\tau_i}, \sigma_{\tau_i})$ . Here,  $\mu_{\tau_i}$  was set to the median and  $\sigma_{\tau_i}$  to the half interquartile range of parameter distributions estimated for 240 real VSDI recordings (ID of experimental subject: 092413) to prevent strong outlier parameter settings. An exemplary realization for artificial bleaching pixel dynamics is illustrated in Fig. 12a.

### Heartbeat

When inspecting pixel-wise dynamics of VSDI data, slow-wave oscillations can be recognized which are related to the experimental subject's heartbeat (Inagaki, 2003; Shoham et al., 1999). Due to periodically changes in blood volume within the cortex (Fukuda et al., 2005), the camera's focus level is changed and therefore the total signal intensity varies (Hofmann, 2020).

---

In accordance with other works by Reynaud et al. (2011) and Chemla et al. (2017), the heartbeat dynamics of a single pixel at coordinate  $(x,y)$  is modeled as combined sine-cosine function  $h$  which is phase-shifted by  $\phi_1$  and  $\phi_2$ :

$$h(t, x, y) = \cos(2\pi \cdot f \cdot t + \phi_1) + \sin(2\pi \cdot f \cdot t + \phi_2) \quad (43)$$

To estimate the heartbeat frequency (in Hz)  $f$  within a single recording, an approach by Hofmann (2020) was followed. As for the dye bleaching component, it is assumed that the heartbeat is having an uniform impact on every pixel of the recorded image area. Accordingly, along the baseline duration of every sequence, the spatial average of pixel time courses is computed and its power spectral density is measured via fast Fourier transform (FFT). Subsequently, the respective frequency with the highest amplitude in the power spectrum is selected within a physiologically plausible interval around the mean electrocardiogram frequency. This frequency is taken as estimate of the real heartbeat frequency for the current recording (Hofmann, 2020).

This approach was repeated over 240 recordings of a real VSDI experiment (ID of experimental subject: 092413), resulting in a distribution of 240 estimated heartbeat frequencies. Again, to ensure a sufficient level of trial-to-trial variability while omitting strong outlier parameter settings for the generation of synthetical heartbeat dynamics, a random draw from  $f \sim N(\mu_f, \sigma_f)$  is taken for every new generated sequence. The median value of the distribution of 240 frequencies is plugged in for  $\mu_f$  and respective half interquartile range is used as  $\sigma_f$ .

To illustrate an exemplary realization for the synthetic heartbeat generation process, please see Fig. 12b.

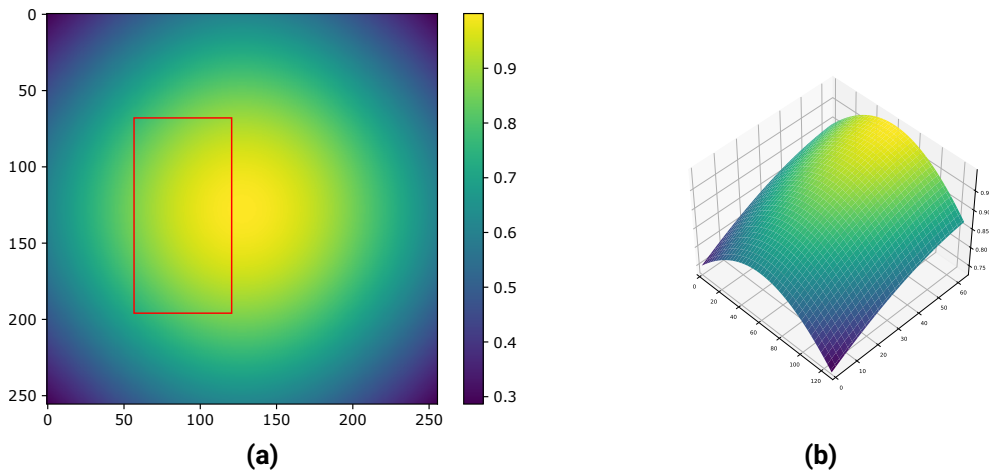
## 2.4.6 Spatial Noise Components

### Illumination

When inspecting real VSDI imaging data, a spatial decrease of signal amplitude towards the image's borders is clearly recognizable in every recorded frame. This is due to the illumination of the convex cortical surface through the recording chamber. To incorporate a comparable spatial surface structure in the synthetic data generation process, a 2-D Gaussian kernel  $g$  is first generated as

$$g(x, y) = A \cdot e^{(a(x-x_0)^2 + 2b(x-x_0)(y-y_0) + c(y-y_0)^2)} \quad (44)$$

with amplitude  $A$ , weighting parameters  $a, b, c$ , and the x-y-coordinates of the kernel's



**Figure 13:** (a) 2-D Gaussian kernel. A rectangular ROI (red) is set according to the target image dimensions (here:  $x = 128$ ,  $y = 64$ ). (b) Surface plot of the 2-D Gaussian kernel cropped to corresponding rectangular ROI.

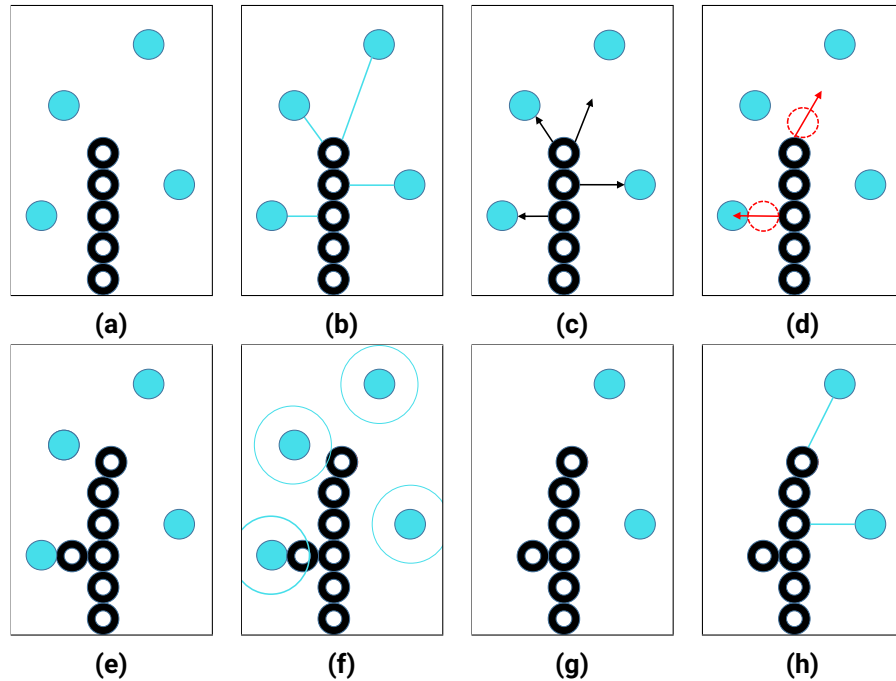
center point  $(x_0, y_0)$ .

The 2-D kernel (Fig. 13a) is then cropped to a rectangular ROI (Fig. 13b) of an equal shape as the target image dimensions. The location of the ROI was kept fixed across every frame and sequence, expressing the simplified assumption of stable and reliable camera positioning over time.

### Blood Vessel Networks

When investigating real sequences of VSDI, branching structures quickly become apparent as image features, which are corresponding to networks of cortical blood vessels. To emulate comparable branching patterns synthetically, the space-colonization algorithm developed by Runions et al. (2005) and further extended by Runions, Lane and Prusinkiewicz (2007) was used.

First, a 2-D envelope has to be defined, which will be colonized by the branching network. In this application for simulating blood vessel structures observed in VSDI data, the 2-D image space marks out this envelope. The image is then seeded by a set of attractor points signaling available empty space. The attractor points are randomly sampled from a discrete 2-D uniform distribution to ensure an even distribution across the full image

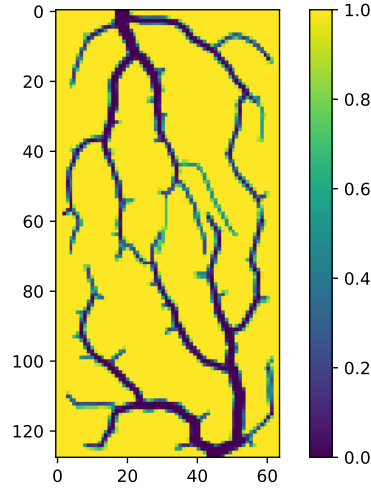


**Figure 14:** Intuition of the space-colonization algorithm. Modified from Runions, Lane and Prusinkiewicz (2007).

space (Fig. 14a). The skeleton of the branching network is grown iteratively, starting at the tree base (Fig. 14b). Nodes are extending their branch segment towards the average direction of all remaining attractors (Fig. 14c). The position of a new node is calculated by normalizing this average direction to a unit vector, then scaling it by a pre-defined segment length (Fig. 14d & 14e). Attractors get pruned when branches are getting too close (Fig. 14f & 14g). The growing process is terminated either when all attractor points were met, no nodes are within the radius of remaining attractor points, or an upper limit of iterations is reached (Fig. 14h). Key parameters to control the branching network generation are the attractor distance, the radius for attractor removal around nodes, and the distance between nodes (Runions et al., 2007).

For the generation of artificial VSDI sequences, an identical blood vessel mask (illustrated in Fig. 15) is shared between all sequences and frames, expressing the simplified assumption of a stable cortex positioning over time.





**Figure 15:** Synthetic blood vessel component. Generated via applications of the space-colonization algorithm by Runions, Lane and Prusinkiewicz (2007). Here, two networks were grown towards the images' center, starting from diagonally opposite tree base points located on the rectangular envelope.

### 2.4.7 Random Noise

During data acquisition, optical imaging exhibits stochastic signal components due to a plethora of different technical reasons, e.g. camera noise or fluctuations in illumination (Grinvald et al., 1999). To incorporate such randomness in the synthetic image sequences, a residual component is generated under the simplifying assumption of being independently and uniformly distributed Gaussian noise along both spatial and temporal dimensions. Accordingly, within frame  $t = 1, \dots, T$ , each pixel value with coordinates  $(x, y)$  is drawn from a standard Gaussian distribution, so that

$$E(t, x, y) \stackrel{iid}{\sim} N(0, 1). \quad (45)$$

### 2.4.8 Composition of Artificial VSDI Sequences

To generate a single artificial VSDI greyscale image sequence  $\hat{S}$  of target shape ( $T = 255$  frames, image height  $X = 128$  pixel, image width  $Y = 64$  pixel), a modification of the

$$\hat{S} = S \odot V + \beta_1 B + \beta_2 H + \beta_3 E + \beta_4 L$$

**Figure 16:** Composition of artificial VSDI greyscale image sequences. A 3-D raw signal  $S$  specified as CRF simulates a single spatio-temporal response pattern from orientation columns in V1 to a grating stimulus. Technical and biological noise components typically observed for VSDI are synthetically generated and applied to  $S$  in terms of a weighted linear combination.

VSDI signal composition model (Sect. 2.2.2) formulated by Reynaud et al. (2011) is used. Here, a weighted linear combination of all synthetic VSDI components is specified. The major difference concerns the inclusion of (artificial) neural response dynamics into the model. Cortical response profiles, which were included solely on temporal dimension in the original model formulation, are now replaced by a fundamental spatio-temporal signal  $S$ , which is a 3-D CRF containing one of the pre-defined artificial response patterns. Latter simulates ground truth information about orientation columns in V1 responding to a grating stimulus. All VSDI-typical noise components are further applied on this signal. As the original model formulation by Reynaud et al. (2011) was layed out for signal decomposition on single-pixel domain, only temporal dynamics covering evoked response profiles as well as technical and biological noise such as dye bleaching  $B$  and oscillatory components like heartbeat  $H$  were considered. By explicitly accounting for confounding spatial structures in the image domain, namely artificial blood vessel networks  $V$  and illumination  $L$ , the model is further extended to

$$\hat{S} = S \odot V + \beta_1 B + \beta_2 H + \beta_3 E + \beta_4 L. \quad (46)$$

As temporal noise components are assumed to have constant spatial dynamics along the image dimensions, they are repeated for every pixel at coordinate  $(x,y)$ . Likewise, spatial structures generated as 2-D arrays are repeated for every available frame  $t = 1, \dots, T$  and are also shared between sequences, as these components are assumed to remain constant over time. Accordingly, every model regressor forms a 3-D matrix which is weighted by its corresponding  $\beta$ . As stated by Reynaud et al. (2011), the usage of a linear model expresses the assumption that all model terms add up linearly. In this context, the only exception is made for the blood vessel component  $V$  which is introduced as multiplicative term as it essentially constitutes a binary image mask. The overall linear composition for generating artificial VSDI sequences is illustrated in Fig. 16.

---

## 3 Results

---

### 3.1 Approaches for Model Evaluation

After the VAE model training has been completed, the performance of the encoder and decoder compartments of the VAE model are assessed by carrying out several evaluation approaches. For that purpose, an arbitrary grayscale image sequence of shape  $[T, M, N]$  with  $t = 1, \dots, T$  frames as well as image height  $M$  and width  $N$  (both in pixel) is chosen as evaluation sequence from the respective validation data partition.

#### Encoder Evaluation

The variational approximate posterior is defined as multivariate Gaussian with diagonal covariance structure, so that

$$\log q_\phi(z|x^{(i)}) = \log \mathcal{N}(z; \mu^{(i)}, \sigma^{2(i)} I). \quad (47)$$

with identity matrix  $I$ . Here, sets of mean and standard deviation parameters  $\mu^{(i)}, \log \sigma^{2(i)}$  represent the output of the encoding neural network for datapoint  $x^{(i)}$  (Kingma and Welling, 2013).

The encoder draws a sample  $z^{(i,\ell)} \sim q_\phi(z|x^{(i)})$  following the approximate posterior  $q_\phi(z|x^{(i)})$  for datapoint  $x^{(i)}$  and latent variable  $\ell = 1, \dots, L$  by applying the reparametrization trick previously described in Sect. 2.3.5, so that

$$z^{(i,\ell)} = g_\phi(x^{(i)}, e^{(\ell)}) = \mu^{(i)} + \sigma^{(i)} \odot \epsilon^{(\ell)}. \quad (48)$$

with  $\epsilon^{(\ell)} \sim \mathcal{N}(0, I)$  and both the prior  $p_\theta(z)$  as well as approximate posterior  $q_\phi(z|x)$  being Gaussian (Kingma and Welling, 2013).

The chosen evaluation image sequence is passed frame-wise to the encoder compartment of the VAE. After all frames have been processed, the latent encodings of shape  $[T, L]$  are comprising a sample for each of the  $L$  latent variables over  $T$  input frames. Each latent dimension can then be inspected separately on two different levels:

- 
- (a) on *sequence*-level via kernel density estimation (KDE) of each latent variable’s samples distribution over all  $T$  frames of the input image sequence, visualizing deviations from the prior distribution specified as  $p_\theta(z) \sim \mathcal{N}(0, 1)$ , in turn indicating the regularization capability of the KL term in the VAE objective;
  - (b) on *frame*-level by plotting  $z^{(i,\ell)}$  against frame index  $t = 1, \dots, T$ , deviations from the prior mean of zero are visualized, by which frames containing relevant features or anomalies can be discovered, e.g. regime-switching behaviour of cortical response dynamics between baseline patterns and stimulus-related patterns.

Please note that no temporal structure can be assumed for  $z^{(i,\ell)} \sim q_\phi(z|x^{(i)})$ . As the VAE model specification at hand does not take temporal dependencies between observations into account, samples are drawn independently for every frame  $t = 1, \dots, T$ . Thus, plots obtained from (b) should only be interpreted as pseudo time series.

### Decoder Evaluation

For examining the capabilities of the VAE model to reconstruct input images from the evaluation sequence, samples in  $z$  are subsequently plugged into the decoder network constituting the generative distribution  $\log p_\theta(x^{(i)}|z^{(i,\ell)})$ . Latter enables the decoder to reconstruct the current image by relying entirely on  $z^{(i,\ell)}$ .

The general image reconstruction quality on single-frame level is operationalized via mean squared error (MSE) between the VAE input and output image. The MSE constitutes a common measure in image processing for assessing the similarity of two images (Wang, Bovik, Sheikh and Simoncelli, 2004). Formally, it is defined as the average squared differences in pixel intensities  $J_1$  and  $J_2$  of two images, so that

$$\text{MSE} = \frac{\sum_{M,N} [J_1(m,n) - J_2(m,n)]^2}{M * N} \quad (49)$$

with row index  $m = 1, \dots, M$  and column index  $n = 1, \dots, N$ .

Furthermore, the impact of each latent dimension on the image reconstructions is assessed in terms of a latent space walking procedure modified from Schau et al. (2019). Here, all but a single latent variable are kept constant while the variable of interest is swept through values drawn from the inverse cumulative distribution (CDF) of the standard Normal distribution, as latter also represents the prior distribution  $p_\theta(z) \sim \mathcal{N}(0, 1)$ . Drawn values from the inverse CDF are then plugged in for the variable of interest. By this approach, a synthetic latent space is generated for every draw. When passing these manipulated

---

encodings into the decoder network, a manifold of images can be generated by which the impact of the target latent dimension on the image reconstruction can be inspected (Schau et al., 2019). In this context, Schau et al. (2019) are keeping all remaining variables fixed at their expected mean value of zero, as  $p_{\theta}(z) \sim N(0, 1)$ . However, this approach does not allow for feasible interpretations when distributions of latent encodings show substantial deviations from their standard Normal prior. This problem can be alleviated when using a more robust point measure instead of the arithmetic mean. As the median is more robust against outliers and skewed distributions, it offers better interpretability of changes in reconstructions and will therefore be used in the following evaluations of latent space walks.

## 3.2 Parameter Studies

### 3.2.1 Overview

When specifying the model architecture of the VAE, its capability to extract features of the underlying activity patterns is essentially determined by two parameters in particular:

1. the *model capacity*, here defined as the number of latent variables  $\ell$  in the bottleneck  $z$ -layer of the VAE,
2. the level of *model regularization*, which in context of the  $\beta$ -VAE approach is specified as the KL weight  $\beta$  within the loss function.

To assess the impact of both parameters on the VAE modeling results, a 2-D grid is defined by the cartesian product of two vectors  $L = [1, 2, 3, 4, 5, 10, 25, 50, 75, 100]$  and  $B = [1, 2, 3, 4, 5, 10, 25, 50, 75, 100]$ , resulting in a total of 100 unique configurations of both  $\ell$  and  $\beta$ . A separate VAE model is trained for each configuration while holding all remaining model parameters constant.

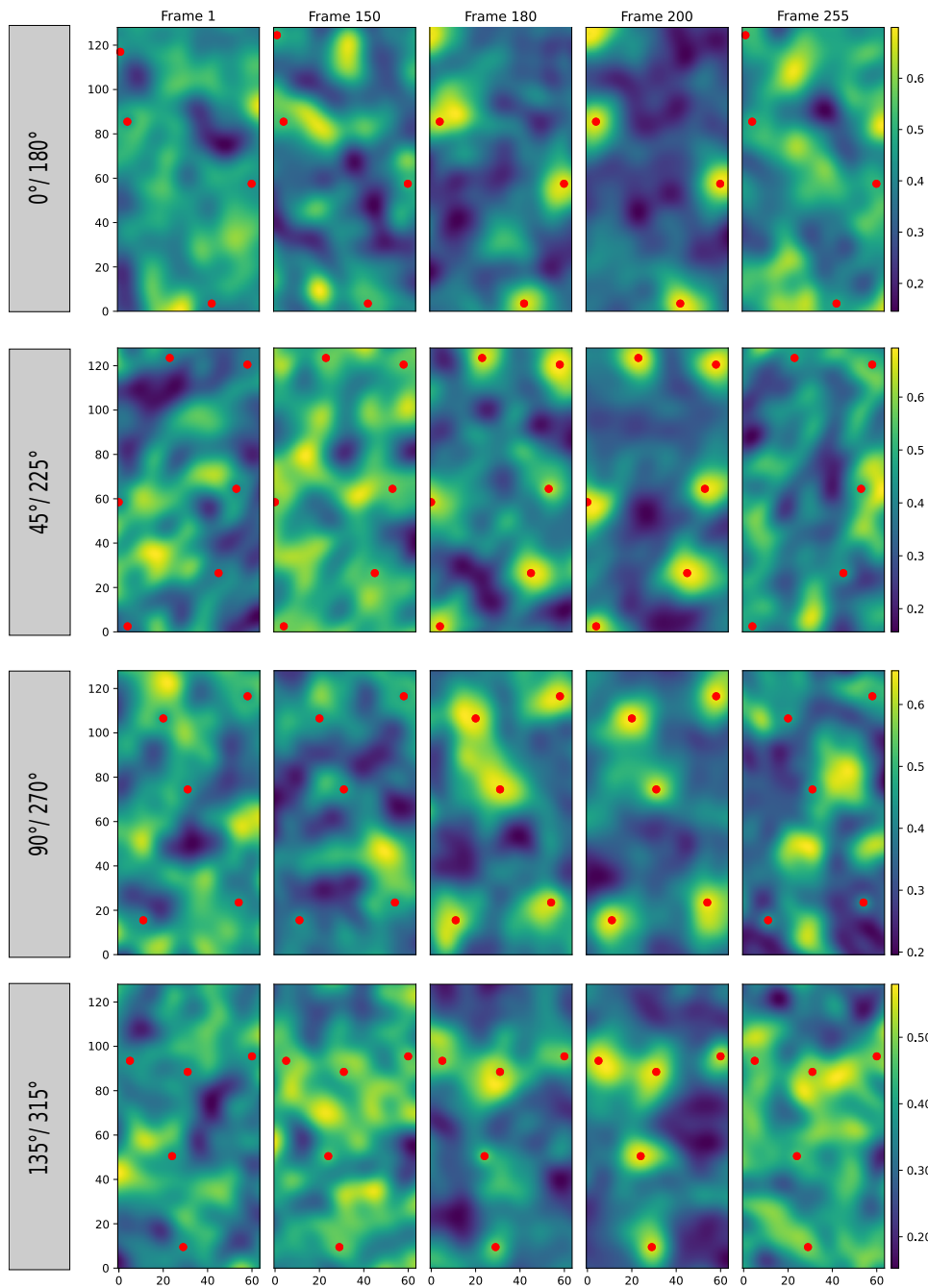
Parameter studies on  $\ell$  and  $\beta$  are carried out on six data settings summarized in Tab. 1, which differ in their respective selection of data basis and applied pre-processing steps. Latter is impacting the total number of available frames. The first three settings are solely artificial image sequences generated via the aforementioned approach (Sect. 2.4), and are based on pre-defined CRF patterns (Sect. 2.4.4) simulating columnar responses to four different orientations ( $0^\circ/180^\circ$ ,  $45^\circ/225^\circ$ ,  $90^\circ/270^\circ$ ,  $135^\circ/315^\circ$ ). Key frames of exemplary CRF realisations are shown in Fig. 17. After 150 baseline frames of unconditioned fluctuations, a regime switch to one of the four pre-defined response patterns slowly emerges. It is accompanied by an amplitude increase of the centroids and spatial inhibition of activity

**Table 1:** Overview of parameter study settings. Considerable differences in available frame sizes and numbers are resulting from the selection of data type and pre-processing steps.

	A	B.1	B.2	C.1	C.2	C.3
<b>Data Type</b>	CRF	Synthetic VSDI	Synthetic VSDI	Real VSDI	Real VSDI	Real VSDI
<b>Pre-Processing Steps</b>						
- Normalization [0,1]	✓	✓	✓	✓	✓	✓
- Spatial Bandpass Filtering	-	-	✓	-	✓	✓
- Baseline Subtraction	-	-	✓	-	✓	-
- GLM [Reynaud et al. 2011]	-	-	-	-	-	✓
<b>Frame Selection</b>						
- Initial ROI Size	128 x 64	128 x 64	128 x 64	320 x 160	320 x 160	320 x 160
- Cropped Image Size	-	-	-	128 x 64	128 x 64	128 x 64
<b>Data Availability</b>						
- Sequences per Recording Block	80	80	80	80	80	80
- Frames per Sequence	255	255	105	255	105	255
- Frame Total	20.400	20.400	8.400	20.400	8.400	20.400

in the surrounding, both peaking around frame 180 and stabilizing until frame 200. The remaining frames of each sequence exhibit the return to the baseline regime.

Across all settings, data from a single recording block comprising 80 image sequences were selected and partitioned into sets for model training (80%  $\hat{=}$  64 sequences) and testing (20%  $\hat{=}$  16 sequences). As it is a common requirement for training deep neural networks, the input values are normalized beforehand (Bishop, 2006). Each image sequence was normalized by its according sequence-specific minimum and maximum value. Due to the sigmoid activation function in the last deconvolution layer of the implemented VAE network (Fig. 32), a target value range of [0, 1] was chosen.



**Figure 17:** Parameter study, settings A, B.1, B.2: CRF patterns. Selected frames for highlighting spatio-temporal dependencies for CRF patterns pre-defined via centroids (red dots).

---

### 3.2.2 Setting A: CRF

#### Data Basis

In this setting, only raw CRFs resembling the ground truth spatio-temporal signal are used for model training and evaluation. Accordingly, no synthetic VSDI noise components are applied yet. This setting therefore resembles a scenario of perfect data quality, or alternatively, having an optimal data pre-processing approach at hand. While in practice this should never be the case, this simulation can nevertheless serve for checking proper functioning of the  $\beta$ -VAE implementation as well as benchmark for all further parameter study settings. An exemplary image sequence for this data basis is shown in Fig. 18, which is holding a single CRF pattern for  $135^\circ/315^\circ$  orientation.

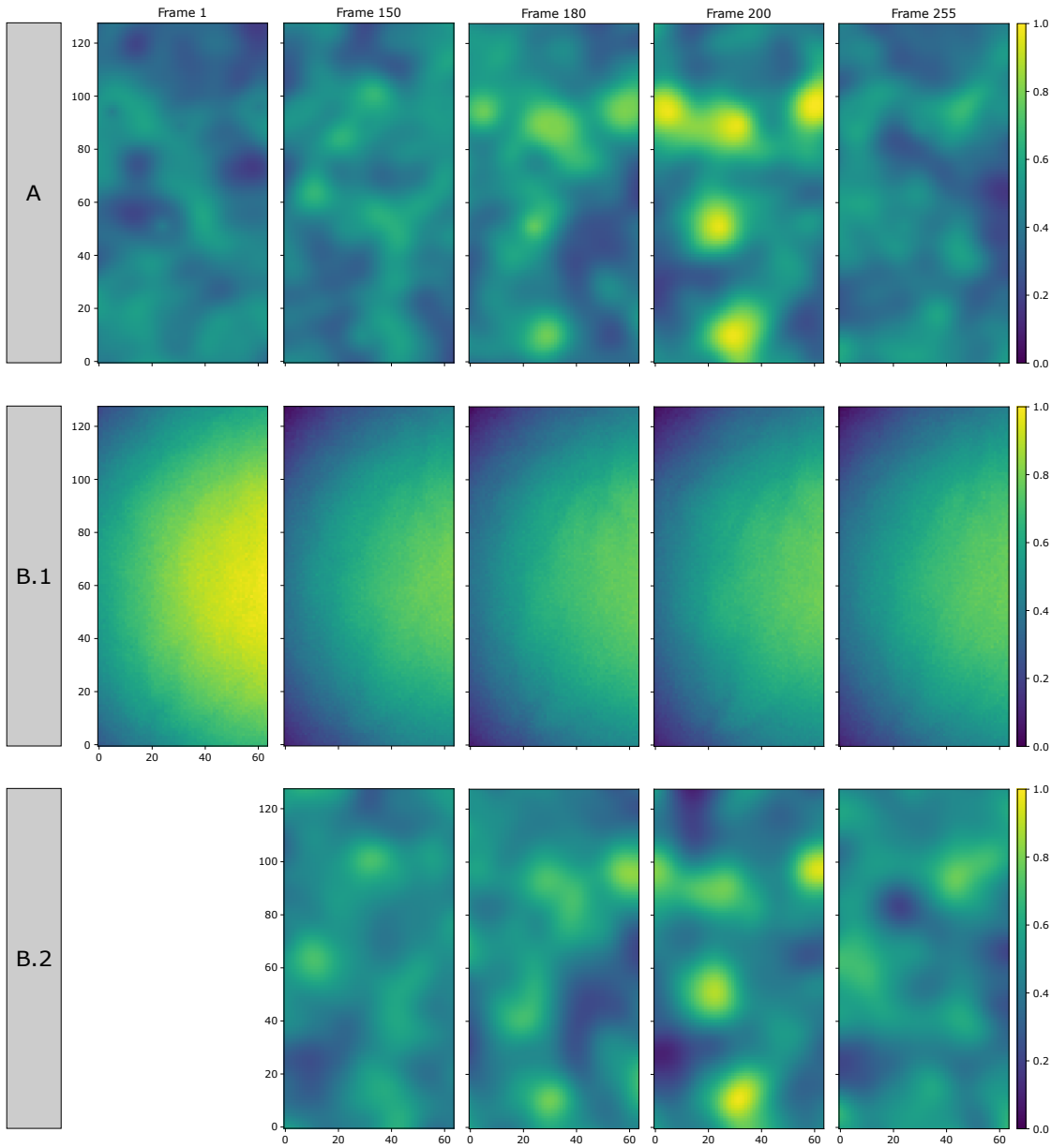
#### Model Fit

For each parameter configuration, the fit of the corresponding  $\beta$ -VAE is assessed by comparing its loss and validation loss curves after model training has been completed (Tab. 2). For the majority of configurations, both curves are decreasing along respective training epochs. As the loss remains below the validation loss throughout all epochs and only a small gap between both curves exists, these findings indicate a good fit of respective model specifications. This especially applies to configurations with lower numbers of latent variables ( $\ell \leq 5$ ) as well as higher model capacities ( $\ell > 50$ ). Interestingly, for medium capacities of  $10 \leq \ell \leq 50$  latent dimensions, symptoms of model overfitting are noticeable, as validation loss curves are increasing towards the end of model training whereas the loss keeps decreasing. This implies that these model specifications do not seem to generalize well on unseen data although performing well on training data. Putting stronger weights on the KL term seems to mitigate this effect, as signs for good model fit can be observed again for higher values of  $\beta$ . For cases with very low numbers of latent dimensions ( $\ell \leq 2$ ) and strong regularization ( $\beta \geq 75$ ), model underfitting is indicated by the loss reaching larger values than validation loss, as well as more pronounced gaps between both curves.

#### Reconstruction Quality

The impact of model capacity as well as the KL weight on the reconstruction quality is assessed in terms of MSE between model inputs and outputs. Here, image sequences held back as validation partition comprising 4.080 frames are fed into the VAE model. Subsequently, the MSE is computed between each input frame and its corresponding reconstruction. The distributions of MSE values are illustrated per parameter configuration via boxplots in Fig. 19. A substantial impact of the model capacity becomes evident within





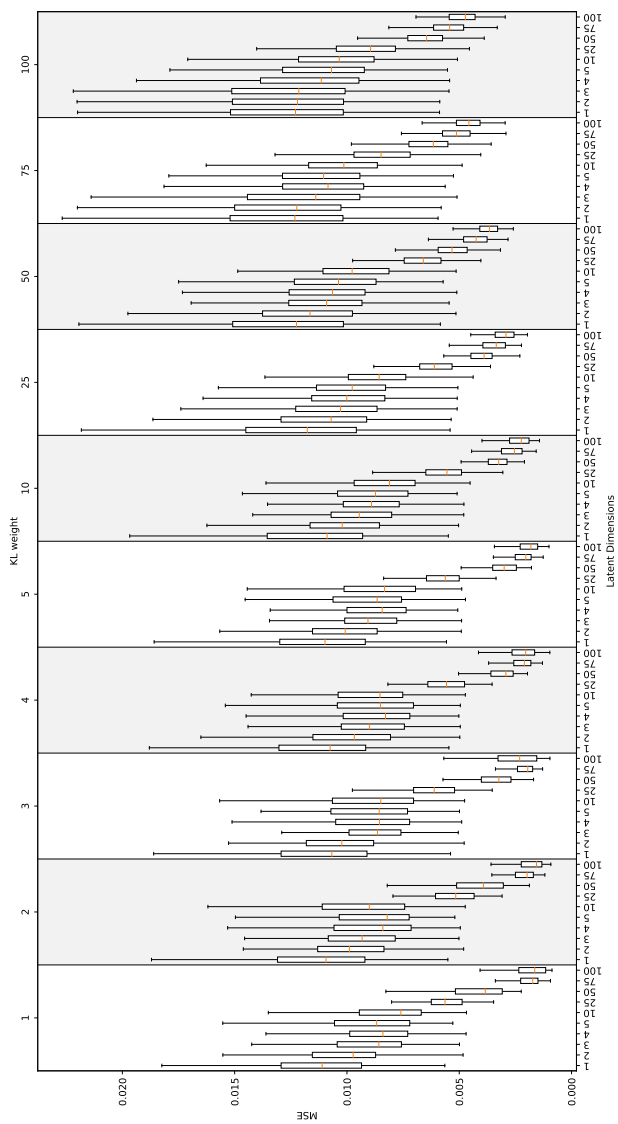
**Figure 18:** Parameter study, settings A, B.1, B.2: Data basis. Comparison of pre-processing approaches for a single CRF pattern ( $135^\circ/315^\circ$ ) A: raw; B.1: CRF & artificial VSDI components; B.2: CRF & artificial VSDI components, denoising via baseline subtraction (hampering interpretability of baseline frames, thus omitted) and 2-D bandpass filtering. Key frames were selected from the artificial image sequence comprising  $T = 255$  frames with image size  $128 \times 64$ .

**Table 2:** Parameter study, setting A: Assessment of model fit. ✓: good fit, /: ambiguous, O: model overfitting, U: model underfitting.

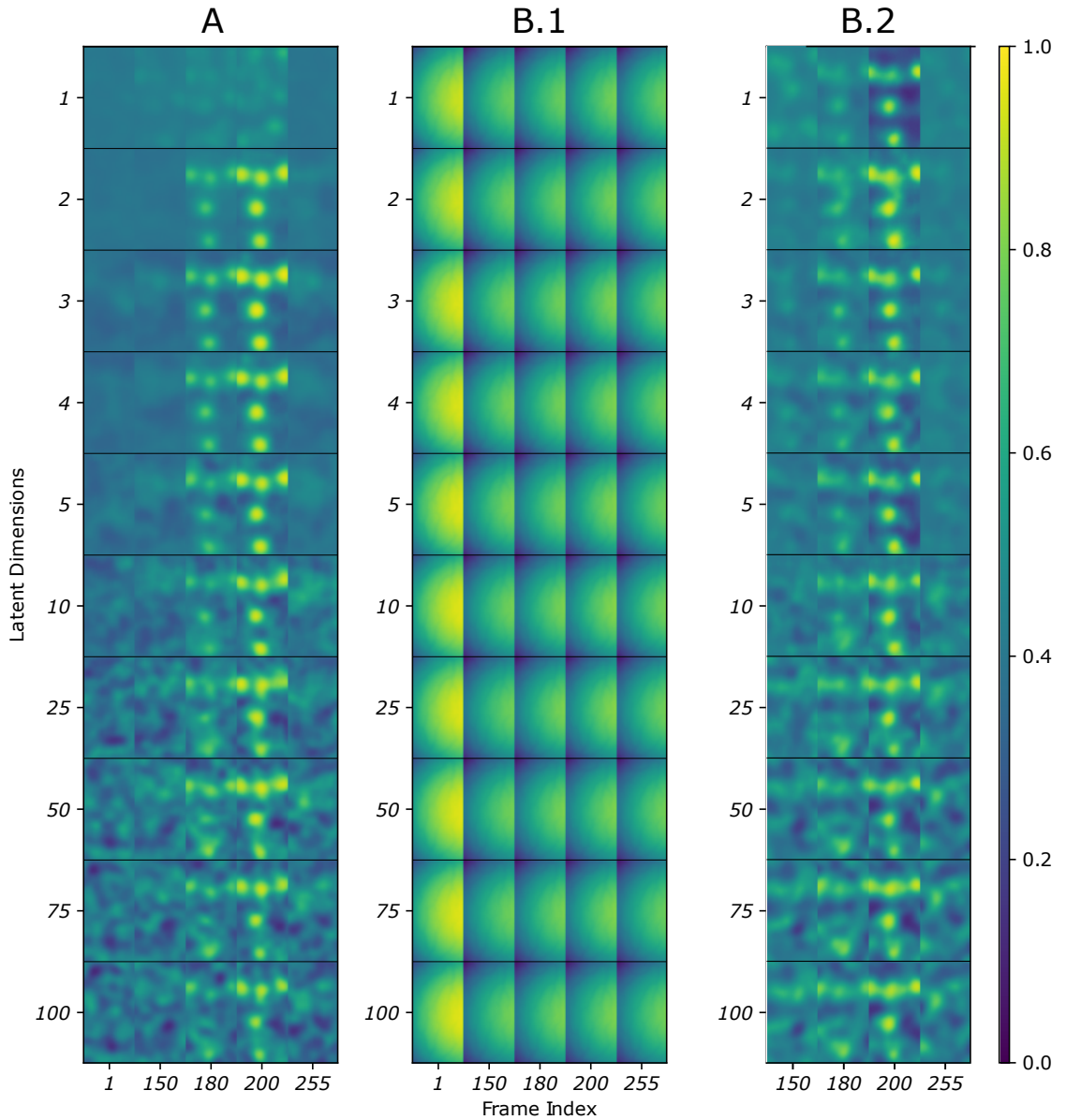
KL Weight	Latent Dimensions										
	1	2	3	4	5	10	25	50	75	100	
1	✓	✓	✓	✓	/	O	O	/	✓	✓	
2	✓	✓	✓	/	/	O	O	/	✓	✓	
3	✓	✓	✓	✓	/	O	O	/	✓	✓	
4	✓	✓	✓	✓	/	O	O	/	✓	✓	
5	✓	/	✓	✓	/	O	O	✓	✓	✓	
10	✓	✓	✓	✓	✓	O	O	/	✓	✓	
25	✓	✓	✓	✓	✓	/	O	/	/	/	
50	/	✓	✓	✓	✓	✓	✓	/	/	✓	
75	U	/	✓	✓	✓	✓	✓	/	/	/	
100	U	U	/	✓	✓	✓	✓	✓	✓	✓	

each parametrization of the KL weight  $\beta$ . For increases in the dimensionality of latent space, descriptive statistics summarizing the central tendency (quartiles and median) of the MSE distributions decrease. This becomes especially apparent for larger step sizes in latent dimensionality (e.g. from  $\ell = 10$  to  $\ell = 25$ ). As the KL weight increases, the reconstruction quality seems to get slightly worse between different specifications of  $\beta$ . In this regard, larger step sizes for  $\beta$  lead to more pronounced increases in MSE. Yet, when comparing boxplots for different  $\beta$  with identical  $\ell$ , this effect does not seem to be significant, as most of respective boxplot pairs show considerable overlap. The variability of MSE values increases with higher KL weights, as the distribution of MSE values are getting more dispersed and spreaded, which is indicated by greater interquartile ranges and distances between whiskers.

The influence of model capacity on reconstruction quality is also evident when models with different numbers of latent variables are compared (Fig. 20). Here, a single CRF containing the spatio-temporal pattern for  $135^\circ / 315^\circ$  orientation is defined as model input. While all model specifications (with the exception of  $\ell = 1$ ) are able to closely reconstruct the pre-defined pattern, significant differences become apparent for baseline frames before and after the response phase. Increasing the number of latent variables ( $\ell \geq 10$ ) seem to enable the model to cover these spatio-temporal fluctuations unrelated to the response pattern with higher level of detail.



**Figure 19:** Parameter study, setting A: MSE. For each parameter configuration between model capacity (in terms of latent dimensions) and regularization (in terms of KL weighting), a separate VAE model is specified. All frames of the validation partition comprising 4.080 frames are passed into the corresponding pre-trained model. The MSE between the current input image and its corresponding reconstruction is computed frame-wise. Subsequently, resulting distributions of MSE values are illustrated as boxplots for every model specification.



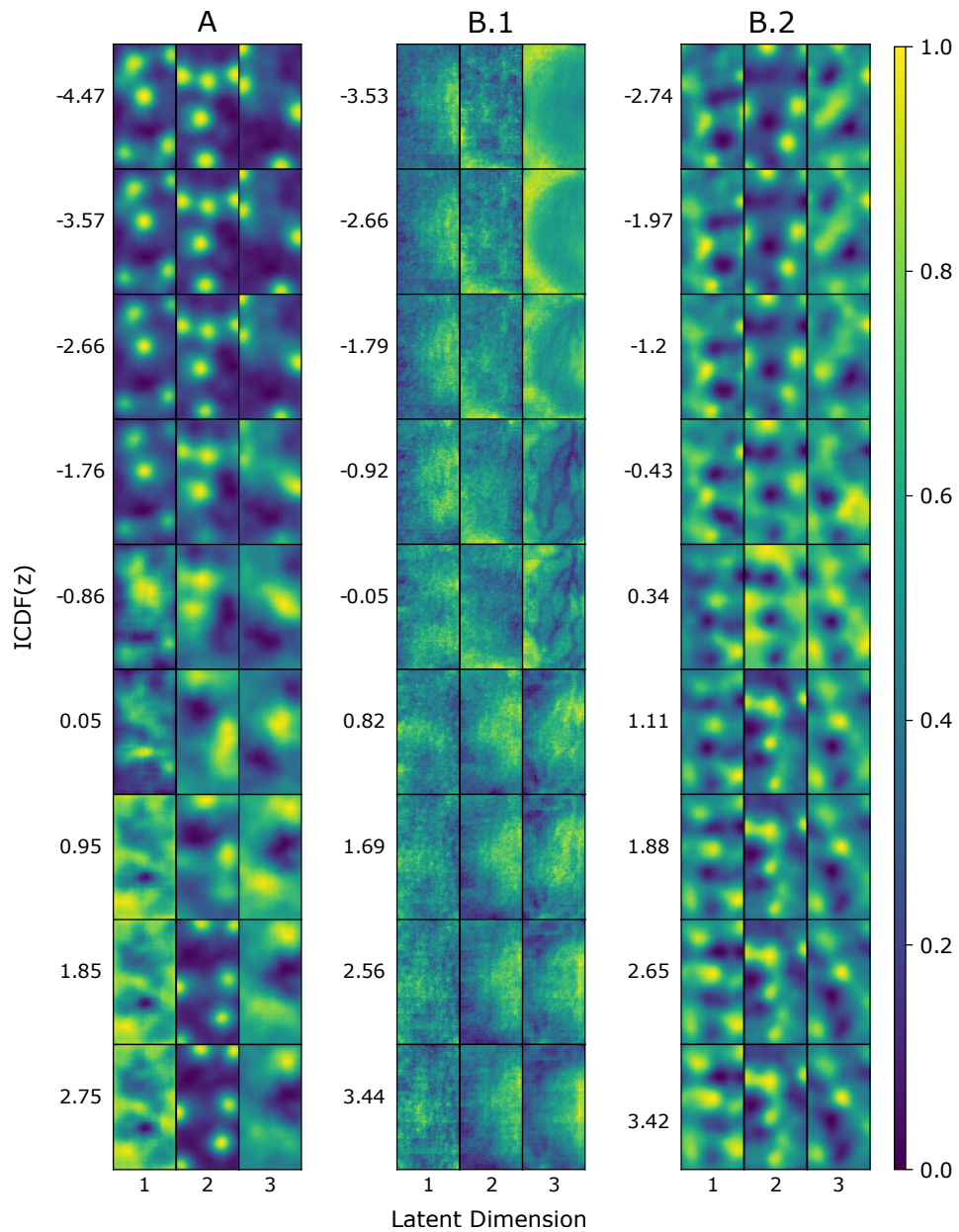
**Figure 20:** Parameter study, settings A, B.1, B.2: Input reconstructions. Different synthetic VSDI components and denoising steps are applied on a single CRF pattern (here:  $135^\circ/315^\circ$  orientation). A: CRF; B.1: CRF & artificial VSDI components; B.2: CRF & artificial VSDI components, denoising via baseline subtraction (hampering interpretability of baseline frames, thus omitted) and 2-D bandpass filtering. Reconstructed key frames (x-axis) from  $\beta$ -VAE specified with varying numbers of latent dimensions in  $z$  (y-axis) and KL weight  $\beta = 1$ .

---

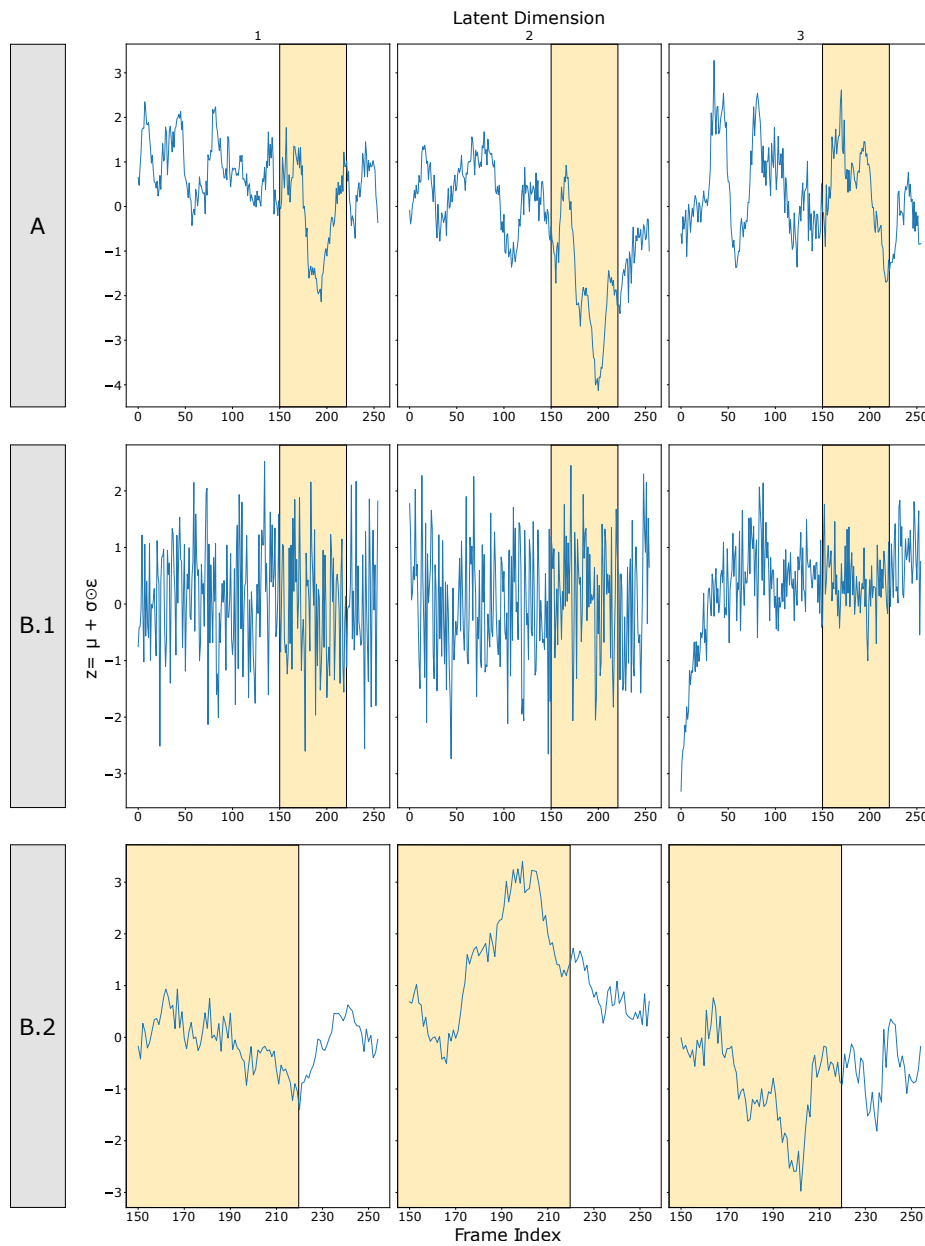
## Feature Extraction

When inspecting latent space walks, it becomes obvious that features of all four pre-defined CRF patterns are extracted by the VAE. Especially for low latent dimensionalities  $\ell$  and small KL weight  $\beta$ , whole patterns are encoded by the tails of a single latent variable's encodings distribution. To exemplify these results, a latent space walk for a single model with parameter configuration  $\beta = 4$ ,  $\ell = 3$  is illustrated in Fig. 21. Here, negative values for the first latent variable are encoding all activation centroids related to the patterns for orientation  $90^\circ / 270^\circ$ , whereas positive values seem to cover random fluctuations as global spatial component. By the second latent dimension, two patterns with mutually exclusive activation centers are extracted: on negative value range, centroids corresponding to orientation  $135^\circ / 315^\circ$  are found, while positive values are assigned to the pattern for  $45^\circ / 225^\circ$ . Encodings close to zero are resulting in more global spatial activations, appearing as hypothetical transition between both orientation patterns. Finally, the remaining pattern for orientation  $0^\circ / 180^\circ$  is covered by the negative value range of the third latent dimension. Positive encoding values of the first and second latent variable cover activation across the whole image domain, seemingly unrelated to the pre-defined CRF patterns.

When instead only a single validation sequence is used as model input (here: CRF with pattern for  $135^\circ / 315^\circ$  orientation), corresponding encodings can be used to track deviations from the prior for each latent dimension on individual frame basis. As exemplified in Fig. 22 for a  $\beta$ -VAE model specification with  $\beta = 4$ ,  $\ell = 3$ , shifts from the prior mean of zero are observable in the second latent variable. Here, the most pronounced deviations can be observed for encodings of the second latent variable around frame index 200, falling into negative value range. Interestingly, respective encodings closely follow the temporal dynamics of the CRF pattern, as the simulated response amplitude is raised from frame 180 to 200, stabilizes until frame 220 and returns to baseline towards the end of the sequence. This further underlines that the second latent dimension seems to encode pre-defined CRF pattern in the given input sequence.



**Figure 21:** Parameter study, settings A, B.1, B.2: Latent space walk. Samples from the inverse CDF of  $p(z) \sim N(0, 1)$  (y-axis) are plugged in for a single latent variable (x-axis) while keeping all remaining variables at their median encoding value. Decodings of latent space walks are resulting from  $\beta$ -VAE model specified with KL weight  $\beta = 4$  and latent dimensionality of  $\ell = 3$ .



**Figure 22:** Parameter study, A, B.1, B.2: Frame-wise encodings. Resulting from  $\beta$ -VAE model ( $\beta = 4$ ,  $\ell = 3$ ). Samples for each latent variable in  $z$  (y-axis) plotted against frame indices (x-axis) of a single input image sequence comprising a CRF pattern for  $135^\circ / 315^\circ$  orientation. Background colors marking simulated baseline (white) and stimulation (yellow) phases.



### 3.2.3 Setting B.1: Synthetic VSDI, raw

#### Data Basis

To assess which features the  $\beta$ -VAE is capable to extract from datasets with low SNR and noise components typical for VSDI, synthetic datasets are used as model inputs. These sequences follow the fully known data-generating process described in Sect. 2.4.8. For this setting, raw sequences are passed as model input without applying any pre-processing steps. Key frames of a selected image sequence comprising the CRF pattern for 135/315 orientation are shown in Fig. 18 for exemplifying this data basis. While the pre-defined response pattern cannot be easily recognized anymore in corresponding frames, spatial confounders related to illumination and blood vessel networks become apparent as dominant signal components. Furthermore, a global decrease in signal amplitude along corresponding frames is simulating the temporal effect of dye bleaching.

#### Model Fit

In Tab. 3, a clear distinction between models with good fit and underfitting can be observed. Interestingly, these problematic fits occur when using medium to high numbers of latent dimensions ( $\ell \geq 25$ ) and low values for the KL weight ( $\beta < 10$ ). Configurations with stronger KL weighting and lower model capacity in turn seem to result in appropriate model fits.

**Table 3:** Parameter study, setting B.1: Assessment of model fit.  $\checkmark$ : good fit,  $/$ : ambiguous, O: model overfitting, U: model underfitting.

KL Weight	Latent Dimensions									
	1	2	3	4	5	10	25	50	75	100
1	/	U	U	U	U	U	U	U	U	U
2	/	$\checkmark$	/	U	U	U	U	U	U	U
3	$\checkmark$	$\checkmark$	$\checkmark$	/	/	U	U	U	U	U
4	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	U	U	U	U	U
5	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	/	U	U	U	U
10	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	/	U	U	U	U
25	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	/	/	U	U
50	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	/	/
75	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
100	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$



---

## Reconstruction Quality

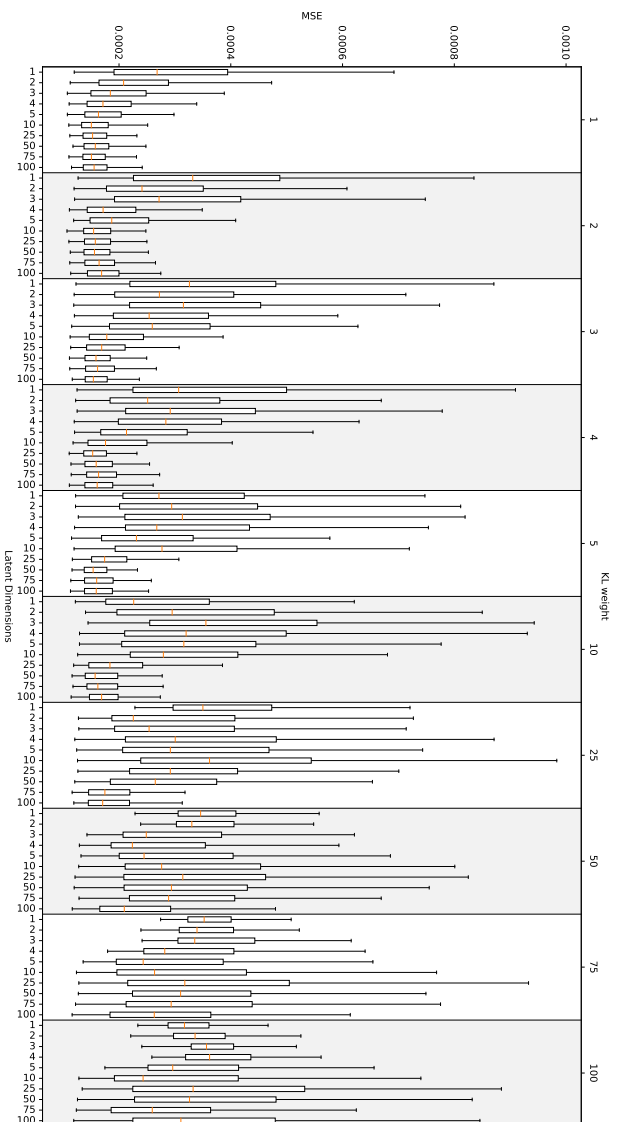
Only the parametrization of regularization weight  $\beta = 1$  leads to a decrease in median MSE as well as dispersion and spread of respective MSE distributions while increasing model capacities. When setting the KL weight to values between  $2 \leq \beta \leq 10$ , differences between boxplots (Fig. 23) for lower model capacities ( $\ell \leq 10$ ) appear to become less significant, as considerable overlaps between interquartile ranges become apparent. Also, median MSE values seem to vary unsystematically with an increase in  $\ell$ . When instead using stronger regularization weights, the interaction effect between model capacity and regularization seems to reverse: interquartile ranges and distance between whiskers are now increasing for configurations with high values for  $\beta$  and  $\ell$ . For configurations with high model regularization ( $\beta \geq 50$ ), the smallest dispersions and spreads of boxplots can now be observed for low latent dimensionality ( $\ell \leq 4$ ).

For all configurations of latent dimensionality, dominant spatial components (illumination, blood vessels) contained in the model inputs also appear in corresponding reconstructions, latter being illustrated in Fig. 20. Aside of these spatial noise components, a decrease in signal amplitude due to dye bleaching can also be seen along each reconstructed sequence.

## Feature Extraction

The latent space walks illustrated in Fig. 21 reveal that the elliptical-shaped illumination is appearing over all latent dimensions, while the branching blood vessel structures seem to be encoded solely by the third variable. In this setting, response-related activity following the pre-defined CRF patterns cannot be clearly identified from visual inspection of the latent space.

When inspecting the frame-wise encodings (Fig. 22) for a single validation sequence comprising the CRF for  $135^\circ/315^\circ$  orientation, samples for the first and second latent variable seem to scatter randomly around the prior mean of zero without showing any systematic deviations related to the response pattern. In contrast, the third variable shows encodings which rise rapidly out of the negative value range from the beginning of the sequence and then settle to a value of zero. As samples are following the dynamics of the artificial dye bleaching very closely, this suggests that this variable is indeed able to cover the amplitude decay across the sequence. This presumption is further underlined when inspecting the latent space walk for corresponding variable (Fig. 21). It appears that the illumination is encoded by negative values while blood vessel structures appear for samples close to zero.



**Figure 23:** Parameter study, setting B.1: MSE. For each parameter configuration between model capacity (in terms of latent dimensions) and regularization (in terms of KL weighting), a separate VAE model is specified. All frames of the validation partition comprising 4.080 frames are passed into the corresponding pre-trained model. The MSE between the current input image and its corresponding reconstruction is computed frame-wise. Subsequently, resulting distributions of MSE values are illustrated as boxplots for every model specification.

---

### 3.2.4 Setting B.2: Synthetic VSDI incl. Pre-Processing (Blank Subtraction)

#### Data Basis

As in the previous setting, artificial VSDI confounders are applied on corresponding datasets. However, data pre-processing steps are subsequently taken on each image, including baseline subtraction and 2-D spatial bandpass filtering. For the baseline subtraction, the last 30 frames of every sequence’s baseline are averaged to a zero-frame which in turn is subtracted from each individual frame from the same sequence. The spatial bandpass filter is implemented as Difference-of-Gaussians via two low-pass 2-D Gaussian filters. Latter are specified with different radii (here:  $160 \mu\text{m}$ ,  $960 \mu\text{m}$ ) for expressing plausible lower and upper limits of column sizes to be emphasized in the input images. Key frames of an exemplary sequence with a CRF pattern for  $135/315$  orientation is shown in Fig. 18. Due to the choice of baseline subtraction as pre-processing technique, frames of the corresponding baseline become essentially uninformative and hence are omitted for further data analysis. This in turn greatly reduces the total number of available images to 8.400 images per recording block.

#### Model Fit

Comparisons between loss and validation loss curves for this setting (Tab. 4) reveal appropriate model fitting for configurations with low to medium KL weighting ( $\beta \leq 25$ ) and either small ( $\ell < 4$ ) or large ( $\ell \geq 50$ ) dimensionality of latent space. By contrast, models with medium number of latent variables tend to be more problematic, which becomes especially apparent for specifications with  $\ell = 10$  exhibiting signs of overfitting the training data. For all configurations with high KL weighting ( $\beta \geq 50$ ), corresponding loss and validation loss curves only show ambiguous tendencies.

#### Reconstruction Quality

For the pre-processed data basis, similar trends as for raw CRF data (setting A) are apparent with respect to the MSE distributions (Fig. 24). A predominant tendency of the median MSE decreasing for higher dimensionalities of the latent space  $\ell$  can be observed, which applies for all configurations of KL weight  $\beta$ . In most cases, the strongest decrease in MSE distributions is seen between  $\ell = 10$  and  $\ell = 25$ . As previously stated for setting A, the variability of MSE distributions increases with higher regularization weights, which is derived from larger dispersion and spread of respective boxplots.

Similar results to the previous setting A are also evident in the input reconstructions (Fig. 20). The given CRF ( $135^\circ/315^\circ$  orientation) around frame 200 is accurately mapped

**Table 4:** Parameter study, setting B.2: Assessment of model fit. ✓: good fit, /: ambiguous, O: model overfitting, U: model underfitting.

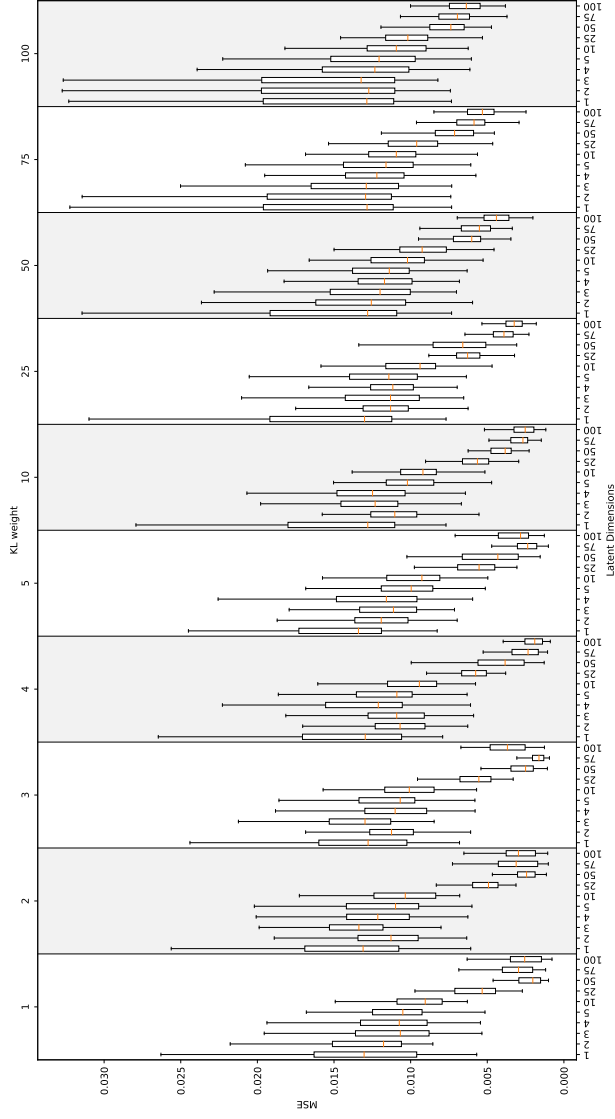
<i>KL Weight</i>	<i>Latent Dimensions</i>									
	1	2	3	4	5	10	25	50	75	100
1	U	✓	/	O	O	O	/	✓	✓	✓
2	U	✓	✓	/	/	O	/	✓	✓	✓
3	✓	✓	/	/	O	O	/	✓	✓	✓
4	✓	✓	✓	/	/	O	/	✓	✓	✓
5	✓	✓	✓	/	/	O	/	✓	✓	✓
10	/	✓	✓	/	O	O	/	✓	✓	✓
25	O	✓	/	/	/	O	/	✓	✓	✓
50	/	/	/	/	/	/	/	/	/	/
75	/	/	/	/	/	/	/	/	/	/
100	/	/	/	/	/	/	/	/	/	/

across all configurations for  $\ell$ . With larger latent dimensions, random fluctuations before and after pattern onset can be covered more accurately.

### Feature Extraction

The latent space walks illustrated in Fig. 21 reveal that after data pre-processing, features of the pre-defined CRF patterns are getting extracted again. Nevertheless, reconstructions show considerably more background noise as well as more blurred activation centers when compared to its counterparts in setting A. The pattern for  $90^\circ / 270^\circ$  orientation is now encoded via positive value range of encodings in the first latent dimension. The second latent variable in turn is covering two orthogonal patterns related to orientations with  $45^\circ / 225^\circ$  (negative values) as well as  $135^\circ / 315^\circ$  (positive values). The third latent variable is substantially more difficult to interpret. It appears that this dimension seems to encode a local singular activation shared by multiple patterns by its positive value range, while mapping residual features on global image domain through negative values. Activation centers related to  $0^\circ / 180^\circ$  orientation can not be readily detected from the latent space walks of any dimension.

When passing a single validation sequence comprising a CRF pattern for  $135^\circ / 315^\circ$  orientation into the encoder, encodings of corresponding latent dimension (second variable) strongly shift towards positive value range during stimulation phase and returns towards the prior mean of zero towards the end of the sequence. Interestingly, while the unrelated



**Figure 24:** Parameter study, setting B.2: MSE. For each parameter configuration between model capacity (in terms of latent dimensions) and regularization (in terms of KL weighting), a separate VAE model is specified. All frames of the validation partition comprising 1.680 frames are passed into the corresponding pre-trained model. The MSE between the current input image and its reconstruction is computed frame-wise. Subsequently, resulting distributions of MSE values are illustrated as boxplots for every model specification.

---

first dimension keeps fluctuating around the prior mean of zero, the third latent variable exhibits considerable deviations into negative value range for these stimulated key frames.

### 3.2.5 Setting C.1: VSDI, raw

#### Data Basis

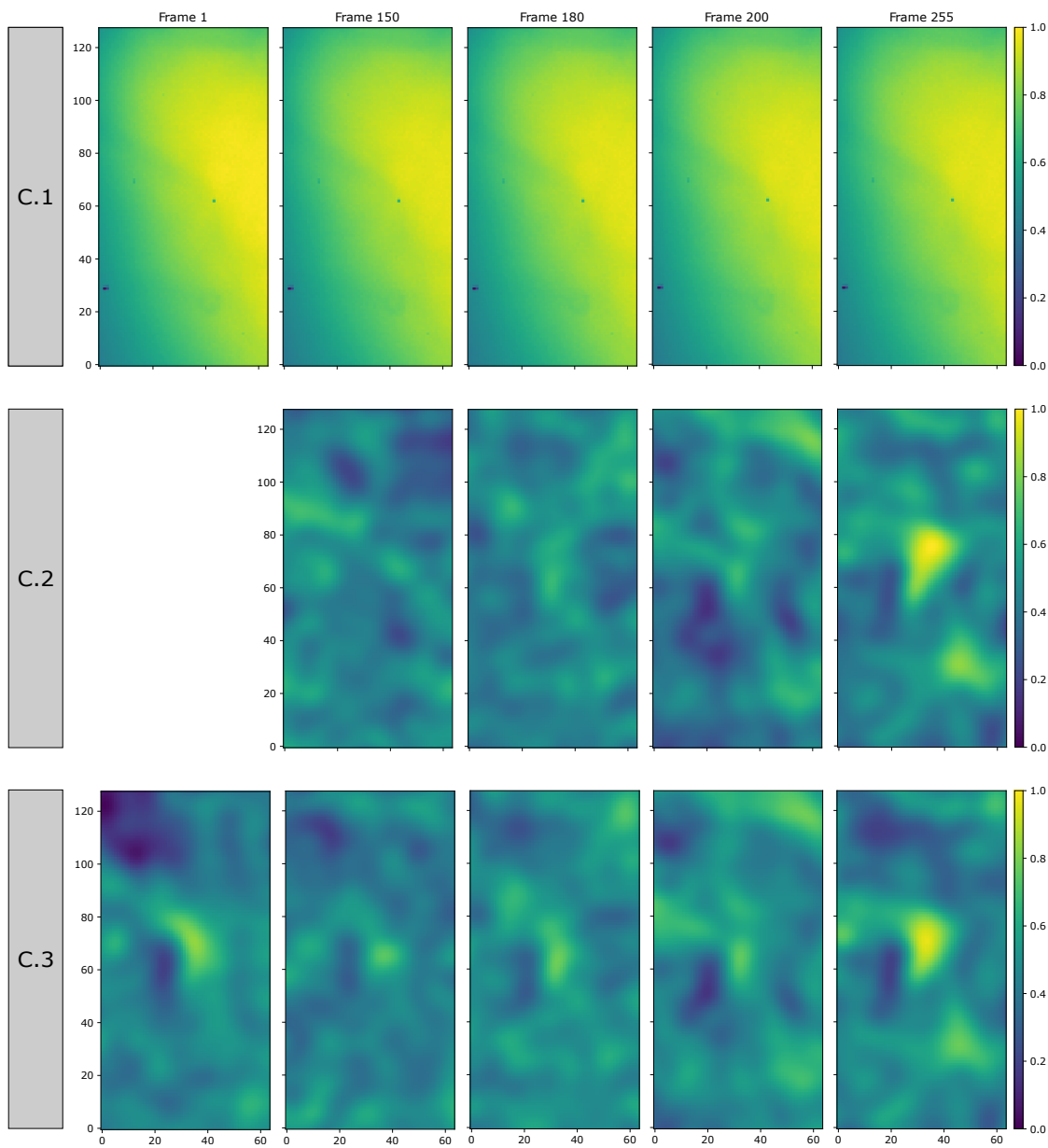
Starting with setting C.1, the data basis is now changed from artificial to real experimental data. For this purpose, data from a single VSDI experiment (subject ID: 092413) are processed, which is following the stimulation paradigm described in Sect. 2.1.2 using gratings with different directions. The first recording block immediately after initial dye staining was selected which contains eight trials, each comprising recordings for eight directions ( $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ$ ) and two additional blank recordings. In total, 80 sequences corresponding to 20.400 images are used. For setting C.1, no denoising method is applied yet. The only pre-processing steps include image cropping to target dimensionality of 128x64 pixels as well as sequence-wise value normalization to range  $[0, 1]$ . Selected frames of an exemplary single recording for  $0^\circ$  are shown in Fig. 25. Due to low SNR and dominant noise components especially in terms of illumination, frames appear nearly identical along the sequence duration.

#### Model Fit

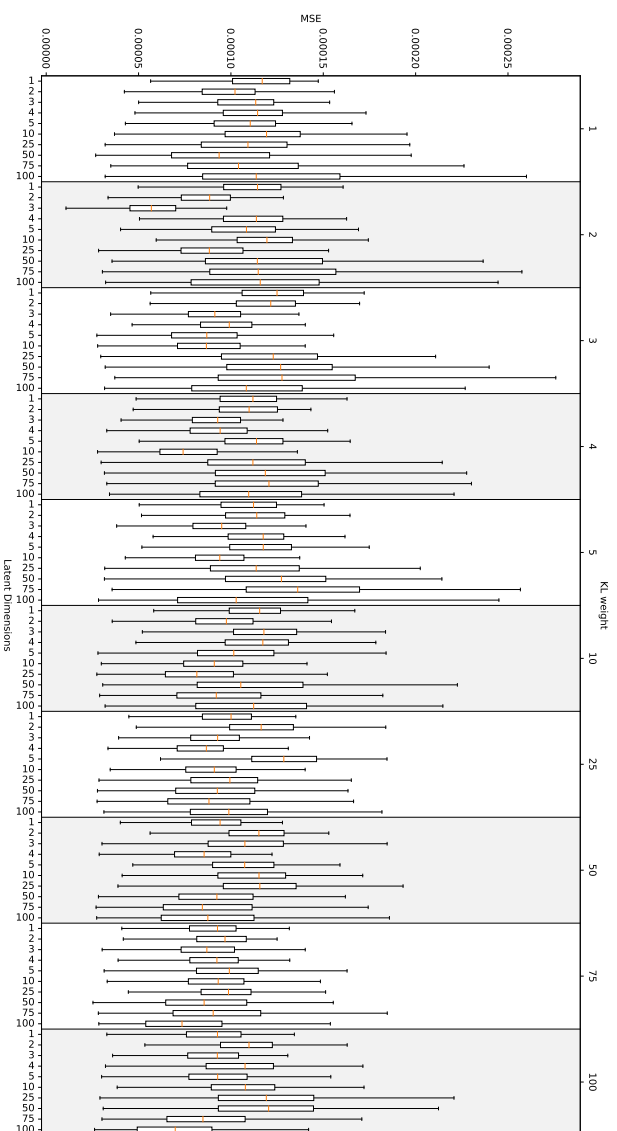
Regarding the fit of  $\beta$ -VAE, a sharp drop in loss curves within the first couple of training epochs can be observed for any configuration of model capacity and regularization. This indicates that the model is learning the most about the input data very early in the beginning of the training process and is gaining comparatively little additional information after this initial phase. When restricting the curves to all but the first epoch, signs of adequate model fit can be detected throughout all parameter choices for  $\beta$  and  $\ell$ , i.e. continuous decrease, loss remains below validation loss and small gaps between both curves.

#### Reconstruction Quality

The boxplots for MSE values between model inputs and corresponding reconstructions show no clear trend for this setting with variations in model capacity or regularization. As shown in Fig. 26, nearly all distributions seem to be largely overlapping, indicating no significant difference after re-parametrization of  $\ell$  or  $\beta$ . Reconstructions of the input sequence show hardly any variations when compared for



**Figure 25:** Parameter study, settings C.1, C.2, C.3: Data basis. Different denoising steps are applied on a single VSDI recording (here:  $0^\circ$  direction). C.1: raw; C.2: denoising via baseline subtraction (hampering interpretability of baseline frames, thus omitted) and 2-D bandpass filtering; C.3: VSDI, denoising via GLM and 2-D bandpass filtering. Key frames selected from real VSDI image sequence comprising  $T = 255$  frames cropped to target image size  $128 \times 64$ .



**Figure 26:** Parameter study, setting C.1: MSE. For each parameter configuration between model capacity (in terms of latent dimensions) and regularization (in terms of KL weighting), a separate VAE model is specified. All frames of the validation partition comprising 4,080 frames are passed into the corresponding pre-trained model. The MSE between the current input image and its corresponding reconstruction is computed frame-wise. Subsequently, resulting distributions of MSE values are illustrated as boxplots for every model specification.



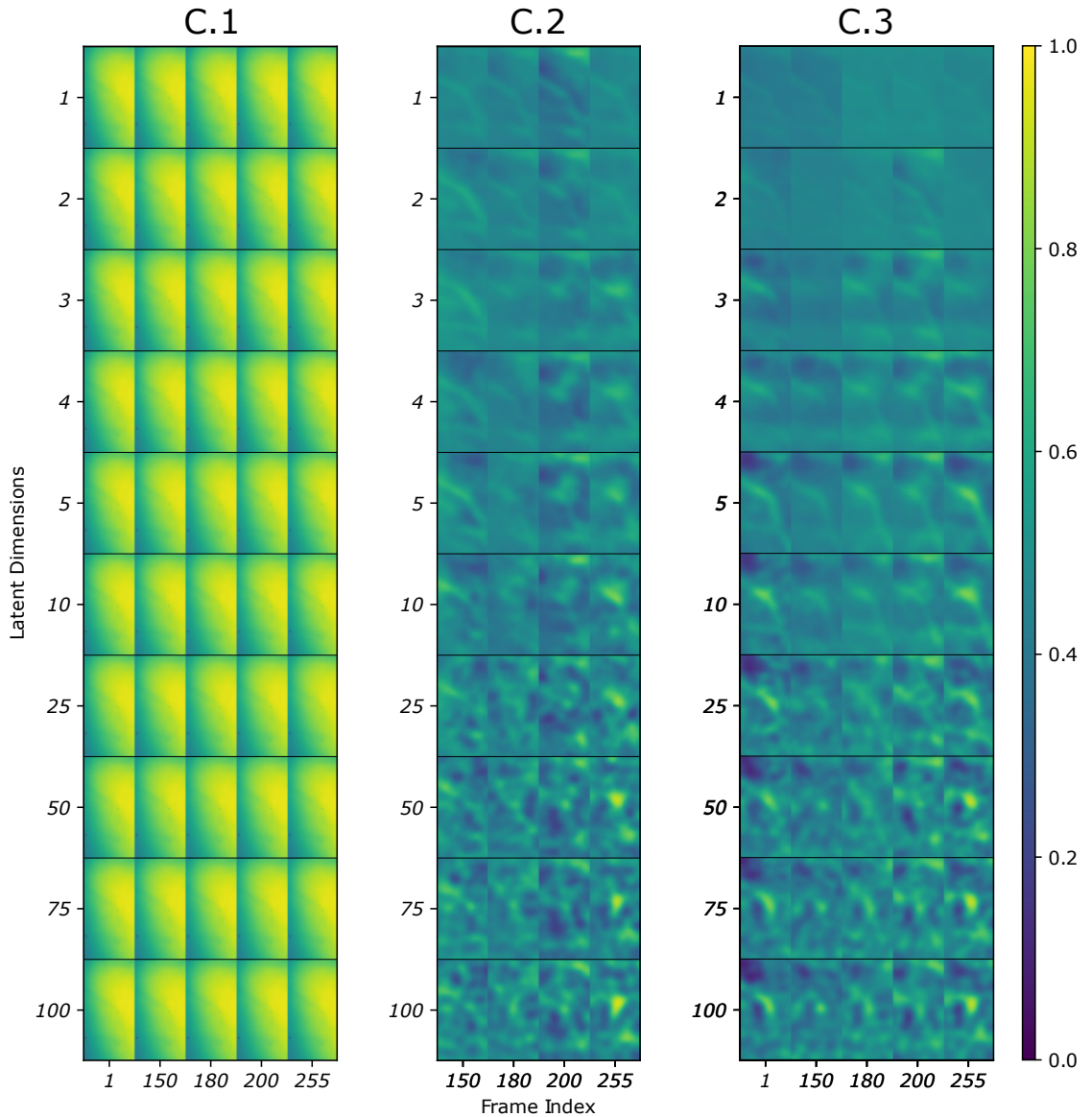
---

different specifications regarding model capacity and regularization. As illustrated in Fig. 27, each prediction shows the same predominant spatial features of the input sequence, especially illumination and blood vessel structures.

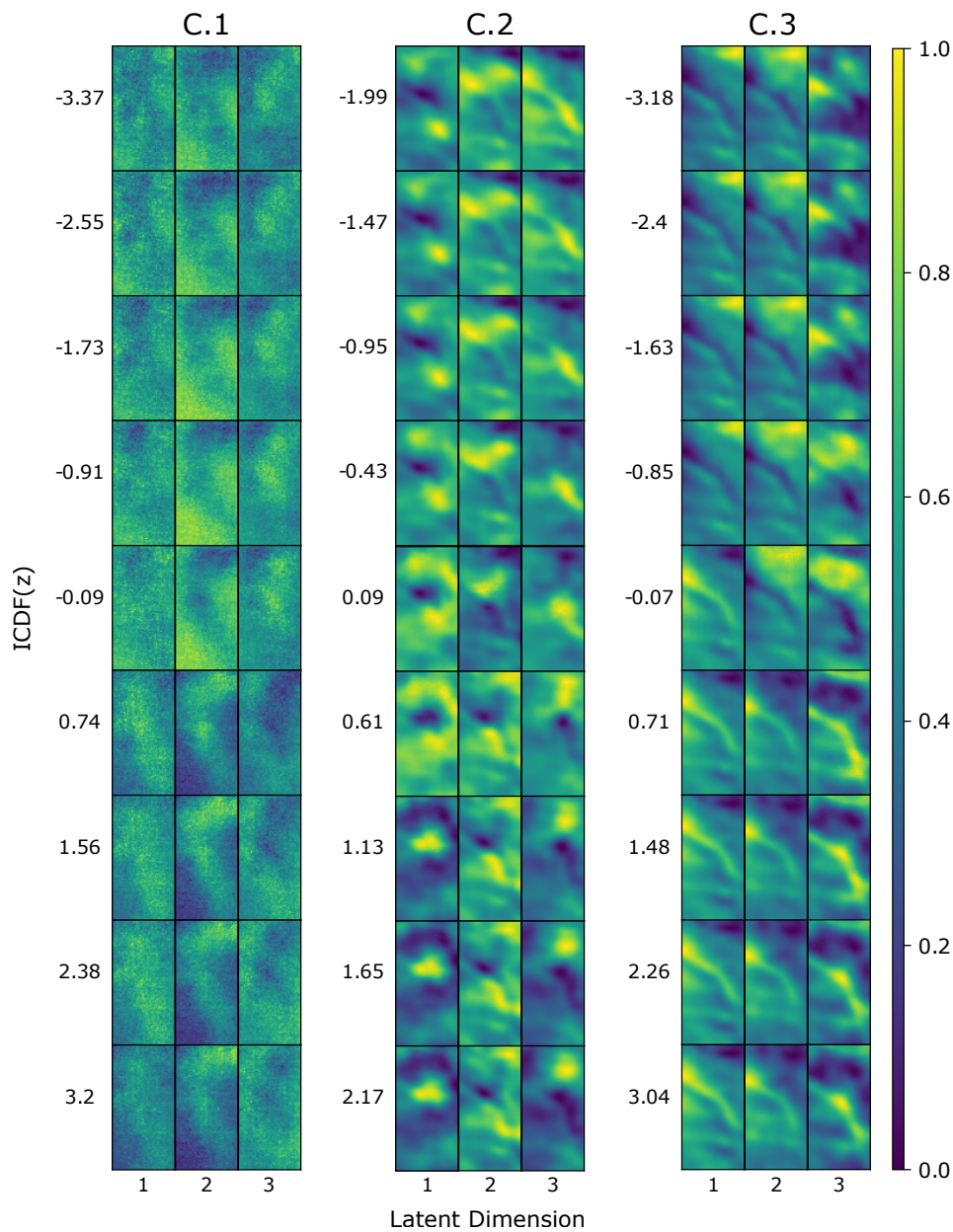
### **Feature Extraction**

The latent space walks in Fig. 28 allow only vague interpretations regarding encoded features of the input sequence. Here, the second latent variable shows indistinct activations spreading diagonally from upper left to lower right image borders. In contrast, no clear tendencies can be recognized for the first and third variables, which represent global activations on whole image domain. Remarkably, blood vessel networks do not appear to be encoded by any latent dimension despite being clearly present in the input reconstructions.

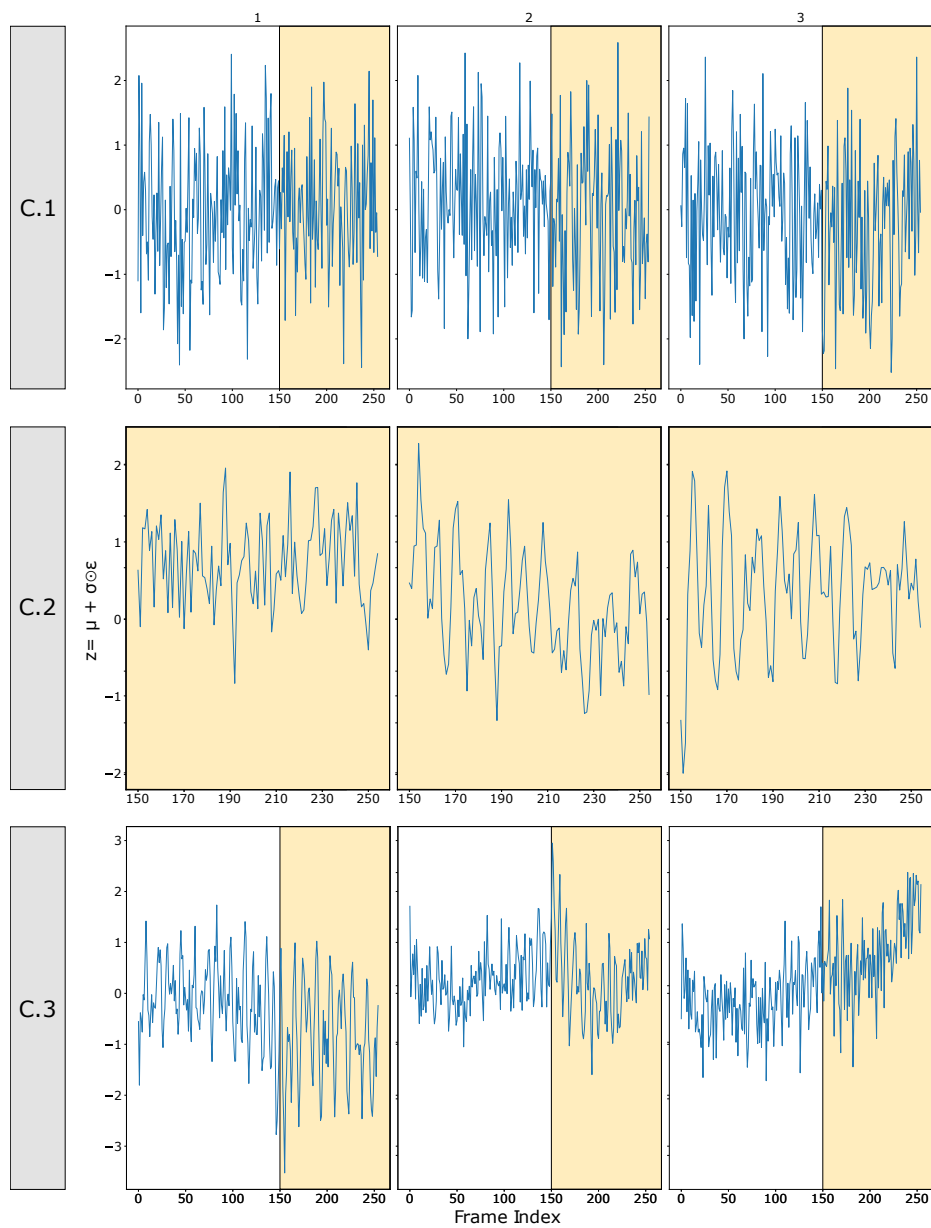
Sampled encodings of all three variables (Fig. 29) tend to closely follow the prior distribution, as values fluctuate around the prior mean of zero. Furthermore, no fundamental differences between baseline and stimulation regimes can be recognized for any latent dimension, neither in terms of varying mean or variance over frame indices.



**Figure 27:** Parameter study, settings C.1, C.2, C.3: Input reconstructions. Different signal components and denoising steps are applied on a single VSDI sequence. C.1: raw VSDI, no denoising; C.2: VSDI, denoising via baseline subtraction (hampering interpretability of baseline frames, thus omitted) and 2-D bandpass filtering; C.3: VSDI, denoising via GLM and spatial bandpass filtering. Reconstructions of selected key frames (x-axis) from  $\beta$ -VAE specified with varying numbers of latent dimensions in  $z$  (y-axis) and KL weight  $\beta = 1$ .



**Figure 28:** Parameter study, settings C.1, C.2, C.3: Latent space walk. Samples from the inverse CDF of  $p(z) \sim N(0, 1)$  (y-axis) are plugged in for a single latent variable (x-axis) while keeping all remaining variables at their median encoding value. Decodings of latent space walks are resulting from  $\beta$ -VAE model specified with KL weight  $\beta = 5$  and latent dimensionality of  $\ell = 3$ .



**Figure 29:** Parameter study, C.1, C.2, C.3: Frame-wise encodings. Resulting from  $\beta$ -VAE model ( $\beta = 5$ ,  $\ell = 3$ ). Samples for each latent variable in  $z$  (y-axis) plotted against frame indices (x-axis) of a single input image sequence comprising a CRF pattern for  $135^\circ / 315^\circ$  orientation. Background colors marking simulated baseline (white) and stimulation (yellow) phases.

### 3.2.6 Setting C.2: VSDI incl. Pre-Processing (Blank Subtraction)

#### Data Basis

In this setting, sequences from a real VSDI grating experiment (Subject ID: 092413) were pre-processed by the same pipeline as in setting B.2. This includes steps for baseline subtraction and 2-D spatial bandpass filtering. Key frames of an exemplary sequence for  $0^\circ$  grating direction are shown in Fig. 25, starting with the onset of stimulation at frame 150. Towards the end of the sequence, activations of a local region seem to stabilize in the image center which approximates the spatial extents of a cortical column.

#### Model Fit

Appropriate model fit is assessed for configurations with increasing latent dimensionality as well as KL weighting when comparing corresponding loss and validation loss curves (Tab. 5). Model configurations with high regularization ( $\beta \geq 25$ ) and low number of latent variables ( $\ell \leq 3$ ) show tendencies for model underfitting, while choices for medium latent dimensionality ( $10 \leq \ell \leq 25$ ) and small KL weights ( $\beta \leq 4$ ) indicate signs of overfitting the training data.

**Table 5:** Parameter study, Setting C.2: Assessment of model fit. ✓: good fit, /: ambiguous, O: model overfitting, U: model underfitting.

KL Weight	Latent Dimensions									
	1	2	3	4	5	10	25	50	75	100
1	/	✓	/	/	/	O	O	/	✓	✓
2	/	O	O	/	/	O	O	/	✓	✓
3	/	✓	✓	✓	/	O	O	✓	✓	✓
4	/	✓	✓	✓	✓	✓	O	/	/	/
5	/	/	✓	✓	✓	/	/	✓	✓	✓
10	/	/	/	✓	✓	✓	O	✓	✓	✓
25	U	/	/	/	✓	✓	✓	✓	✓	✓
50	U	U	U	/	✓	✓	✓	✓	✓	✓
75	U	U	/	/	/	✓	✓	✓	✓	✓
100	U	U	U	/	/	✓	✓	✓	✓	✓

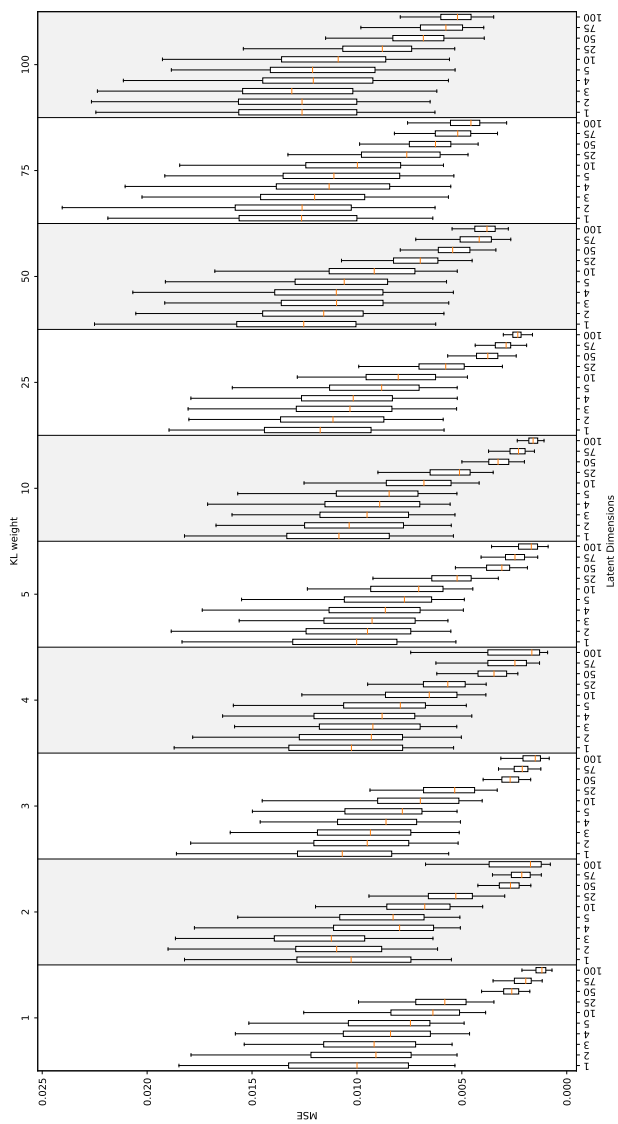
---

## Reconstruction Quality

Comparing the MSE boxplots for different model configurations (Fig. 30) reveals that the median MSE is decreasing for higher dimensionalities in  $\ell$ . This observation applies for all choices of KL weight  $\beta$ . For this setting, the strongest decrease in MSE distributions is recognized between  $\ell = 25$  and  $\ell = 50$ . With higher regularization weights, the variability of MSE distributions increases in terms of dispersion and spread of corresponding boxes. The first frames of each reconstructed image sequences shown in Fig. 27 are introduced with spatial activations forming an elongated branching structure which overlays the blood vessel network. Towards the end of the sequence, the activated region in the image center are getting reconstructed for model configurations with  $\ell \geq 3$ . Larger dimensionalities of latent space enable the model to reconstruct random fluctuations in the surrounding image areas more accurately.

## Feature Extraction

The latent space walk (Fig. 28) for the first variable shows two local activation centers at the upper left and lower right image for negative encoded values. Samples towards zero value exhibit global spatial activations with the exception of the image center, while positive values show a single activation region in the central image region. For the second latent dimension, negative values encode horizontal activations in the upper image area formed by two local components, which seem to merge when sampled values approximate zero. Positive values encode vertical activations formed by two centers in the upper right and lower right of the image region. The third variable covers a elongated branching pattern by negative encoding values, which is stretching diagonally from upper left to lower right and is overlaying the blood vessel network. Samples towards zero encode more a locally confined activation in the right image half, while positive values lead to reconstructions showing a single activation center at the top right image border. When passing a single validation sequence under  $0^\circ$  orientation stimulation as model input, corresponding encodings (Fig. 29) in the first latent dimension considerably diverge from prior mean of zero, as samples are constantly shifted towards a value of one. By contrast, samples for the second variable starting with values around one for the first couple of frames, while quickly returning and remaining close to zero. The third latent variable varies around its prior mean of zero, yet with its variance considerably decreasing for the last 20-30 frames.



**Figure 30:** Parameter study, setting C.2: MSE. For each parameter configuration between model capacity (in terms of latent dimensions) and regularization (in terms of KL weighting), a separate VAE model is specified. All frames of the validation partition comprising 1.680 frames are passed into the corresponding pre-trained model. The MSE between the current input image and its corresponding reconstruction is computed frame-wise. Subsequently, resulting distributions of MSE values are illustrated as boxplots for every model specification.

### 3.2.7 Setting C.3: VSDI incl. Pre-Processing (GLM)

#### Data Basis

The last data setting, pre-processing is applied in form of the GLM model described in Sect. 2.2.2 on domain of pixel time series followed by 2-D spatial bandpass filtering on frame-basis. The bandpass filter is again implemented via Difference-of-Gaussians of two low-pass 2-D Gaussian filters with respective radii of  $160 \mu\text{m}$  and  $960 \mu\text{m}$ . In contrast to baseline subtraction, the usage of GLM allows for preserving frames from the baseline. This ensures an identical frame count of 20.400 frames per recording block as in the raw data setting (C.1). Key frames of a sequence under  $0^\circ$  grating orientation is shown in Fig. 25.

#### Model Fit

Comparing loss and validation loss (Tab. 6) indicates similar results as for setting C.2. While choosing a specific latent dimensionality of  $\ell = 25$  with low KL weighting of  $\beta \leq 5$  can lead to overfitting of the training data, models with low number of latent variables ( $\ell \leq 4$ ) and high KL weighting ( $\beta \geq 50$ ) show signs of underfitting. All other specifications, which make up the majority of selected models, show signs of appropriate model fit.

**Table 6:** Parameter study, setting C.3: Assessment of model fit. ✓: good fit, /: ambiguous, O: model overfitting, U: model underfitting.

KL Weight	Latent Dimensions									
	1	2	3	4	5	10	25	50	75	100
1	/	/	✓	✓	✓	/	O	/	✓	✓
2	✓	✓	✓	✓	✓	✓	/	✓	✓	✓
3	✓	✓	✓	✓	✓	✓	O	✓	✓	✓
4	✓	✓	✓	✓	✓	✓	/	✓	✓	✓
5	✓	✓	✓	✓	✓	✓	/	✓	✓	✓
10	/	✓	✓	✓	✓	✓	✓	✓	✓	✓
25	U	/	✓	✓	✓	✓	✓	✓	✓	✓
50	U	U	U	/	/	✓	✓	✓	✓	✓
75	U	U	U	/	/	✓	✓	✓	✓	✓
100	U	U	U	U	/	/	✓	✓	✓	✓



---

## Reconstruction Quality

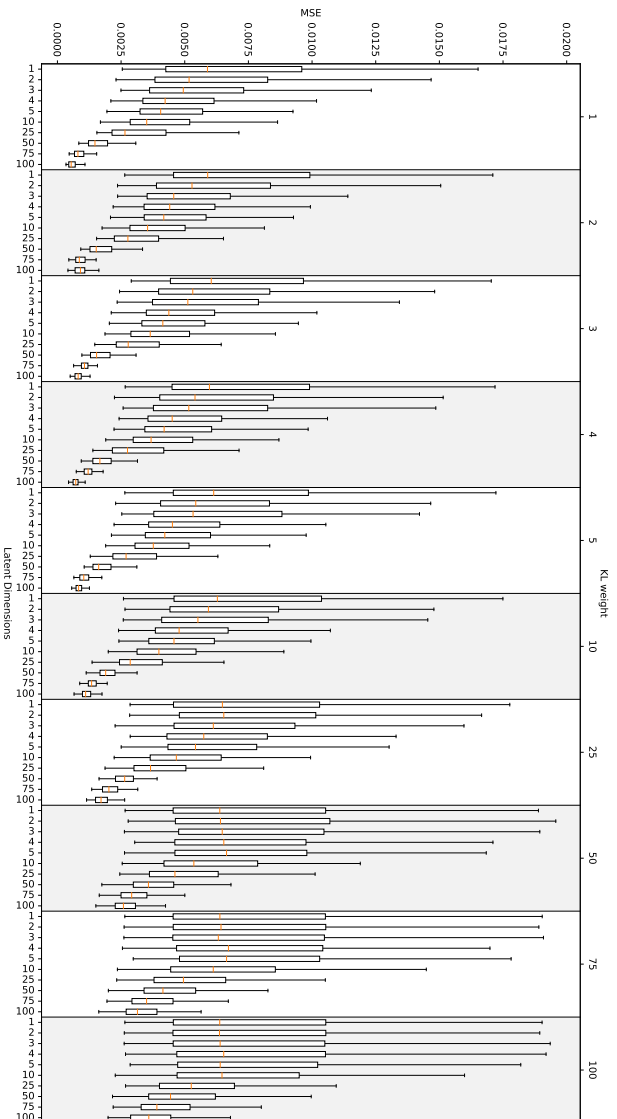
The MSE distributions for different configurations between model capacity and regularization (Fig. 31) show a decrease in median MSE as well variability in terms of box spread and dispersion with increasing latent dimensionality. While this holds true for specifications with low to medium regularization ( $\beta \leq 25$ ), it seems that this effect is mitigated when choosing higher values for the KL weight. For these models with  $\beta \geq 50$ , a decrease in MSE values is only seen when choosing a sufficiently high number of latent variables with  $\ell \geq 25$ . If this is not the case, no fundamental difference in MSE distributions can be found between specifications with low to medium number of variables ( $\ell \leq 10$ ).

Reconstructions of a single image sequence under stimulation with a  $0^\circ$  grating orientation show substantial differences between models with varying latent dimensionality while keeping the KL weight fixed at  $\beta = 1$ . Increasing  $\ell$  allows for more accurate predictions of non-stationary fluctuations. In turn, models with low to medium number of latent variables ( $\ell \leq 25$ ) show reconstructions of a branching spatial pattern diagonally spreading from the top left to bottom right of the image domain, overlaying a prominent blood vessel, which is also recognizable in the corresponding latent space walks (Fig. 28).

## Feature Extraction

By inspecting the latent space walks shown in Fig. 28, it becomes apparent that the first two dimensions encode similar activation regions, which are more locally confined in the image domain. Negative values of the first latent dimension cover a single activation center on top right image border, while positive values encode an activation center on the upper left border connected with an elongated structure overlaying the blood vessel. For these variables, both extracted activation patterns seem to switch rather abruptly in the latent space walk without an obvious transition, rather leaving the impression of bipolarity. Negative values for the third latent dimension result in local activations on the left border, as well as partially displayed at the bottom and top right of the image. Values around zero encode more global and spread activity on the upper image half. Positive values again cover the branching structure overlaying the blood vessel in connection with a local activation center in the central image region.

In contrast to setting C.1, encodings for baseline and stimulation frames seem to exhibit distinct properties when using a single sequence from current data basis with  $0^\circ$  grating stimulation as input for a  $\beta$ -VAE, latter being specified by  $\beta = 5$  and  $\ell = 3$ . While baseline encodings in the first dimension fluctuate around the prior mean of zero,



**Figure 31:** Parameter study, setting C.3: MSE. For each parameter configuration between model capacity (in terms of latent dimensions) and regularization (in terms of KL weighting), a separate VAE model is specified. All frames of the validation partition comprising 4.080 frames are passed into the corresponding pre-trained model. The MSE between the current input image and its corresponding reconstruction is computed frame-wise. Subsequently, resulting distributions of MSE values are illustrated as boxplots for every model specification.

---

the central tendency for frames during stimulus presentation seems to be shifted more towards values around  $-1$  while showing considerably higher variance. In the second latent variable, after the stimulation onset starting at frame 150, there are signs of a high short-term increase followed by an decaying phase towards values undershooting the baseline level. Towards the end of the input sequence, encodings are returning to values close to prior mean. For the third latent variable, a sharp increase in encoding values is seen which persists from the beginning of the stimulation until the end of the sequence.

### 3.3 Computational Performance

#### Hardware & Software Specifications

For this thesis, all custom code for data handling, processing and visualization as well as model building, training and evaluations was written and executed in Python 3.8 (Van Rossum and Drake, 2009). Here, multiple custom code repositories were generated and maintained by the author of this thesis. Their respective purpose, version and dependencies are listed in Tab. A.1 in Appendix A.

All code was executed on a conventional PC system having the following specifications: Microsoft Windows 10 Professional, 64-bit, Intel<sup>®</sup> Core<sup>™</sup> i7-3820 Quad-Core CPU @ 3.60GHz, 49152MHz DDR-3 RAM, 12 GB NVIDIA<sup>®</sup> GeForce<sup>®</sup> RTX 3060 GPU @ 1320MHz / Boost: 1777MHz.

#### Synthetic Data Generation

The generation of a single CRF of dimensionality  $[T, M, N]$  with  $T = 255$  frames, image height  $M = 128$  and width  $N = 64$  in pixel is relying on functions of the *GSTools* library (S. Müller, Schüler, Zech and Heße, 2021). Because associated kriging operations are computationally demanding, the Python module *Joblib* is used for creating several CRFs in parallel by distributing subtasks on multiple threads. This substantially decreases the overall duration for data generation, limiting the required computation time for a single image sequence to 31 s and for building the full dataset comprising 240 sequences in total to ca. 124 min.

In context of generating artificial VSDI components, *spatial* confounders related to artificial blood vessel networks and illumination are created only once as these are shared between all sequences, taking ca. 33.8 s of computation time. Sequence-specific components comprising *temporal* VSDI components (such as artificial dye bleaching and heartbeat) as well as random noise were generated nearly instantaneously in  $< 0.01$  s. Subsequent

---

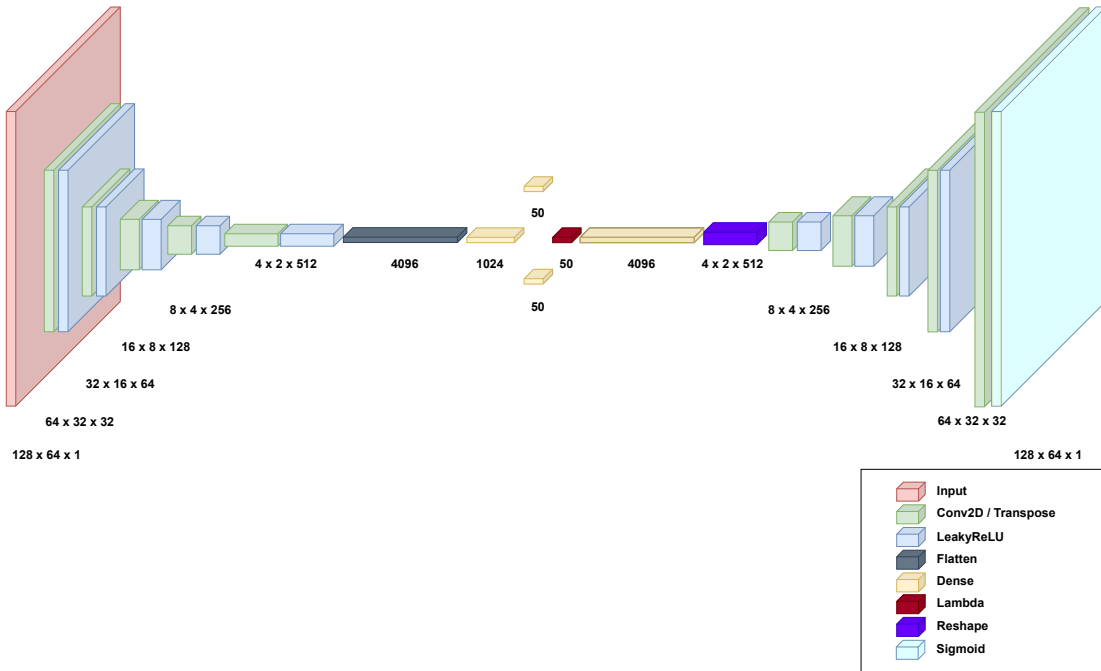
application of the linear composition model described in Sect. 2.4.8 for combining all VSDI-related components with a single CRF resembling the ground truth signal accordingly takes ca. 0.18 s per sequence.

### VAE Implementation

The  $\beta$ -VAE model described in Section 2.3.5 is implemented via the Python deep learning API *Keras* (Chollet, 2015). Its corresponding layer architecture is illustrated in Fig. 32. The encoder network compartment is built by stacking multiple 2-D convolutional neural network (CNN) layers, each coupled with a leaky rectified linear unit (LeakyReLU) activation layer. By this approach, the encoder is able to first capture local features within the input images before learning global correlations of the data when passed through hierarchically higher layers. The network depth is limited by two essential aspects: on the one hand by the dimensionality of input data, in this case an image size of  $128 \times 64$  pixel, which only allows for a maximum layer depth of seven layers to still be capable to capture the rectangularity of the input image ( $2 \times 1$  in the last layer); on the other hand, the rapid growth in trained parameter numbers which is accompanying an increase in network depth.

The specification of CNN kernel sizes has potential impact on both model accuracy and efficiency, as large kernels are suited to capture high-resolution patterns and small kernel sizes for capturing low-resolution patterns (Tan and Le, 2019). While popular deep convolutional network architectures like *Xception* (Chollet, 2016) and *MobileNetV2* (Sandler, Howard, Zhu, Zhmoginov and Chen, 2018) tend to use sequential stacks of smaller kernels ( $1 \times 1$ ,  $3 \times 3$ ), larger sizes such as  $5 \times 5$  and  $7 \times 7$  as well as combinations of different kernel sizes are reported to benefit in terms of model accuracy and efficiency (Tan and Le, 2019). For this purpose, a mix of different sizes ( $[7 \times 7, 7 \times 7, 5 \times 5, 3 \times 3, 3 \times 3]$ ) is used accordingly for the sequence of encoder CNN layers, and in reverse order for the decoder network. This approach is combined with the usage of small strides by two units of the convolution along spatial image dimensions. The number of output filters in the convolution is rather small, starting from 16 in the first layer and is then increasing by a factor of two at each convolutional layer in the encoding network. Vice versa, this holds for the deconvolution layers in the decoder network. Zero-padding is further used at every (de-)convolutional layer for controlling dimension shrinkage after applying kernels larger than  $1 \times 1$  and for avoiding information loss at the image borders.

For defining initial values for the parameters in the network prior to model training, a weight initialization method by He, Zhang, Ren and Sun (2015) is used, which is



**Figure 32:** Implemented VAE layer architecture. The layer terminology is following conventions of deep learning API Keras. For illustration purpose, the size of both input and output layers corresponds to an image size of 128 x 64 pixel. Multiple 2-D convolutional layers (Conv2D) with LeakyReLU activation are stacked together in the encoder network, latter specifying the inference model  $q_{\phi}(z|x)$ . It is followed by a separate fully-connected (commonly denoted as Dense) layer, taking the convolved feature maps as input. Two fully-connected layers provide the parameter vectors of the latent distribution  $p(z) \sim N(\mu, \sigma^2)$ : the first accounting for the set of means  $\mu$ , the second parametrizing the set of standard deviations  $\sigma$  of the Gaussian latent distributions. From each of those distributions, a point is randomly drawn in the Lambda layer via reparametrization trick  $z = \mu + \sigma \odot \epsilon$  with  $\epsilon \sim N(0, 1)$ . The vector of latent representations  $z$  is then fed into the subsequent decoding network specifying the generative model  $p_{\theta}(x|z)$ . The decoder is mirroring the encoder architecture by a stack of 2-D deconvolutional layers (Conv2DTranspose), finally reconstructing the input data  $x$  solely on basis of the given latent representations  $z$ .

---

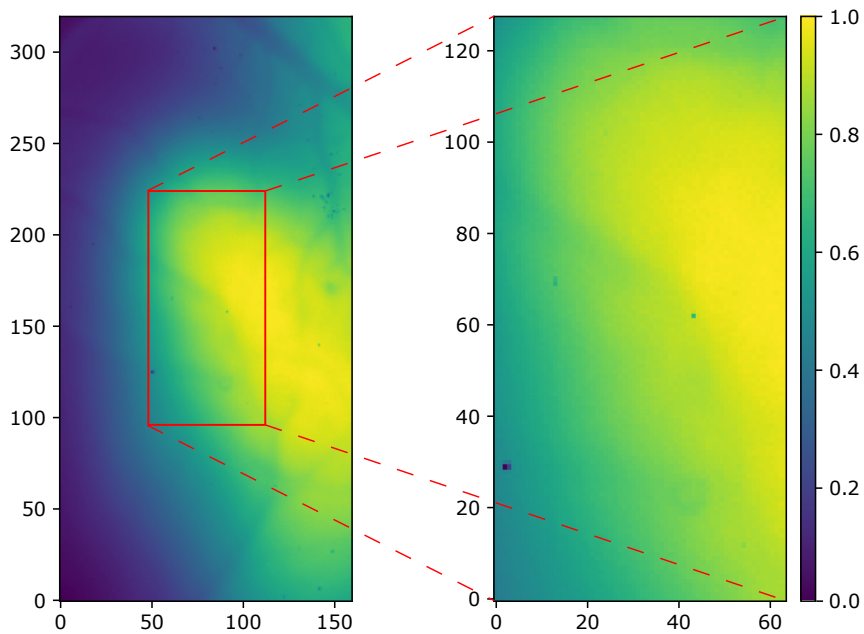
specifically designed for rectified activation units and calculated for the current weight  $w$  as random draw  $w \sim N(0, \sqrt{2/n})$ , with  $n$  denoting the number of input connections to the node.

Adaptive moment estimation, commonly known as *Adam* (Kingma and Ba, 2014), is used as gradient-based stochastic optimization method during model training. Here, individual adaptive learning rates for different parameters from estimates of the first and second moments of the gradients are computed. This method is computationally efficient with low memory requirements, is mostly invariant to rescaling of gradients, and scales well to large-scale high-dimensional data problems (Kingma and Ba, 2014). As suggested by Kingma and Ba (2014), the exponential decay rate for the first- and second-moment estimates are kept at their respective default values of 0.9 and 0.999. As higher learning rates lead to numerical issues during model training and subsequent NaN loss values (probably occurring due to half precision data format of FLOAT-16 to circumvent out-of-memory errors), the learning rate was kept at  $1e-4$ .

## Model Training

All image sequences were split into a training and testing partition with a fixed test proportion set to 20% of available data. When choosing a batch size of 255 frames (the equivalent of a full image sequence), a training epoch requires ca. 24 s to complete.

The overall duration of VAE training on synthetic datasets over the course of 100 epochs is ca. 40 min. By contrast, the original resolution of real VSDI images comprises a larger image dimensionality of  $320 \times 160$  pixels. As compability with all (de-)convolutional layers of the VAE has to be ensured in terms of recovering the correct rectangular image size of the encoder input in the decoder output, dimensions have to follow power-of-two numbers. Accordingly, images were cropped to the nearest compatible resolution of  $256 \times 128$  pixels. However, this still lead to crashes during neural network training due to memory leakage. To circumvent this problem, respective training data was passed sequence-wise from corresponding files into RAM by using a custom Keras *DataGenerator* class. While this approach fixed the memory leakage problem, this lead to heavier reliance on SSD read speed, in turn increasing computation time for neural network training by several orders of magnitude. A single training epoch of identical batch size of 255 frames now took ca. 180 s fo finish. Eventually, a smaller ROI of  $128 \times 64$  pixels indicated in Fig. 33 was chosen to limit computation times for all parameter studies to reasonable durations, improving comparability between the results for the different data types by ensuring identical data dimensionalities, as well as decreasing the impact of the illumination artefact on model learning. For a single grayscale image with height  $M = 128$  and width  $N = 64$  as input of the implemented VAE model, the encoder part takes ca. 14.19 ms for sampling the



**Figure 33:** Parameter study, settings C.1, C.2, C.3: ROI. From the VSDI sequences of a real experiment (subject: 092413) with image dimensions of  $320 \times 160$  pixel (*left*), a ROI of  $128 \times 64$  pixel was defined around the original image center.

latent encodings, while the decoder part is predicting the image reconstructions in ca. 13.09 ms. When instead using a full image sequence of dimensionality  $[T, M, N]$  with  $T = 255$  frames as input, the additional computational costs are drastically limited due to the amortized variational inference approach described in Sect. 2.3.5. Consequently, the encoder only requires ca. 87.87 ms and the decoder takes ca. 127.58 ms for processing all frames of the input sequence.

---

## 4 Discussion

---

### 4.1 Implications of Parameter Studies

The results of the parameter studies presented above (Sect. 3.2) substantially vary depending on the respective data basis and pre-processing approaches, as well as parameter choices for model capacity and regularization weight.

Benchmarking with known ground truth (setting A) shows promising results of the  $\beta$ -VAE for a broad variety of parameterizations, for which spatial features related to the pre-defined CRFs are found in the latent space walks and are also trackable on single-frame basis in the sampled encodings for respective variables. Furthermore, the reconstructions of input frames exhibit the intended regime-switching behaviour between baseline and activation phases. When increasing the model capacity in terms of dimensionality of the latent space, this improves the reconstruction quality with regard to a decrease in MSE values. This is mainly attributable to more precise reconstructions of random spatio-temporal fluctuations. Aside of the benchmark setting A, these results can also be recognized in settings incorporating any form of data pre-processing pipeline for improving the SNR (B.2, C.2, C.3).

Signs of model underfitting can be observed for comparatively few configurations, mainly when using very low numbers of latent variables in combination with strong model regularization. The model then seems to have too little capacity for capturing the high-dimensional image sequences, which is accompanied by inflexibilities of latent encodings to deviate from the prior  $p(z)$ . This is expressed by reconstructions only switching between a discrete number of archetypical spatial patterns, which are mostly unrelated to the actual input images.

Interestingly, signs of model overfitting are less observed for higher dimensionalities of the bottleneck z-layer. Instead, this behaviour is seen for more specific parameterizations of intermediate model capacity ( $\ell = [10, 25]$ ). This behaviour still occurs when incorporating any form of data pre-processing for both artificial as well as real image sequences with low SNR and prevalent artifacts (settings B.2, C.2, C.3). One possible explanation for this



---

observation is that with a high-dimensional latent space, the VAE architecture has the ability to sufficiently capture the underlying signal structure as well as noise components or other artifacts that might be present. This means that the model should be complex enough to generalize well to unseen data. However, while with an intermediate number of latent variables there might be enough capacity to capture the most dominant structures in the training data, at the same time it is not generalizing well because it is unable to additionally cover all sources of noise and artifacts. This in turn can lead to overfitting. Further information about the impact of model capacity on model fit specifically for VAE architectures is given by Dai and Wipf (2019), as well as for other deep generative models by Loaiza-Ganem, Ross, Cresswell and Caterini (2022).

Another possible explanation is that the model may only use a subset of latent dimensions to represent the data. By effectively discarding irrelevant dimensions during model training, this would pose an information bottleneck (Tishby, Pereira and Bialek, 2000; Tishby and Zaslavsky, 2015), which later would leave room for covering new features from unseen data. When instead having a latent space with too low dimensionality, the network might be forced to fully utilize all available dimensions, possibly leading to overfitting as well.

Following from these previous points, the optimization of the VAE's objective function might be relatively simple for low and high dimensionality of latent space, because both reconstruction and KL divergence term are more balanced for these parameter choices. However, for intermediate model capacities, the optimization could become harder, as the VAE's objective function might become more sensitive to the parameterization, leading to imbalances between reconstruction and KL term in the ELBO and in turn is impeding the optimization to find good minima of the objective function. As observed for datasets with sufficiently high SNR (settings A, B.2., C.2, C.3), an adequate model fit can often be achieved again by increasing the weight on the KL term. Specifically for setting B.2, too strict model regularizations by high choices of  $\beta$  is resulting in more ambiguous fit behaviour, which might benefit from a longer model training procedure by increasing the number of epochs.

In the context of model fit, setting B.1 using synthetic VSDI sequences without application of pre-processing techniques is an exceptional case. Here, a peculiar relationship between model capacity and regularization weight becomes apparent: with increasing number of latent variables and simultaneously lower weight of the KL term, signs of model underfitting can be recognized more frequently. This poses a counter-intuitive finding at first glance, since usually with higher model capacity the risk of model overfitting tends to increase (Goodfellow, Bengio and Courville, 2016). One possible explanation would be the lack of diversity of data. In case of low variability, the model may not be able to

---

learn different features or variation in the data, leading to underfitting. In the particular setting B.1, over-dominant artifacts, especially the illumination component, are masking the underlying ground-truth signal related to the CRF patterns, ultimately depressing inter-frame variability.

Overall, further inspections of the learned features and representations of the model might be necessary for fully understanding the reasons for model under-/overfitting specifically in terms of the data settings at hand. This could lead to more informed hyperparameter tuning such as the dimensionality of the VAE bottleneck layer and regularization weights. In this regard, well-known optimization techniques such as cross-validation (Hastie, Tibshirani and Friedman, 2009) or Bayesian optimization (Snoek, Larochelle and Adams, 2012) might also improve model complexity parameter  $\ell$  and  $\beta$  in future applications.

## 4.2 Limitations & Extensibility

### 4.2.1 Stimulation Paradigm

For this thesis, only sequences under stimulation via drifting full-field gratings were analyzed. It is therefore important to consider possible biases of neural activation with a specific response behaviour for selected properties of this particular visual pattern, which may substantially differ from natural scenes. In this context, Akasaki, Sato, Yoshimura, Ozeki and Shimegi (2002) investigated the impact of the receptive field surrounding on neural activity in V1 of anaesthetized cats. The study used extracellular recordings to measure the activity of V1 neurons in response to sinusoidal grating stimuli presented in the center and surrounding of the receptive field. The results showed that the activity of V1 neurons was suppressed when stimuli were presented in the surrounding of the receptive field. The study also found that this suppressive effect was strongest when stimulus properties such as orientation-, direction-contrast or relative spatial phase difference were similar in the center and surrounding. Vice versa, this modulatory effect was weaker for stimuli that were dissimilar. The authors concluded that the surround plays a critical role in shaping the activity of V1 neurons, and that this role is determined by the similarity between the center and surround stimuli (Akasaki et al., 2002).

Random dot patterns such as velocity- or direction-change-dots pose an common alternative for visual stimulation. By the usage of moving point clouds, activity is primarily evoked from neuron populations that are specifically selective for velocity and direction of motion, as a point has no orientation information. Yet, response amplitudes for dot patterns may be substantially smaller when compared with stimulation through gratings

---

(Hofmann, 2020; Peter, 2019).

Another problematic aspect of the chosen stimulation paradigm is caused by an over-representation of baseline frames. As each recorded VSDI sequence is introduced by a substantially longer baseline phase of 1000 ms than the following 700 ms of stimulus presentation, any subsequent data analysis or modeling approach is possibly skewed by this imbalanced frame selection from both regimes. Resampling techniques like random undersampling (Liu, Wu and Zhou, 2009; Mishra, 2017) can be an effective method for addressing this issue, as this would equalize the number of frames from both unstimulated and stimulated recording phases. This allows for reducing the risk for a skewed training of deep learning models towards the majority of data, in this case the baseline activity. It is also accompanied by computational benefits in terms of less memory and storage requirements due to the reduced number of datapoints, and consequently faster model training. It is yet important to underline that a considerable amount of information may be lost in the undersampling approach. Latter aspect could hinder research projects specifically addressing dynamics in this particular recording phase, such as ongoing activity in absence of any stimulation.

#### **4.2.2 Data Acquisition**

Despite aforementioned benefits of using voltage-sensitive dyes for optical imaging (Sect. 2.1.1), several problems complicate a reliable extraction of response-related features from recorded image sequences.

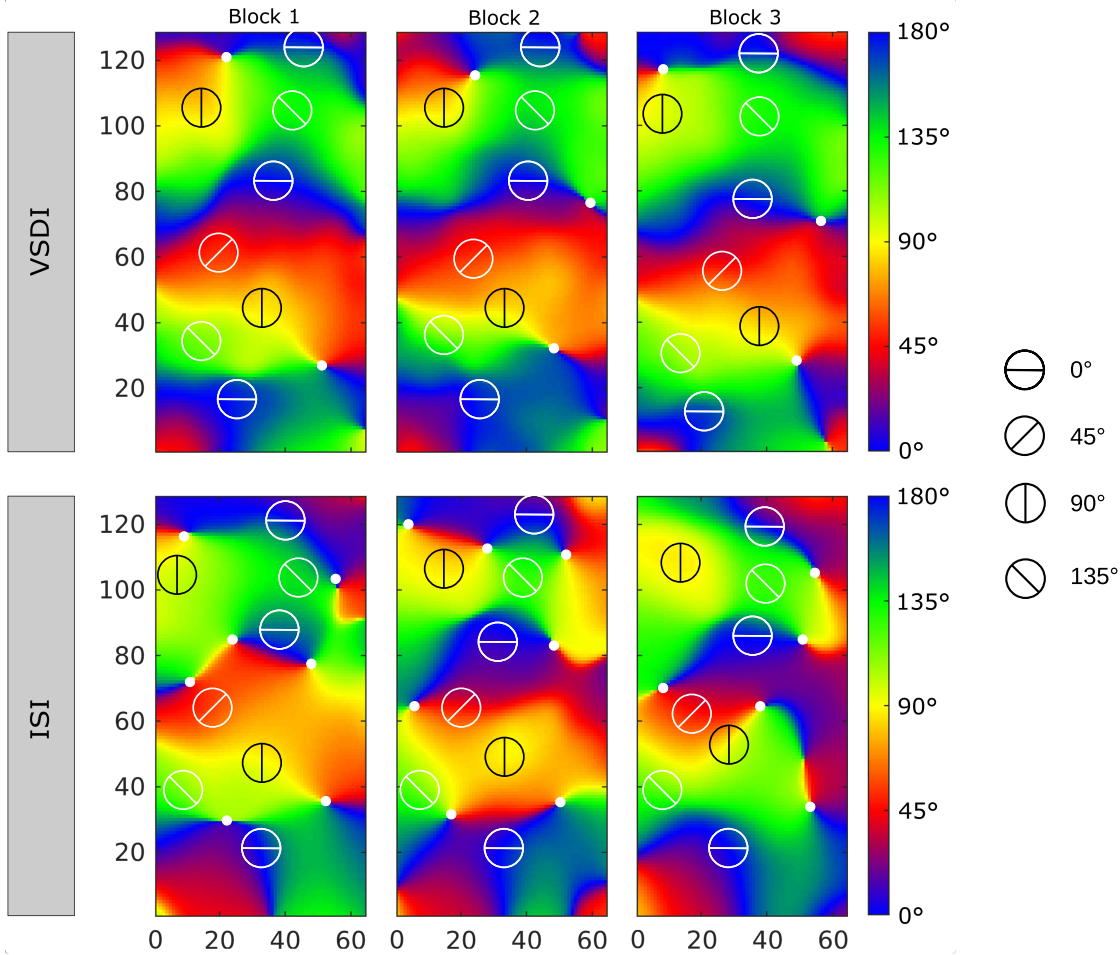
On the one hand, the fluorescence signal represents activity composed as an aggregate of different signal sources. In this context, other biological structures (e.g. ocular-dominance columns, cytochrome oxidase blobs) may be involved in or even superimpose the evoked response (Bartfeld and Grinvald, 1992). In case of real VSDI experimental data, this is still evident after application of pre-processing steps (settings C.2, C.3), as the shifts in amplitude during the stimulus regime can overlap with blood vessels on local spatial domain. In theory, no interference between the neural response and the hemodynamic signal should exist, as the excitation wavelength of the used dye RH-1691 overlaps only marginally with that of hemoglobin, as previously stated in Sec. 2.1.1. This observation is not clearly attributable, and this may be related either to neurovascular coupling, as the local blood flow is directly correlated with local energy consumption in the brain matter (Klein, Kuschinsky, Schrock and Vetterlein, 1986; Liao et al., 2013), or changes of the camera focal plane in relation to the heartbeat (Hofmann, 2020).

On the other hand, cortical response dynamics are highly variable (for a review, see Singer,

---

2013). This is expressed by increased trial-to-trial variability of cortical response, which to this day is still poorly understood. Possible explanations account this inter alia to either pre-stimulus oscillations (Fries, 2005) or ongoing spontaneous hemodynamic fluctuations (Saka, Berwick and Jones, 2012).

To ensure that the VSDI setup in use is indeed feasible of recording neural activity, cortical maps in V1 are compared between the used VSDI setup with those resulting from another measurement technique for the identical experimental subject (ID: 092413). In this case, intrinsic signal imaging (ISI) is chosen, as it rather relies on changes in hemoglobin concentration and oxygenation as reporter for neuronal activity instead of changes in membrane potential indicated by fluorescent dyes in VSDI (Grinvald et al., 1986). This validation approach is, therefore, similar to that of Shoham et al. (1999). When assessing orientation preference maps for both VSDI and ISI, as illustrated in Fig. 34, two aspects become particularly apparent: on the one hand, similar spatial structures can be seen when comparing maps of both acquisition methods. Here, patches indicating certain orientation preferences as well as pinwheels share similar locations to a high extent, thus indicating validity of VSDI measurements. On the other hand, a substantial level of variability is recognizable for both recording techniques, as spatial shape and connectedness of individual patches can slightly differ between recording blocks. Aside of neural variability of the response signal, another more technical explanation would consider the (re-)staining of the fluorescent dye applied prior to each recording block.



**Figure 34:** Orientation preference maps obtained from VSDI & ISI. For both data acquisition methods, recordings of the identical experimental subject (ID: 092413) are used. Response maps are calculated for contralateral moving grating patterns in each of the four main orientations ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ). Subsequently, response maps are averaged and spatially bandpass-filtered. For each major orientation, response amplitudes are vector-summed per pixel and their corresponding angles (in degrees) are color-coded.

---

### 4.2.3 Data Simulation

In this thesis, an approach for generating synthetic VSDI data on basis of single-trial recordings is presented. Several assumptions were made regarding the included noise components and their composition, which will be explained in more detail below.

So far, every synthetic image sequence comprises individual dye bleaching dynamics, as parameters of the related double exponential model are re-drawn per sequence. This implicitly assumes independency of the bleaching component across sequences. Yet for real VSDI recordings, bleaching kinetics endure over the course of light exposure in the range of minutes (Takagaki et al., 2008) and therefore comprise multiple consecutive sequences, which rather implies a shared temporal dependency structure. Therefore, alternative model specifications would either consider a global parameterization of the double exponential for multiple sequences, or introduce a dedicated model for the temporal parameter variations over consecutive sequences. The latter requires a deeper understanding of dye-related chemical properties, which is not within the scope of this thesis.

Heartbeat-related dynamics are modeled entirely in the temporal domain as originally introduced by Reynaud et al. (2011), while structures related to blood vessel networks are specified only in spatial domain. This ignores possible confounding signals related to heartbeat and/or cerebral blood flow. While related shifts in the absorption spectrum of hemoglobin is reportedly less problematic for the used “blue” dye RH-1691 (Lippert et al., 2007), artifacts due to physical pulsatory movements of blood vessels may still be contained in the recordings (Ferezou, Mátyás and Petersen, 2009). Also, interaction effects due to neural-hemodynamic or neurovascular coupling (Liao et al., 2013; Martindale et al., 2005) may be convolved in the fluorescent signal and are not yet accounted for in the current approach.

The generation of synthetic single-trial data is based on 3-D CRFs, which are interpreted as the fundamental spatio-temporal signal. These fields are generated by constraining the respective kriging approach via user-defined key locations and values, allowing for the precise control of extent and timing of corresponding response patterns. As artificial orientation maps (Macke et al., 2009, 2011) are introduced as prior information, pre-knowledge about spatial properties of columnar responses in V1 to oriented stimuli is already accounted for in the present approach.

The associated kriging approach is executed solely on whole-pattern domain, as it is assumed that there exists strong response synchrony of cells with similar feature selectivity (Brosch, Bauer and Eckhorn, 1995; Eckhorn et al., 1988; Ts'o et al., 1986) due to

---

horizontal connections between cortical domains with similar tuning properties (Gilbert and Wiesel, 1989; Malach, Amir, Harel and Grinvald, 1993). This in turn ignores temporal phase shifts and latencies between individual activity centers, which per se might represent a neural encoding mechanism (Fries et al., 2001).

Interactions between visual areas are still subject of current research, which the data synthesis approach is not yet accounting for. On the one hand, these can be expressed through synchronized activity (Engel, Kreiter, König and Singer, 1991; Ghisovan, Nemri, Shumikhina and Molotchnikoff, 2008), which might be modulated by attention-selective mechanisms associated with synchronization of oscillatory responses specifically within Gamma frequency band (Fries, 2005). On the other hand, interactions can exhibit structures of travelling waves (Cowey, 1964; Ermentrout and Kleinfeld, 2001), whose substrate may lie in long-range horizontal connections (Sato, Nauhaus and Carandini, 2012). When presenting full-field contrast reversal gratings, waves are propagating from area 18 to 17 in cat's visual cortex with systematic phase shifts, while being independent of stimulus orientation. However, in absence of stimulation they can travel in both directions between both areas (Zheng and Yao, 2012). These findings suggest that global waves have a more general function of integrating information over large regions of space instead of encoding for specific visual features (Sato et al., 2012).

Therefore, future insights about cortical interactions should be incorporated in the synthetic data-generating model for VSDI data. For example, different shapes of spatio-temporal wave patterns observed in mesoscale optical imaging, such as simple planar waves to more complex source-sink, spiral-in or saddle patterns (Afrashteh, Inayat, Mohsenvand and Mohajerani, 2017; Townsend and Gong, 2018; Townsend et al., 2015), can be introduced in terms of velocity vector fields as additional spatio-temporal components interacting with the already established response-related model components.

The parametrization of each model term introduced in the linear signal composition (Sect. 2.4.8) is based on averaged parameter estimations from sequences of a single VSDI experiment. While this provides an approximation of realistic weightings between individual signal- and noise-related components, taking into account parameter information from further experimental subjects should improve the generalizability of the data-generating model. However, due to the invasiveness of the measurement procedure, data availability is severely limited for VSDI. Even in the hypothetical case of additional data acquisition, rather strict requirements have to be considered to ensure comparability. Latter aspect is referring, among others, to the usage of an identical camera setup, dye structure and stimulation paradigm, as well as comparable dye staining quality.



---

#### 4.2.4 Data Pre-Processing

When comparing both applied data pre-processing approaches, namely baseline subtraction (setting C.2) and GLM (setting C.3), similar spatial locations of response activity are emphasized for real VSDI sequences. As verified for the synthetic data basis in setting B.2, subtracting the baseline can substantially improve the SNR and alleviate the identification of pre-defined CRF patterns from noise-contaminated image sequences.

Nevertheless, it is important to note that both approaches come with specific assumptions and drawbacks, which make them more or less suitable for different research questions and applications. Baseline subtraction draws its main advantages from removing constant background fluorescence from the image without posing strong statistical assumptions on signal- or noise-related dynamics. Additionally, it can correct for inhomogeneities in the image, such as uneven illumination or dye bleaching behaviour. As it does not pose strong statistical assumptions about neither the data-generating process nor individual components related to VSDI signal or noise components, it is readily available for ad-hoc data pre-processing, which predestines it especially for the use of online closed-loop experiments. However, this method implicitly assumes that baseline fluorescence of both the blank and stimulus phases are identical (Raguette et al., 2016). If the baseline is not chosen carefully, the application of baseline subtraction (or division) can reduce or even obscure the amplitude of the neural response. This results from the low amplitude of neuronal activity, which is only in the order of a thousandth compared to the baseline fluorescence level (Grinvald et al., 1999). Furthermore, as it is applied on global image domain by default, it cannot take into account for different baseline fluorescence levels in distinct image regions, potentially introducing additional errors and biases.

When instead using the GLM, ideally only random white noise should be left in case that all deterministic elements from the recorded VSDI signal are considered in the specification of the denoising model. In this case, the residuals can be attributed to both, spontaneous neural activity and other non-physiological sources of randomness (Reynaud et al., 2011). Yet, when the requirement of a priori knowledge of all relevant components of the VSDI signal is not met, the GLM will suffer from biased parameter estimates due to omitted variable bias, which in turn leads to false conclusions about stable neural response dynamics (Stevenson, 2018). Since the shape of each regressor has to be defined a priori (Chemla et al., 2017), this opens up dangers of misspecification which, especially in the case of dominant noise components, may ultimately lead to false interpretation of the neural signal. General model misspecification has also to be considered, as in the original GLM formulation neither spatial components nor spatio-temporal patterns (such



---

as cortical oscillations) are introduced as regressors yet. Furthermore, the application of the GLM is based on typical assumptions for linear modeling, concerning linear independence and additivity of regressors as well as homoscedasticity and Gaussianity of residuals (Greene, 2008), which may not hold for real-world scenarios.

A more recent approach for denoising VSDI sequences is described by Carmi et al. (2021). Here, extracting and locating responses from VSDI recordings in rat's primary visual cortex is achieved by a combination of the GLM model by Reynaud et al. (2011) with Temporally Structured Component Analysis (TSCA), latter initially proposed by Blumenfeld (2010). Evaluations and comparisons were carried out for seven different methods: frame averaging, multi-parametric thresholding system (Gross, Ivzan, Farah and Mandel, 2019), maximal cross-correlation delay (Polack and Contreras, 2012), correlation to delayed theoretical response, TSCA, GLM, as well as the combination of GLM and TSCA. Performance was assessed via cluster separation metrics, namely Silhouette Index and Davies-Bouldin Index, and further validations were done through simulated data. In terms of both metrics, the combination of GLM with subsequent TSCA outperformed the other compared approaches in terms of locating cortical responses and the generation of retinotopic maps (Carmi et al., 2021).

#### **4.2.5 Data Analysis / Modeling**

With respect to the performed parameter studies, the choice for a VAE model architecture (Sect. 2.3.5) turned out to be advantageous in several aspects.

Across all data settings, a high quality of input reconstructions can be observed in terms of MSE. This is already achievable while using low numbers of training epochs ( $\leq 100$  epochs) and low amount of training data (in this case 20.400 images  $\hat{=}$  80 sequences per recording block), which in the majority of model configurations seem to be sufficient for good model fitting properties. Also, the commonly known problem of VAE models concerning blurry output images cannot be observed for the reconstructions in any data setting.

From the parameter studies, it can also be deduced that the VAE is agnostic to the input data, as no strong statistical assumptions need to be made towards the data-generating process, which implies that it can be trained on different datasets without any adaptation of the model architecture.

Further positive aspects of the VAE cover not only its robustness against spatial non-linearities or non-stationarities, but also its ability to express variabilities of image features within the latent encodings. The latter aspect in particular poses a fundamental prerequisite when processing image data acquired through VSDI.

---

Lastly, the VAE offers computational efficiency when using pre-trained models for latent feature extraction and input reconstructions, which poses a strict requirement of future closed-loop experiments.

Despite aforementioned arguments in favor of the VAE, several limitations and drawbacks have to be noted, which require further extensions of the model architecture. The clearest disadvantage of the current VAE implementation is the neglect of temporal dependencies in the image sequences. Each model input (in this case 2-D frames) is processed independently, as each dynamic sequence is treated as static, and the VAE is applied under the assumption of  $p_\theta(x_{1:T}) = \prod_{t=1}^T p(x_t)$ . This indeed can lead to adequate model fitting for single observations, which is also seen in the aforementioned parameter studies in terms of individual frame reconstructions quality. However, this also implies that the model is not capable of accounting for temporal structures within the input sequences yet, as it is implicitly assumed that  $p_\theta(z_{1:T}) = \prod_{t=1}^T p(z_t)$ . To achieve this, it would be necessary to introduce temporal dependencies in the latent states  $z_{1:T}^i$  with  $i = 1, \dots, N$  sequences and  $t = 1, \dots, T$  frames by defining the prior for the latent variables of a sequence as  $p_\theta(z_{1:T})$  (Fraccaro, 2018).

In this regard, sequential extensions of the VAE with either dynamic Bayesian networks such as state-space models for stochastic dynamics (R. Krishnan, Shalit and Sontag, 2017), or recurrent neural networks for deterministic dynamics (Bayer and Osendorfer, 2014) are subject of current research (for a review, see Girin et al., 2021). This includes network architectures such as variational recurrent neural networks (Chung et al., 2015), deep Kalman filters (R. G. Krishnan, Shalit and Sontag, 2015), stochastic recurrent neural networks (Goyal, Sordoni, Côté, Ke and Bengio, 2017), Kalman variational autoencoders (Fraccaro, Kamronn, Paquet and Winther, 2017), recurrent variational autoencoders (Leglaive, Alameda-Pineda, Girin and Horaud, 2020) and dynamical variational autoencoders (Girin et al., 2021).

As previously stated, all image data was collected only from a single experimental subject (ID: 092413). The inclusion of data from additional subjects would require changes to the training regime in terms of the composition of training and testing data partitions, resulting in either (i) joint training on all available image datasets across subjects, or (ii) separate model trainings per individual subject. In case of (i), the VAE will most likely focus on image components explaining the most data variance between subjects, more specifically artifact-related characteristics such as individual blood vessel patterns. These features are mostly undesired for the present research question at hand, and necessitates a substantial increase in latent variables to further cover features related to neural dynamics, in turn leading to aggravated interpretability of latent space. By contrast, (ii) will require

---

a unification of subject-specific latent spaces to corresponding image components observed across all subjects, which per se poses a complex assignment problem.

Alternative deep learning frameworks for the purpose of generative modeling have recently emerged, each holding its respective advantages and drawbacks when compared to VAE. This includes among others:

- generative adversarial networks (GAN) (Goodfellow et al., 2014),
- flow-based models like normalizing flows (Dinh, Krueger and Bengio, 2014; Rezende and Mohamed, 2015),
- diffusion models (Ho, Jain and Abbeel, 2020; Sohl-Dickstein, Weiss, Maheswaranathan and Ganguli, 2015; Song and Ermon, 2019).

GANs can offer improved image quality compared to VAE, but suffer from the lack of an explicit representation of the generative data distribution, as well as difficult model training due to possible mode collapse (also known as Helvetica scenario), non-convergence to a Nash equilibrium and instability leading to the problem of vanishing gradients, and overfitting caused by imbalances between generator and discriminator (Kossale, Airaj and Darouichi, 2022).

Normalizing flows are able to optimize the exact data log-likelihood  $\log p(x)$  and infer exact values for each latent variable  $z$  by transforming distributions through a series of invertible parameterized functions; this comes while offering constant gradient computations scaling respective to depth, and only requires an encoder to be learned (Kingma and Dhariwal, 2018). Yet, it requires the invertibility and efficiency of computing the determinant of the Jacobian, and usually provides lower quality of generative results when compared to VAE and GAN (Bond-Taylor, Leach, Long and Willcocks, 2021).

Diffusion models rely on two stages comprising (i) a forward diffusion stage gradually introducing Gaussian noise to the input data in multiple steps, as well as (ii) a parametrized reverse (or backward) process stage utilizing a generative model to reconstruct the original input data from the diffused (noisy) data by training the generative model to gradually reverse the diffusion process. While these models produce state-of-the-art image quality and offer advantages such as tractability and a stationary training objective, generating samples can still be computationally expensive due to the reliance on a long Markov chain of diffusion steps, although recent methods have been proposed to improve this process (Croitoru, Hondru, Ionescu and Shah, 2022).

---

## 4.2.6 Model Evaluation

The current approach for evaluating the VAE model results so far comprises the assessment of the following aspects:

- model fit in terms of comparisons between loss and validation loss, as well as the KL loss function;
- image reconstruction quality in terms of the distance between model input and output images measured via MSE, which acts as reconstruction term in the VAE objective function;
- latent feature extraction in terms of (i) frame-specific sampled encoding values for assessing possible regime-switching behaviour within the input data and (ii) variable-specific latent space walks for visualizing the individual contribution of each latent dimension on ; both are inspected while considering the dependence on the size of latent space as well as the balance between reconstruction quality and model regularization.

Regarding the image quality of model reconstructions, the MSE has beneficial properties, such as simplicity and computational inexpensiveness. It further satisfies common requirements of optimization, such as non-negativity, convexity, symmetry and differentiability. Yet, it implicitly assumes independence from temporal and spatial relationships and equal importance of of samples, which can be problematic in certain fields of applications (Wang and Bovik, 2009). Aside from the MSE, a plethora of alternative statistical measures for image comparison exist, such as peak-SNR, contrast-to-noise ratio, mean structure similarity index, correlation coefficient as well as correlation parameter (Salinas and Fernandez, 2007; Wang et al., 2004), which were already compared on the basis of VSDI data (Carmi et al., 2021).

So far, inspections of the extracted features are carried out only on the basis of encodings for single latent variables, which in turn ignores the inter-relationship between dimensions. A simple extension could be made by applying t-distributed stochastic neighbor embedding(t-SNE) (van der Maaten and Hinton, 2008) for mapping the encodings of all latent variables into a low-dimensional target space, latter usually comprising two or three dimensions. The t-SNE approach is well-known as visualization tool for high-dimensional data by using a variation of stochastic neighbor embedding (Hinton and Roweis, 2002).

---

## 4.3 Related Work

The first essential aspect of this work lies in the generation of synthetic VSDI image sequences on single-trial basis. Alternative simulation approaches with scope of VSDI are either focused exclusively on the spatial domain in terms of generating artificial orientation maps (Macke et al., 2009, 2011), or solely on the temporal domain in terms of the composition of signal dynamics on individual pixel level (Chemla et al., 2017; Reynaud et al., 2011).

To date, however, there is no methodology explicitly targeting data simulations of single-trial image sequences of spatio-temporal VSDI dynamics on mesoscopic level. Flotho et al. (2019) describe a broadly related approach to the present data simulation model. In the aforementioned publication, a semi-synthetic multispectral VSDI recording from somatosensory cortex of adult Sprague-Dawley rats was developed. Here, the focus lies on benchmarking several compensation strategies of motion artefacts, the latter being related to adsorption and reflection pulsation, as well as various physiological and ambient movements. A dummy VSDI response was introduced as checkerboard pattern, which was temporally varied by weighting it by a multiplicative and spatially constant function. As background texture, a notch-filtered image of the cortical surface illuminated under red light was used, which comprised surface texture details as well as blurred vessel structures. Additional terms for the pulsation as well as white noise related to the sensor and imaging system were further considered in respective model (Flotho et al., 2019).

The second substantial aspect of this thesis deals with unsupervised feature extraction from VSDI recordings, which is achieved by application of the VAE as a deep generative model. Deep learning techniques in general have been used extensively in a wide range of related fields such as medical image analysis, for instance for image classification and segmentation tasks for cancer diagnostics (as reviewed by Tandel et al. (2019)) or delineating pathological brain regions in MRI studies for diagnostics of Alzheimer's disease (Yamanakkanavar, Choi and Lee, 2020). Furthermore, they have also been used in recognition tasks for predicting diverse cell and tissue structures in fluorescence images (D. Schmidt, Rausch and Schanze, 2020), as well as for the segmentation of neurons from two-photon calcium imaging recordings (Soltanian-Zadeh, Sahingur, Blau, Gong and Farsiu, 2019).

The use of deep learning models has increased exponentially particularly in the case of medical imaging. Extensive developments were also made in brain image analysis for the purpose of diagnostic and classification of strokes, psychiatric disorders, epilepsy, neurodegenerative disorders, and demyelinating diseases (Ravi et al., 2017; Zhu et al., 2019). In this regard, generative modeling via VAE were carried out, inter alia, for brain

---

lesion detection (Akrami, Joshi, Li, Aydore and Leahy, 2020) and anomaly detection (Chatterjee et al., 2022) on basis of MRI data. However, the general lack of deep learning applications specifically targeting VSDI recordings is still apparent to this day.

Alternative approaches for source separation of in vivo VSDI recordings often extend the GLM model by Reynaud et al. (2011). In this regard, Yavuz (2012) first separates temporal components using the GLM into two different groups of components, where the first group holds all artefacts except bleaching, while the second includes neural response and residuals. Spatial PCA is then performed for both groups, as well as on artefacts found during blank recordings. The extracted components are then compared to the blank components by assessing correlation between their respective (temporal) coefficients. When showing low correlations with the blank, artifact components are re-classified as neural activity components, and vice versa.

Raguet et al. (2016) on the other hand use a set of convex non-smooth regularization priors adapted to the morphology of the sources and artifacts to extend the GLM for VSDI recordings from somatosensory cortex in mice under a whisker stimulation paradigm.

Lastly, Afrashteh et al. (2017) use optical-flow analyses for the quantification of spatiotemporal dynamics of mesoscale brain activity from VSDI recordings in mice under sensory forelimb and auditory stimulation. Analyses were carried out after data pre-processing via temporal filtering by finite-impulse-response low-pass filter and subsequent spatial Gaussian filtering.

## 4.4 Conclusion & Outlook

The fundamental objective of this work is the development of a processing pipeline with respect to VSDI image sequences. Here, the focus lies on unsupervised feature extraction in the context of spatio-temporal response activity of neural orientation columns in cat's primary visual cortex.

For this task, the balance between two opposing constraints has to be established: on the one hand, temporal efficiency of the processing algorithms despite high data dimensionality to ensure compatibility with future closed-loop experimental designs; on the other hand, robustness of the methods especially with regard to substantial variability of signal- and noise-related components, as well as low SNR prevalent in VSDI recordings. To this end, a VAE architecture poses an attractive solution combining advantages of both deep learning and probabilistic modeling, and offers a compromise in both aforementioned constraints.


---

Since neither ground truth information nor data labeling are available for real VSDI recordings, model training and evaluation was extended to synthetic image sequences for validating and benchmarking the implemented  $\beta$ -VAE model. Here, prior knowledge about real disturbance terms and artifacts typically observed in VSDI are introduced in the data-generating process via temporal and spatial components. The overall signal composition was achieved in terms of a weighted linear model, by which conditioned random fields were used as hypothetical spatio-temporal activation dynamics with similar properties as orientation columns in V1 responding to grating stimuli.

The results of  $\beta$ -VAE training are fundamentally depending on the balance between input reconstruction quality and model regularization. Extensive parameter studies on both synthetic as well as real VSDI sequences were carried out accordingly for finding proper parameterizations of latent dimensionality, the latter being directly impacting the quality of model output images, as well as the weighting of the regularization term within the VAE loss function. The application of any data pre-processing poses an indispensable requirement of VAE model training for the purpose of improving the SNR and reducing the influence of dominant artifacts. Depending on the choice of pre-processing methods, the VAE training shows visible ranges and border areas of parameter configurations in terms of proper model fit, MSE values as well as distributions of latent encodings of prevalent signal- and artifact-related components.

Future extensions of the synthetic data-generating model could cover new insights about dye-related dynamics, interaction effects (e.g. due to neurovascular coupling), as well as temporal phase shifts and latencies between individual activity centers. Further regressors could be introduced with regards to interactions between different brain areas, which might be expressed in the form of synchronized oscillations or travelling waves. Extension to the VAE model, especially on temporal domain via state-space models or recurrent neural networks, should be considered in future research.

Integrating the implemented pipeline into a closed-loop interactive approach will allow for investigating interactions between sensory-evoked and spontaneously emerging activity states as well as the origin of response variability in activity patterns caused by visual stimuli (e.g., a moving grating pattern). By conditioning the stimulation on the spontaneous occurrence of extracted patterns and assessing their impact on evoked activity states, it is possible to test predictions of a subtractive predictive coding framework (Aitchison and Lengyel, 2017). This approach can lead to a better understanding of neural coding strategies as well as the interdependence of complex internal states and responses



to sensory input. Moreover, these strategies enables robust reconstruction of functional networks, their real-time tracking via sequential models as well as subsequent optimal experimental intervention.



## Appendix A - Code Repository Overview

**Table A.1:** Overview of custom Python repositories. Implemented, tested and executed in Python (v3.8). Repositories will be made available after publication at <https://www.github.com/dekili/>.

Repository	Version	Purpose of Repository	Dependencies	Dependency Version
kh_tools	0.5.0	General Data Handling VSDI (Pre-)Processing Data Visualization	Matplotlib	3.4.3
			NumPy	1.19.5
			pathlib	1.0.1
			SciPy	1.7.1
simSRF	0.1.0	Generation of SRF / CRF	emcee	3.0.2
			gp-maps-python	0.0.3
			GSTools	1.3.3
			Joblib	1.1.0
			Matplotlib	3.4.3
			Numpy	1.19.5
			OpenCV	4.5.3.56
PyVista	0.36.1			
simVSDI	0.1.0	Generation of Synthetic VSDI Artifacts Composition of Synthetic VSDI Sequences	Matplotlib	3.4.3
			NumPy	1.19.5
			OpenCV	4.5.3.56
			Pandas	1.3.2
			scikit-learn	0.24.2
GISMO	0.5.0	DNN Implementation DNN Compiling Model Training Model Evaluation	h5Py	3.1.0
			hankel	1.2.1
			imbalanced-learn	0.8.0
			Keras	2.6.0
			Matplotlib	3.4.3
			NumPy	1.19.5
			OpenCV	4.5.3.56
			pandas	1.3.2
			pathlib	1.0.1
			pillow	9.2.0
			pydot	1.4.2
			PyEtk	1.5.0
			SciKit-Learn	0.24.2
			SciPy	1.7.1
Tensorflow-GPU	2.6.0			
tqdm	4.64.0			



---

## Bibliography

---

- Afrashteh, N., Inayat, S., Mohsenvand, M. & Mohajerani, M. H. (2017). Optical-flow analysis toolbox for characterization of spatiotemporal dynamics in mesoscale optical imaging of brain activity. *NeuroImage*, 153(March), 58–74. doi:10.1016/j.neuroimage.2017.03.034
- Aitchison, L. & Lengyel, M. (2017). With or without you: Predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology*, 46, 219–227. doi:10.1016/j.conb.2017.08.010
- Akasaki, T., Sato, H., Yoshimura, Y., Ozeki, H. & Shimegi, S. (2002). Suppressive effects of receptive field surround on neuronal activity in the cat primary visual cortex. *Neuroscience Research*, 43(3), 207–220. doi:10.1016/S0168-0102(02)00038-X
- Akrami, H., Joshi, A. A., Li, J., Aydore, S. & Leahy, R. M. (2020, April). Brain lesion detection using a robust variational autoencoder and transfer learning. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (pp. 786–790). doi:10.1109/ISBI45749.2020.9098405
- Alberts, B., Wilson, J. & Hunt, T. (2008). *Molecular biology of the cell* (5th ed.). New York: Garland Science.
- Arieli, A., Shoham, D., Hildesheim, R. & Grinvald, A. (1995). Coherent spatiotemporal patterns of ongoing activity revealed by real-time optical imaging coupled with single-unit recording in the cat visual cortex. *Journal of Neurophysiology*, 73(5), 2072–2093. doi:10.1152/jn.1995.73.5.2072
- Arieli, A., Sterkin, A., Grinvald, A. & Aertsen, A. (1996). Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science*, 273(5283), 1868–1871. doi:10.1126/science.273.5283.1868
- Attwell, D. & Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow and Metabolism*, 21(10), 1133–1145. doi:10.1097/00004647-200110000-00001
- Bartfeld, E. & Grinvald, A. (1992). Relationships between orientation-preference pinwheels, cytochrome oxidase blobs, and ocular-dominance columns in primate striate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 89(24), 11905–11909. doi:10.1073/pnas.89.24.11905

- 
- Bathellier, B., Ushakova, L. & Rumpel, S. (2012). Discrete neocortical dynamics predict behavioral categorization of sounds. *Neuron*, 76(2), 435–449. doi:10.1016/j.neuron.2012.07.008
- Bathellier, B., Van De Ville, D., Blu, T., Unser, M. & Carleton, A. (2007). Wavelet-based multi-resolution statistics for optical imaging signals: Application to automated detection of odour activated glomeruli in the mouse olfactory bulb. *NeuroImage*, 34(3), 1020–1035. doi:10.1016/j.neuroimage.2006.10.038
- Bayer, J. & Osendorfer, C. (2014). Learning stochastic recurrent networks, 1–9. arXiv: 1411.7610
- Baylor, D. A., Fuortes, M. G. F. & O’Byrne, P. M. (1971). Receptive fields of cones in the retina of the turtle. *The Journal of Physiology*, 214(2), 265–294. doi:10.1113/jphysiol.1971.sp009432
- Bengio, Y., Courville, A. & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798–1828. doi:10.1109/TPAMI.2013.50
- Bengio, Y., Laufer, E., Alain, G. & Yosinski, J. (2014). Deep generative stochastic networks trainable by backprop. In E. P. Xing & T. Jebara (Eds.), *Proceedings of the 31st International Conference on Machine Learning* (Vol. 32, pp. 226–234). Beijing, China: PMLR.
- Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, 87(1), 527–570. doi:10.1007/BF01797193
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. (1st ed.). New York, N.Y.: Springer New York.
- Blasdel, G. G. (1992). Differential imaging of ocular dominance and orientation selectivity in monkey striate cortex. *The Journal of Neuroscience*, 12(8), 3115–3138. doi:10.1523/JNEUROSCI.12-08-03115.1992
- Blasdel, G. G. & Salama, G. (1986). Voltage-sensitive dyes reveal a modular organization in monkey striate cortex. *Nature*, 321(6070), 579–585. doi:10.1038/321579a0
- Blei, D. M., Kucukelbir, A. & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518), 859–877. doi:10.1080/01621459.2017.1285773
- Blumenfeld, B. (2010). An algorithm for the analysis of temporally structured multidimensional measurements. *Frontiers in Computational Neuroscience*, 3. doi:10.3389/neuro.10.028.2009
- Bond-Taylor, S., Leach, A., Long, Y. & Willcocks, C. G. (2021). Deep generative modelling: A comparative review of VAEs, GANs, normalizing flows, energy-based and autoregressive models. doi:10.1109/TPAMI.2021.3116668. arXiv: 2103.04922

- 
- Bonhoeffer, T. & Grinvald, A. (1991). Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature*, 353(6343), 429–431. doi:10.1038/353429a0
- Bosking, W. H., Zhang, Y., Schofield, B. & Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *The Journal of Neuroscience*, 17(6), 2112–2127. doi:10.1523/JNEUROSCI.17-06-02112.1997
- Bowman, S. R., Vilnis, L., Vinyals, O., Dai, A. M., Jozefowicz, R. & Bengio, S. (2016). Generating sentences from a continuous space. In *Proceedings of the 20th SIGLL Conference on Computational Natural Language Learning (CONLL) 2016* (pp. 10–21). doi:https://doi.org/10.48550/arXiv.1511.06349
- Boycott, B. B. & Wässle, H. (1974). The morphological types of ganglion cells of the domestic cat's retina. *The Journal of Physiology*, 240(2), 397–419. doi:10.1113/jphysiol.1974.sp010616
- Bradski, G. R. & Kaehler, A. (2008). *Learning OpenCV - Computer vision with the OpenCV library: Software that sees*. (1st ed.). Sebastopol, CA: O'Reilly Media, Inc.
- Brodmann, K. (1903). Beiträge zur histologischen Lokalisation der Grosshirnrinde. Zweite Mitteilung: Der Calcarinatypus. *J. Psychol. Neurol.*, 2, 133–159.
- Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Leipzig: Barth.
- Brosch, M., Bauer, R. & Eckhorn, R. (1995). Synchronous high-frequency oscillations in cat area 18. *European Journal of Neuroscience*, 7(1), 86–95. doi:10.1111/j.1460-9568.1995.tb01023.x
- Burel, G. (1992). Blind separation of sources: A nonlinear neural algorithm. *Neural Networks*, 5(6), 937–947. doi:10.1016/S0893-6080(05)80090-5
- Burke, W., Dreher, B. & Wang, C. (1998). Selective block of conduction in Y optic nerve fibres: significance for the concept of parallel processing. *European Journal of Neuroscience*, 10(1), 8–19. doi:10.1046/j.1460-9568.1998.00025.x
- Burkhardt, D. A. (1993). Synaptic feedback, depolarization, and color opponency in cone photoreceptors. *Visual Neuroscience*, 10(6), 981–989. doi:10.1017/S0952523800010087
- Calabrese, A., Schumacher, J. W., Schneider, D. M., Paninski, L. & Woolley, S. M. N. (2011). A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. *PLoS ONE*, 6(1), e16104. doi:10.1371/journal.pone.0016104
- Callaway, E. M. (2005). Structure and function of parallel pathways in the primate early visual system. *The Journal of Physiology*, 566(1), 13–19. doi:10.1113/jphysiol.2005.088047
- Carandini, M. (2004). Amplification of trial-to-trial response variability by neurons in visual cortex. *PLoS Biology*, 2(9), e264. doi:10.1371/journal.pbio.0020264

- 
- Carmi, O., Gross, A., Ivzan, N., Franca, L. L., Farah, N., Zalevsky, Z. & Mandel, Y. (2021). Evaluation and optimization of methods for generating high-resolution retinotopic maps using visual cortex voltage-sensitive dye imaging. *Frontiers in Cellular Neuroscience*, 15. doi:10.3389/fncel.2021.713538
- Castro-Alamancos, M. A. (2004). Absence of rapid sensory adaptation in neocortex during information processing states. *Neuron*, 41(3), 455–464. doi:10.1016/S0896-6273(03)00853-5
- Caton, R. (1875). The electric currents of the brain. *British Medical Journal*, 2, 278.
- Cauchy, A.-L. (1846). Sur les intégrales qui s'étendent à tous les points d'une courbe fermée. *Comptes rendus de l'Académie des sciences*, 23, 251–255.
- Chatterjee, S., Sciarra, A., Dünwald, M., Tummala, P., Agrawal, S. K., Jauhari, A., ... Nürnberger, A. (2022). StRegA: Unsupervised anomaly detection in brain MRIs using a compact context-encoding variational autoencoder. doi:10.1016/j.combiomed.2022.106093
- Chemla, S., Muller, L., Reynaud, A., Takerkart, S., Destexhe, A. & Chavane, F. (2017). Improving voltage-sensitive dye imaging: With a little help from computational approaches. *Neurophotonics*, 4(3), 031215. doi:10.1117/1.NPh.4.3.031215
- Chen, W., Park, K., Pan, Y., Koretsky, A. P. & Du, C. (2020). Interactions between stimulus-evoked cortical activity and spontaneous low frequency oscillations measured with neuronal calcium. *NeuroImage*, 210, 116554. doi:10.1016/j.neuroimage.2020.116554
- Chen, Y., Geisler, W. S. & Seidemann, E. (2006). Optimal decoding of correlated neural population responses in the primate visual cortex. *Nature Neuroscience*, 9(11), 1412–1420. doi:10.1038/nn1792
- Chen, Z., Gomperts, S. N., Yamamoto, J. & Wilson, M. A. (2014). Neural representation of spatial topology in the rodent hippocampus. *Neural Computation*, 26(1), 1–39. doi:10.1162/NECO\_a\_00538
- Chollet, F. (2015). Keras. <https://github.com/fchollet/keras>. GitHub.
- Chollet, F. (2016). Xception: Deep learning with depthwise separable convolutions. arXiv: 1610.02357
- Chung, J., Kastner, K., Dinh, L., Goel, K., Courville, A. C. & Bengio, Y. (2015). A recurrent latent variable model for sequential data. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada* (pp. 2980–2988).
- Clelland, B. G., Harding, T. H. & Tulunay-Keesey, U. (1979). Visual resolution and receptive field size: Examination of two kinds of cat retinal ganglion cell. *Science*, 205(4410), 1015–1017. doi:10.1126/science.472720

- 
- Cleland, B. G., Levick, W. R. & Wässle, H. (1975). Physiological identification of a morphological class of cat retinal ganglion cells. *The Journal of Physiology*, 248(1), 151–171. doi:10.1113/jphysiol.1975.sp010967
- Cohen, L. B., Salzberg, B. M., Davila, H. V., Ross, W. N., Landowne, D., Waggoner, A. S. & Wang, C. H. (1974). Changes in axon fluorescence during activity: Molecular probes of membrane potential. *The Journal of Membrane Biology*, 19(1), 1–36. doi:10.1007/BF01869968
- Cook, P. B. & McReynolds, J. S. (1998). Lateral inhibition in the inner retina is important for spatial tuning of ganglion cells. *Nature Neuroscience*, 1(8), 714–719. doi:10.1038/3714
- Cowey, A. (1964). Projection of the retina on to striate and prestriate cortex in the squirrel monkey, *Saimiri Sciureus*. *Journal of Neurophysiology*, 27(3), 366–393. doi:10.1152/jn.1964.27.3.366
- Cremer, C., Li, X. & Duvenaud, D. (2018). Inference suboptimality in variational autoencoders. arXiv: 1801.03558
- Croitoru, F.-A., Hondru, V., Ionescu, R. T. & Shah, M. (2022). Diffusion models in vision: A survey. arXiv: 2209.04747
- Cunningham, J. P. & Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, 17(11), 1500–1509. doi:10.1038/nn.3776
- Curto, C., Sakata, S., Marguet, S., Itskov, V. & Harris, K. D. (2009). A Simple Model of Cortical Dynamics Explains Variability and State Dependence of Sensory Responses in Urethane-Anesthetized Auditory Cortex. *Journal of Neuroscience*, 29(34), 10600–10612. doi:10.1523/JNEUROSCI.2053-09.2009
- Dai, B. & Wipf, D. (2019). Diagnosing and enhancing VAE models. *ICLR 2019*. arXiv: 1903.05789
- Deco, G. & Brauer, W. (1995). Nonlinear higher-order statistical decorrelation by volume-conserving neural architectures. *Neural Networks*, 8(4), 525–535. doi:10.1016/0893-6080(94)00108-X
- Dinh, L., Krueger, D. & Bengio, Y. (2014). NICE: Non-linear independent components estimation. arXiv: 1410.8516. Retrieved from <http://arxiv.org/abs/1410.8516>
- Dreher, B., Wang, C. & Burke, W. (1996). Limits of parallel processing: Excitatory convergence of different information channels on single neurons in striate and extrastriate visual cortices. *Clinical and Experimental Pharmacology and Physiology*, 23(10-11), 913–925. doi:10.1111/j.1440-1681.1996.tb01143.x
- Dreher, B., Wang, C., Turlajski, K. J., Djavadian, R. L. & Burke, W. (1996). Areas PMLS and 21 a of cat visual cortex: Two functionally distinct areas. *Cerebral Cortex*, 6(4), 585–599. doi:10.1093/cercor/6.4.585

- 
- Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M. & Reitboeck, H. J. (1988). Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biological Cybernetics*, 60(2), 121–130. doi:10.1007/BF00202899
- Einstein, A. (1905). Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik*, 322(6), 132–148. doi:10.1002/andp.19053220607
- Engel, A. K., Kreiter, A. K., König, P. & Singer, W. (1991). Synchronization of oscillatory neuronal responses between striate and extrastriate visual cortical areas of the cat. *Proceedings of the National Academy of Sciences*, 88(14), 6048–6052. doi:10.1073/pnas.88.14.6048
- Enroth-Cugell, C. & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *The Journal of Physiology*, 187(3), 517–552. doi:10.1113/jphysiol.1966.sp008107
- Ermentrout, G. & Kleinfeld, D. (2001). Traveling Electrical Waves in Cortex. *Neuron*, 29(1), 33–44. doi:10.1016/S0896-6273(01)00178-7
- Fekete, T., Omer, D. B., Naaman, S. & Grinvald, A. (2009). Removal of spatial biological artifacts in functional maps by local similarity minimization. *Journal of Neuroscience Methods*, 178(1), 31–39. doi:10.1016/j.jneumeth.2008.11.020
- Felleman, D. J. & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1), 1–47. doi:10.1093/cercor/1.1.1
- Fenko, L., Yizhar, O. & Deisseroth, K. (2011). The development and application of optogenetics. *Annual Review of Neuroscience*, 34(1), 389–412. doi:10.1146/annurev-neuro-061010-113817
- Ferezou, I., Mátyás, F. & Petersen, C. (2009). Imaging the brain in action: Real-time voltage-sensitive dye imaging of sensorimotor cortex of awake behaving mice. *In vivo optical imaging of brain function*, 171–192. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK20229/>
- Fiser, J., Berkes, P., Orbán, G. & Lengyel, M. (2010). Statistically optimal perception and learning: From behavior to neural representations. *Trends in Cognitive Sciences*, 14(3), 119–130. doi:10.1016/j.tics.2010.01.003
- Flotho, P., Haab, L., Eckert, D., Takahashi, K., Schwerdtfeger, K. & Strauss, D. J. (2019). Semi-synthetic dataset for the evaluation of motion compensation approaches for voltage sensitive dye imaging. *International IEEE/EMBS Conference on Neural Engineering, NER, 2019-March*, 1134–1137. doi:10.1109/NER.2019.8716905
- Fraccaro, M. (2018). *Deep Latent Variable Models for Sequential Data* (Doctoral dissertation, Technical University of Denmark).
- Fraccaro, M., Kamronn, S., Paquet, U. & Winther, O. (2017). A disentangled recognition and nonlinear dynamics model for unsupervised learning. arXiv: 1710.05741



- 
- Frey, B. J. (1998). *Graphical Models for Machine Learning and Digital Communication (Adaptive Computation and Machine Learning series)*. doi:10.7551/mitpress/3348.001.0001
- Frey, B. J., Hinton, G. E. & Dayan, P. (1995). Does the wake-sleep algorithm produce good density estimators? In *Proceedings of the 8th International Conference on Neural Information Processing Systems* (pp. 661–667). Cambridge, MA, USA: MIT Press.
- Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, 9(10), 474–480. doi:10.1016/j.tics.2005.08.011
- Fries, P., Neuenschwander, S., Engel, A. K., Goebel, R. & Singer, W. (2001). Rapid feature selective neuronal synchronization through correlated latency shifting. *Nature Neuroscience*, 4(2), 194–200. doi:10.1038/84032
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D. & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4), 189–210. doi:10.1002/hbm.460020402
- Frostig, R. D., Lieke, E. E., Ts'o, D. Y. & Grinvald, A. (1990). Cortical functional architecture and local coupling between neuronal activity and the microcirculation revealed by in vivo high-resolution optical imaging of intrinsic signals. *Proceedings of the National Academy of Sciences*, 87(16), 6082–6086. doi:10.1073/pnas.87.16.6082
- Fu, H., Li, C., Liu, X., Gao, J., Celikyilmaz, A. & Carin, L. (2019). Cyclical annealing schedule: A simple approach to mitigating KL vanishing. arXiv: 1903.10145v3
- Fukuda, M., Rajagopalan, U. M., Homma, R., Matsumoto, M., Nishizaki, M. & Tanifuji, M. (2005). Localization of activity-dependent changes in blood volume to submillimeter-scale functional domains in cat visual cortex. *Cerebral Cortex*, 15(6), 823–833. doi:10.1093/cercor/bhh183
- Galuske, R. A. W., Munk, M. H. J. & Singer, W. (2019). Relation between gamma oscillations and neuronal plasticity in the visual cortex. *Proceedings of the National Academy of Sciences*, 116(46), 23317–23325. doi:10.1073/pnas.1901277116
- Galuske, R. A. W., Schmidt, K. E., Goebel, R., Lomber, S. G. & Payne, B. R. (2002). The role of feedback in shaping neural representations in cat visual cortex. *Proceedings of the National Academy of Sciences*, 99(26), 17083–17088. doi:10.1073/pnas.242399199
- Ganguli, S. & Sompolinsky, H. (2012). Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annual Review of Neuroscience*, 35(1), 485–508. doi:10.1146/annurev-neuro-062111-150410
- Gavrilyuk, S., Polyutov, S., Jha, P. C., Rinkevicius, Z., Ågren, H. & Gel'mukhanov, F. (2007). Many-photon dynamics of photobleaching. *The Journal of Physical Chemistry A*, 111(47), 11961–11975. doi:10.1021/jp074756x

- 
- Gelman, A., Carlin, J., Stern, H. & Rubin, D. (2003). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.
- Gershman, S. J. & Goodman, N. D. (2014). Amortized inference in probabilistic reasoning. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 517–522). Quebec City, CA.
- Ghisovan, N., Nemri, A., Shumikhina, S. & Molotchnikoff, S. (2008). Synchrony between orientation-selective neurons is modulated during adaptation-induced plasticity in cat visual cortex. *BMC Neuroscience*, 9(1), 60. doi:10.1186/1471-2202-9-60
- Gilbert, C. D., Das, A., Ito, M., Kapadia, M. & Westheimer, G. (1996). Spatial integration and cortical dynamics. *Proceedings of the National Academy of Sciences*, 93(2), 615–622. doi:10.1073/pnas.93.2.615
- Gilbert, C. D. & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, 14(5), 350–363. doi:10.1038/nrn3476
- Gilbert, C. D. & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron*, 54(5), 677–696. doi:10.1016/j.neuron.2007.05.019
- Gilbert, C. D. & Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *The Journal of Neuroscience*, 9(7), 2432–2442. doi:10.1523/JNEUROSCI.09-07-02432.1989
- Girin, L., Leglaive, S., Bie, X., Diard, J., Hueber, T. & Alameda-Pineda, X. (2021). Dynamical variational autoencoders: A comprehensive review. *Foundations and Trends® in Machine Learning*, 15(1-2), 1–175. doi:10.1561/22000000089
- Goodchild, A. K., Ghosh, K. K. & Martin, P. R. (1996). Comparison of photoreceptor spatial density and ganglion cell morphology in the retina of human, macaque monkey, cat, and the marmoset *Callithrix jacchus*. *The Journal of Comparative Neurology*, 366(1), 55–75. doi:10.1002/(SICI)1096-9861(19960226)366:1<55::AID-CNE5>3.0.CO;2-J
- Goodfellow, I. (2016). NIPS 2016 tutorial: Generative adversarial networks. arXiv: 1701.00160
- Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep Learning*. MIT Press. Retrieved from <http://www.deeplearningbook.org>
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (Vol. 27), Curran Associates, Inc.
- Goyal, A., Sordoni, A., Côté, M.-A., Ke, N. R. & Bengio, Y. (2017). Z-forcing: Training stochastic recurrent networks. doi:<https://doi.org/10.48550/arXiv.1711.05411>. arXiv: 1711.05411

- 
- Grandy, T. H., Greenfield, S. A. & Devonshire, I. M. (2012). An evaluation of in vivo voltage-sensitive dyes: Pharmacological side effects and signal-to-noise ratios after effective removal of brain-pulsation artifacts. *Journal of Neurophysiology*, 108(11), 2931–2945. doi:10.1152/jn.00512.2011
- Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., ... Vanrumste, B. (2008). Review on solving the inverse problem in EEG source analysis. *Journal of NeuroEngineering and Rehabilitation*, 5(1), 25. doi:10.1186/1743-0003-5-25
- Greene, W. H. (2008). *Econometric analysis*. Upper Saddle River, N.J.
- Grinvald, A., Anglister, L., Freeman, J. A., Hildesheim, R. & Manker, A. (1984). Real-time optical imaging of naturally evoked electrical activity in intact frog brain. *Nature*, 308(5962), 848–850. doi:10.1038/308848a0
- Grinvald, A., Frostig, R. D., Siegel, R. M. & Bartfeld, E. (1991). High-resolution optical imaging of functional brain architecture in the awake monkey. *Proceedings of the National Academy of Sciences*, 88(24), 11559–11563. doi:10.1073/pnas.88.24.11559
- Grinvald, A. & Hildesheim, R. (2004). VSDI: A new era in functional imaging of cortical dynamics. *Nature Reviews Neuroscience*, 5(11), 874–885. doi:10.1038/nrn1536
- Grinvald, A., Hildesheim, R., Farber, I. C. & Anglister, L. (1982). Improved fluorescent probes for the measurement of rapid changes in membrane potential. *Biophysical Journal*, 39(3), 301–308. doi:10.1016/S0006-3495(82)84520-7
- Grinvald, A., Lieke, E., Frostig, R. D., Gilbert, C. D. & Wiesel, T. N. (1986). Functional architecture of cortex revealed by optical imaging of intrinsic signals. *Nature*, 324(6095), 361–364. doi:10.1038/324361a0
- Grinvald, A., Lieke, E. E., Frostig, R. D. & Hildesheim, R. (1994). Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *The Journal of Neuroscience*, 14(5), 2545–2568. doi:10.1523/JNEUROSCI.14-05-02545.1994
- Grinvald, A., Shoham, D., Shmuel, A., Glaser, D. E., Vanzetta, I., Shtoyerman, E., ... Arieli, A. (1999). In-vivo optical imaging of cortical architecture and dynamics. In *Modern techniques in neuroscience research* (pp. 893–969). doi:10.1007/978-3-642-58552-4\_34
- Gross, A., Ivzan, N. H., Farah, N. & Mandel, Y. (2019). High-resolution VSDI retinotopic mapping via a DLP-based projection system. *Biomedical Optics Express*, 10(10), 5117. doi:10.1364/BOE.10.005117
- Harris, K. D. & Thiele, A. (2011). Cortical state and attention. *Nature Reviews Neuroscience*, 12(9), 509–523. doi:10.1038/nrn3084
- Hastie, T., Tibshirani, R. & Friedman, J. (2009). Model assessment and selection. In *The elements of statistical learning: data mining, inference, and prediction* (pp. 219–259). doi:10.1007/978-0-387-84858-7\_7

- 
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97–109. doi:10.1093/biomet/57.1.97
- He, K., Zhang, X., Ren, S. & Sun, J. (2015). Delving deep into rectifiers surpassing human-level performance on ImageNet classification. arXiv: 1502.01852
- Heggelund, P. & Albus, K. (1978). Response variability and orientation discrimination of single cells in striate cortex of cat. *Experimental Brain Research*, 32(2). doi:10.1007/BF00239727
- Heße, F., Prykhodko, V., Schlüter, S. & Attinger, S. (2014). Generating random fields with a truncated power-law variogram: A comparison of several numerical methods. *Environmental Modelling & Software*, 55, 32–48. doi:10.1016/j.envsoft.2014.01.013
- Hickey, T. L. & Gullery, R. W. (1974). An autoradiographic study of retinogeniculate pathways in the cat and the fox. *The Journal of Comparative Neurology*, 156(2), 239–253. doi:10.1002/cne.901560207
- Higgins, I., Matthey, L., Glorot, X., Pal, A., Uria, B., Blundell, C., ... Lerchner, A. (2016). Early visual concept learning with unsupervised deep learning. arXiv: 1606.05579
- Higgins, I., Matthey, L., Pal, A., Burgess, C. P., Glorot, X., Botvinick, M. M., ... Lerchner, A. (2017). beta-VAE: Learning basic visual concepts with a constrained variational framework. In *Proceedings of the 5th International Conference on Learning Representations (ICLR)*, Toulon, FRA.
- Hill, D. K. & Keynes, R. D. (1949). Opacity changes in stimulated nerve. *The Journal of physiology*, 108(3), 278–81.
- Hinton, G. E. & Roweis, S. (2002). Stochastic neighbor embedding. In S. Becker, S. Thrun & K. Obermayer (Eds.), *Advances in Neural Information Processing Systems* (Vol. 15), MIT Press. Retrieved from <https://proceedings.neurips.cc/paper/2002/file/6150ccc6069bea6b5716254057a194ef-Paper.pdf>
- Ho, J., Jain, A. & Abbeel, P. (2020). Denoising diffusion probabilistic models. arXiv: 2006.11239
- Hofmann, D. J. (2020, February). *Interareale raumzeitliche Dynamiken und Variabilität der Aktivität neuronaler Populationen im visuellen Kortex der Katze* (Doctoral dissertation, Technische Universität Darmstadt). doi:<https://doi.org/10.25534/tuprints-00011440>
- Horsley, V. & Clarke, R. H. (1908). The structure and functions of the cerebellum examined by a new method. *Brain*, 31(1), 45–124. doi:10.1093/brain/31.1.45
- Hossein-Zadeh, G.-A., Ardekani, B. A. & Soltanian-Zadeh, H. (2003). A signal subspace approach for modeling the hemodynamic response function in fMRI. *Magnetic Resonance Imaging*, 21(8), 835–843. doi:10.1016/S0730-725X(03)00180-2
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417–441. doi:10.1037/h0071325

- 
- Hubel, D. H. & Wiesel, T. N. (1961). Integrative action in the cat's lateral geniculate body. *The Journal of Physiology*, 155(2), 385–398. doi:10.1113/jphysiol.1961.sp006635
- Hubel, D. H. & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1), 106–154. doi:10.1113/jphysiol.1962.sp006837
- Hubel, D. H. & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195(1), 215–243. doi:10.1113/jphysiol.1968.sp008455
- Hubel, D. H. & Wiesel, T. N. (1972). Laminar and columnar distribution of geniculocortical fibers in the macaque monkey. *The Journal of Comparative Neurology*, 146(4), 421–450. doi:10.1002/cne.901460402
- Hubel, D. H. & Wiesel, T. N. (1974). Uniformity of monkey striate cortex: A parallel relationship between field size, scatter, and magnification factor. *The Journal of Comparative Neurology*, 158(3), 295–305. doi:10.1002/cne.901580305
- Hubel, D. H. & Wiesel, T. N. (1998). Early exploration of the visual cortex. *Neuron*, 20(3), 401–412. doi:10.1016/S0896-6273(00)80984-8
- Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P. & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, 394(6695), 784–787. doi:10.1038/29537
- Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3), 626–634. doi:10.1109/72.761722
- Hyvärinen, A. & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4-5), 411–430. doi:10.1016/S0893-6080(00)00026-5
- Inagaki, S. (2003). Isolation of neural activities from respiratory and heartbeat noises for in vivo optical recording in guinea pigs using independent component analysis. *Neuroscience Letters*, 352, 9–12. doi:10.1016/s0304-3940(03)00998-4
- Jékely, G. (2009). Evolution of phototaxis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1531), 2795–2808. doi:10.1098/rstb.2009.0072
- Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). New York: Springer.
- Jolliffe, I. T. & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. doi:10.1098/rsta.2015.0202
- Jordan, M. I., Ghahramani, Z., Jaakkola, T. S. & Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning*, 37(2), 183–233. doi:10.1023/A:1007665907178
- Kandel, E. R., Koester, J. D., Mack, S. H. & Siegelbaum, S. (Eds.). (2021). *Principles of Neural Science* (6th ed.). New York, N.Y.: McGraw Hill.

- 
- Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A. & Arieli, A. (2003). Spontaneously emerging cortical representations of visual attributes. *Nature*, 425(6961), 954–956. doi:10.1038/nature02078
- Kingma, D. P. & Ba, J. L. (2014). Adam: A method for stochastic optimization. doi:https://doi.org/10.48550/arXiv.1412.6980
- Kingma, D. P. & Dhariwal, P. (2018). Glow: Generative flow with invertible 1x1 convolutions. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 31), Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper/2018/file/d139db6a236200b21cc7f752979132d0-Paper.pdf
- Kingma, D. P. & Welling, M. (2013, December). Auto-Encoding Variational Bayes. In *Proceedings of the 2nd International Conference on Learning Representations (ICLR)* (pp. 1–14). arXiv: 1312.6114. Retrieved from http://arxiv.org/abs/1312.6114
- Kingma, D. P. & Welling, M. (2019). An introduction to variational autoencoders. *Foundations and Trends in Machine Learning*, 12(4), 307–392. doi:10.1561/22000000056. arXiv: 1906.02691
- Klein, B., Kuschinsky, W., Schrock, H. & Vetterlein, F. (1986). Interdependency of local capillary density, blood flow, and metabolism in rat brains. *American Journal of Physiology-Heart and Circulatory Physiology*, 251(6), H1333–H1340. doi:10.1152/ajpheart.1986.251.6.H1333
- Kohn, A. & Smith, M. A. (2005). Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *Journal of Neuroscience*, 25(14), 3661–3673. doi:10.1523/JNEUROSCI.5106-04.2005
- Kompanets, I. & Zalyapin, N. (2020). Methods and devices of speckle-noise suppression (Review). *Optics and Photonics Journal*, 10(10), 219–250. doi:10.4236/opj.2020.1010023
- Konnerth, A. & Orkand, R. K. (1986). Voltage-sensitive dyes measure potential changes in axons and glia of the frog optic nerve. *Neuroscience Letters*, 66(1), 49–54. doi:10.1016/0304-3940(86)90164-3
- Kossale, Y., Airaj, M. & Darouichi, A. (2022, October). Mode collapse in generative adversarial networks: An Overview. In *2022 8th International Conference on Optimization and Applications (ICOA)* (pp. 1–6). doi:10.1109/ICOA55659.2022.9934291
- Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52(6), 119–139.
- Krishnan, R., Shalit, U. & Sontag, D. (2017). Structured inference networks for nonlinear state space models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1). doi:10.1609/aaai.v31i1.10779



- 
- Krishnan, R. G., Shalit, U. & Sontag, D. (2015). Deep Kalman filters. arXiv: 1511.05121. Retrieved from <http://arxiv.org/abs/1511.05121>
- Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16(1), 37–68. doi:10.1152/jn.1953.16.1.37
- Kullback, S. & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86. doi:10.1214/aoms/1177729694
- Latimer, K. W., Yates, J. L., Meister, M. L. R., Huk, A. C. & Pillow, J. W. (2015). Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science*, 349(6244), 184–187. doi:10.1126/science.aaa4056
- Leglaive, S., Alameda-Pineda, X., Girin, L. & Horaud, R. (2020, May). A recurrent variational autoencoder for speech enhancement. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 371–375). doi:10.1109/ICASSP40776.2020.9053164
- Leventhal, A. G., Rodieck, R. W. & Dreher, B. (1981). Retinal ganglion cell classes in the old world monkey: Morphology and central projections. *Science*, 213(4512), 1139–1142. doi:10.1126/science.7268423
- Liao, L.-D., Tsytsarev, V., Delgado-Martínez, I., Li, M.-L., Erzurumlu, R., Vipin, A., ... Thakor, N. V. (2013). Neurovascular coupling: In vivo optical techniques for functional brain imaging. *BioMedical Engineering OnLine*, 12(1), 38. doi:10.1186/1475-925X-12-38
- Lippert, M. T., Takagaki, K., Xu, W., Huang, X. & Wu, J.-Y. (2007). Methods for voltage-sensitive dye imaging of rat cortical activity with high signal-to-noise ratio. *Journal of Neurophysiology*, 98(1), 502–512. doi:10.1152/jn.01169.2006
- Liu, X.-Y., Wu, J. & Zhou, Z.-H. (2009). Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(2), 539–550. doi:10.1109/TSMCB.2008.2007853
- Livingstone, M. & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science*, 240(4853), 740–749. doi:10.1126/science.3283936
- Loaiza-Ganem, G., Ross, B. L., Cresswell, J. C. & Caterini, A. L. (2022). Diagnosing and fixing manifold overfitting in deep generative models. arXiv: 2204.07172
- Lomber, S. G., Payne, B. R., Cornwell, P. & Long, K. D. (1996). Perceptual and cognitive visual functions of parietal and temporal cortices in the cat. *Cerebral Cortex*, 6(5), 673–695. doi:10.1093/cercor/6.5.673
- Long, T., Cao, Y. & Cheung, J. C. K. (2019). On posterior collapse and encoder feature dispersion in sequence VAEs. arXiv: 1911.03976

- 
- Lucas, J., Tucker, G., Grosse, R. & Norouzi, M. (2019). Understanding posterior collapse in generative latent variable models. In *Proceedings of the 7th International Conference on Learning Representations (ICLR)* (pp. 1–16).
- Luczak, A., Barthó, P. & Harris, K. D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron*, 62(3), 413–425. doi:10.1016/j.neuron.2009.03.014
- Macke, J. H., Gerwinn, S., White, L. E., Kaschube, M. & Bethge, M. (2009). Bayesian estimation of orientation preference maps. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams & A. Culotta (Eds.), *Proceedings of the 22nd International Conference on Neural Information Processing Systems* (pp. 1195–1203). doi:10.5555/2984093.2984228
- Macke, J. H., Gerwinn, S., White, L. E., Kaschube, M. & Bethge, M. (2011). Gaussian process methods for estimating cortical maps. *NeuroImage*, 56(2), 570–581. doi:10.1016/j.neuroimage.2010.04.272
- Maeda, S., Inagaki, S., Kawaguchi, H. & Song, W.-J. (2001). Separation of signal and noise from in vivo optical recording in Guinea pigs using independent component analysis. *Neuroscience Letters*, 302(2-3), 137–140. doi:10.1016/S0304-3940(01)01678-0
- Malach, R., Amir, Y., Harel, M. & Grinvald, A. (1993). Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex. *Proceedings of the National Academy of Sciences*, 90(22), 10469–10473. doi:10.1073/pnas.90.22.10469
- Mangel, S. C. (1991). Analysis of the horizontal cell contribution to the receptive field surround of ganglion cells in the rabbit retina. *The Journal of Physiology*, 442(1), 211–234. doi:10.1113/jphysiol.1991.sp018790
- Martindale, J., Berwick, J., Martin, C., Kong, Y., Zheng, Y. & Mayhew, J. (2005). Long duration stimuli and nonlinearities in the neural–haemodynamic coupling. *Journal of Cerebral Blood Flow and Metabolism*, 25(5), 651–661. doi:10.1038/sj.jcbfm.9600060
- Masland, R. H. (2012). The tasks of amacrine cells. *Visual Neuroscience*, 29(1), 3–9. doi:10.1017/S0952523811000344
- Mishra, S. (2017). Handling imbalanced data: SMOTE vs. random undersampling. *International Research Journal of Engineering and Technology (IRJET)*, 4(8), 317–320.
- Mountcastle, V. B. (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex. *Journal of Neurophysiology*, 20(4), 408–434. doi:10.1152/jn.1957.20.4.408
- Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain*, 120(4), 701–722. doi:10.1093/brain/120.4.701
- Müller, S. & Schüler, L. (2020, April). GeoStat-Framework/GSTools: Volatile Violet v1.2.1 (Version v1.2.1). doi:https://doi.org/10.5281/zenodo.3751743



- 
- Müller, S., Schüler, L., Zech, A. & Heße, F. (2021). GStools v1.3: A toolbox for geostatistical modelling in Python. *Geoscientific Model Development Discussions*, 2021, 1–33. doi:10.5194/gmd-2021-301
- Müller, W. A., Frings, S. & Möhrlein, F. (2019). *Tier- und Humanphysiologie* (6th ed.). doi:10.1007/978-3-662-58462-0
- Murphy, K., Jones, D. & Van Sluyters, R. (1995). Cytochrome-oxidase blobs in cat primary visual cortex. *The Journal of Neuroscience*, 15(6), 4196–4208. doi:10.1523/JNEUROSCI.15-06-04196.1995
- Nauhaus, I., Busse, L., Carandini, M. & Ringach, D. L. (2009). Stimulus contrast modulates functional connectivity in visual cortex. *Nature Neuroscience*, 12(1), 70–76. doi:10.1038/nn.2232
- Neal, R. M. (2001). Annealed importance sampling. *Statistics and Computing*, 11(2), 125–139. doi:10.1023/A:1008923215028
- Ng, A. Y. & Jordan, M. I. (2001). On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. In *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic* (pp. 841–848). Cambridge, MA, USA: MIT Press.
- Nikolenko, S. I. (2021). *Synthetic Data for Deep Learning* (1st ed.). doi:10.1007/978-3-030-75178-4
- Nonnenmacher, M., Turaga, S. C. & Macke, J. H. (2017). Extracting low-dimensional dynamics from multiple large-scale neural population recordings by learning to predict correlations. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30 (NIPS 2017)* (Vol. 30, pp. 5702–5712). Curran Associates, Inc.
- O’Brien, D. F. (1982). The chemistry of vision. *Science*, 218(4576), 961–966. doi:10.1126/science.6291153
- Orbach, H. & Cohen, L. B. (1983). Optical monitoring of activity from many areas of the in vitro and in vivo salamander olfactory bulb: a new method for studying functional organization in the vertebrate central nervous system. *The Journal of Neuroscience*, 3(11), 2251–2262. doi:10.1523/JNEUROSCI.03-11-02251.1983
- Pace-Schott, E. F. & Hobson, J. A. (2002). The neurobiology of sleep: Genetics, cellular physiology and subcortical networks. *Nature Reviews Neuroscience*, 3(8), 591–605. doi:10.1038/nrn895
- Parsons, L., Haque, E. & Liu, H. (2004). Subspace clustering for high dimensional data. *ACM SIGKDD Explorations Newsletter*, 6(1), 90–105. doi:10.1145/1007730.1007731
- Pascual-Leone, A. & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, 292(5516), 510–512. doi:10.1126/science.1057099

- 
- Payne, B. R. & Lomber, S. G. (2003). Quantitative analyses of principal and secondary compound parieto-occipital feedback pathways in cat. *Experimental Brain Research*, 152(4), 420–433. doi:10.1007/s00221-003-1554-x
- Payne, B. R. & Peters, A. (2002). *The Cat Primary Visual Cortex* (1st ed.). San Diego, CA: Academic Press.
- Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, 3, 96–146. doi:10.1214/09-SS057
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559–572. doi:10.1080/14786440109462720
- Peter, M. (2019, July). *Repräsentation und Analyse von Bewegungsinformationen in den kortikalen Komponenten des visuellen Systems der Katze* (Doctoral dissertation, Technische Universität Darmstadt).
- Petersen, C. C. H., Grinvald, A. & Sakmann, B. (2003). Spatiotemporal dynamics of sensory responses in layer 2/3 of rat barrel cortex measured in vivo by voltage-sensitive dye imaging combined with whole-cell voltage recordings and neuron reconstructions. *The Journal of Neuroscience*, 23(4), 1298–309. doi:23/4/1298[pii]
- Petersen, C. C. H., Hahn, T. T. G., Mehta, M., Grinvald, A. & Sakmann, B. (2003). Interaction of sensory responses with spontaneous depolarization in layer 2/3 barrel cortex. *Proceedings of the National Academy of Sciences*, 100(23), 13638–13643. doi:10.1523/JNEUROSCI.23-04-01298.2003
- Planck, M. (1901). Ueber das Gesetz der Energieverteilung im Normalspectrum. *Annalen der Physik*, 309(3), 553–563. doi:10.1002/andp.19013090310
- Polack, P.-O. & Contreras, D. (2012). Long-range parallel processing and local recurrent activity in the visual cortex of the mouse. *Journal of Neuroscience*, 32(32), 11120–11131. doi:10.1523/JNEUROSCI.6304-11.2012
- Poulet, J. F. A. & Petersen, C. C. H. (2008). Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. *Nature*, 454(7206), 881–885. doi:10.1038/nature07150
- Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., McNamara, J. O. & Williams, S. M. (2001). *Neuroscience* (2nd ed.). Sunderland (MA): Sinauer Associates.
- Raguet, H., Monier, C., Foubert, L., Ferezou, I., Fregnac, Y. & Peyré, G. (2016). Spatially structured sparse morphological component separation for voltage-sensitive dye optical imaging. *Journal of Neuroscience Methods*, 257, 76–96. doi:10.1016/j.jneumeth.2015.09.024
- Rasmussen, C. E. & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning* (2nd ed.). Cambridge, MA: The MIT Press.

- 
- Ratzlaff, E. H. & Grinvald, A. (1991). A tandem-lens epifluorescence microscope: Hundred-fold brightness advantage for wide-field imaging. *Journal of Neuroscience Methods*, 36(2-3), 127–137. doi:10.1016/0165-0270(91)90038-2
- Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B. & Yang, G.-Z. (2017). Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1), 4–21. doi:10.1109/JBHI.2016.2636665
- Reidl, J., Starke, J., Omer, D. B., Grinvald, A. & Spors, H. (2007). Independent component analysis of high-resolution imaging data identifies distinct functional domains. *NeuroImage*, 34(1), 94–108. doi:10.1016/j.neuroimage.2006.08.031
- Reinoso-Suárez, F. (1961). *Topographischer Hirnatlas der Katze für experimental-physiologische Untersuchungen*. Darmstadt: Merck.
- Reynaud, A., Takerkart, S., Masson, G. & Chavane, F. (2011). Linear model decomposition for voltage-sensitive dye imaging signals. *NeuroImage*, 54(2), 1196–1210. doi:10.1016/j.neuroimage.2010.08.041
- Rezende, D. J. & Mohamed, S. (2015). Variational inference with normalizing flows. arXiv: 1505.05770
- Rezende, D. J., Mohamed, S. & Wierstra, D. (2014). Stochastic backpropagation and approximate inference in deep generative models. arXiv: 1401.4082. Retrieved from <http://arxiv.org/abs/1401.4082>
- Ringach, D. L. (2009). Spontaneous and driven cortical activity: implications for computation. *Current Opinion in Neurobiology*, 19(4), 439–444. doi:10.1016/j.conb.2009.07.005
- Robert, C. P. & Casella, G. (2004). *Monte Carlo statistical methods* (2nd ed.). New York, NY: Springer New York.
- Rockland, K. S. & Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Research*, 179(1), 3–20. doi:10.1016/0006-8993(79)90485-2
- Rodieck, R. W. & Watanabe, M. (1993). Survey of the morphology of macaque retinal ganglion cells that project to the pretectum, superior colliculus, and parvocellular laminae of the lateral geniculate nucleus. *The Journal of Comparative Neurology*, 338(2), 289–303. doi:10.1002/cne.903380211
- Runions, A., Fuhrer, M., Lane, B., Federl, P., Rolland-Lagan, A. G. & Prusinkiewicz, P. (2005). Modeling and visualization of leaf venation patterns. *ACM Transactions on Graphics*, 24(3), 702–711. doi:10.1145/1073204.1073251
- Runions, A., Lane, B. & Prusinkiewicz, P. (2007). Modeling trees with a space colonization algorithm. *Natural Phenomena*, 63–70.

- 
- Rydhmer, K. & Selvan, R. (2021). Dynamic  $\beta$ -VAEs for quantifying biodiversity by clustering optically recorded insect signals. *Ecological Informatics*, 66, 101456. doi:10.1016/j.ecoinf.2021.101456
- Sabri, M. M. & Arabzadeh, E. (2018). Information processing across behavioral states: Modes of operation and population dynamics in rodent sensory cortex. *Neuroscience*, 368, 214–228. doi:10.1016/j.neuroscience.2017.09.016
- Saka, M., Berwick, J. & Jones, M. (2012). Inter-trial variability in sensory-evoked cortical hemodynamic responses: The role of the magnitude of pre-stimulus fluctuations. *Frontiers in Neuroenergetics*, 4. doi:10.3389/fnene.2012.00010
- Salinas, H. & Fernandez, D. (2007). Comparison of PDE-based nonlinear diffusion approaches for image enhancement and denoising in optical coherence tomography. *IEEE Transactions on Medical Imaging*, 26(6), 761–771. doi:10.1109/TMI.2006.887375
- Sanchez-Vives, M. V. & McCormick, D. A. (2000). Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nature Neuroscience*, 3(10), 1027–1034.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. arXiv: 1801.04381
- Sato, T. K., Nauhaus, I. & Carandini, M. (2012). Traveling waves in visual cortex. *Neuron*, 75(2), 218–229. doi:10.1016/j.neuron.2012.06.029
- Schau, G. F., Thibault, G., Dane, M. A., Gray, J. W., Heiser, L. M. & Chang, Y. H. (2019). Variational autoencoding tissue response to microenvironment perturbation. *Proceedings of SPIE—the International Society for Optical Engineering*, 10949. doi:10.1117/12.2512660
- Schmidt, D., Rausch, A. & Schanze, T. (2020). Deep learning-based recognition of cell structures in fluorescence microscopy sequences with respect to their morphology on cells infected with Marburg virus. *Current Directions in Biomedical Engineering*, 6(3), 501–504. doi:10.1515/cdbme-2020-3129
- Schmidt, K. E., Lomber, S. G., Payne, B. R. & Galuske, R. A. W. (2011). Pattern motion representation in primary visual cortex is mediated by transcortical feedback. *NeuroImage*, 54(1), 474–484. doi:10.1016/j.neuroimage.2010.08.017
- Shoham, D., Glaser, D. E., Arieli, A., Kenet, T., Wijnbergen, C., Toledo, Y., ... Grinvald, A. (1999). Imaging cortical dynamics at high spatial and temporal resolution with novel blue voltage-sensitive dyes. *Neuron*, 24(4), 791–802. doi:10.1016/S0896-6273(00)81027-2
- Shoham, D., Hübener, M., Schulze, S., Grinvald, A. & Bonhoeffer, T. (1997). Spatio-temporal frequency domains and their relation to cytochrome oxidase staining in cat visual cortex. *Nature*, 385(6616), 529–533. doi:10.1038/385529a0

- 
- Simmons, G. (1996). *Calculus with analytic geometry* (2nd ed.). New York, N.Y: McGraw-Hill.
- Singer, W. (2013). Cortical dynamics revisited. *Trends in Cognitive Sciences*, 17(12), 616–626. doi:10.1016/j.tics.2013.09.006
- Smith, M. A. & Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. *Journal of Neuroscience*, 28(48), 12591–12603. doi:10.1523/JNEUROSCI.2929-08.2008
- Snoek, J., Larochelle, H. & Adams, R. P. (2012, June). Practical Bayesian optimization of machine learning algorithms. In F. Pereira, C. J. Burges, L. Bottou & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (Vol. 25), Curran Associates, Inc. arXiv: 1206.2944
- Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N. & Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. arXiv: 1503.03585
- Soltanian-Zadeh, S., Sahingur, K., Blau, S., Gong, Y. & Farsiu, S. (2019). Fast and robust active neuron segmentation in two-photon calcium imaging using spatiotemporal deep learning. *Proceedings of the National Academy of Sciences*, 116(17), 8554–8563. doi:10.1073/pnas.1812995116
- Song, Y. & Ermon, S. (2019). Generative modeling by estimating gradients of the data distribution. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 32), Curran Associates, Inc. Retrieved from <https://proceedings.neurips.cc/paper/2019/file/3001ef257407d5a371a96dcd947c7d93-Paper.pdf>
- Steinberg, R. H., Reid, M. & Lacy, P. L. (1973). The distribution of rods and cones in the retina of the cat (*Felis domesticus*). *The Journal of Comparative Neurology*, 148(2), 229–248. doi:10.1002/cne.901480209
- Steriade, M., Nunez, A. & Amzica, F. (1993). A novel slow (< 1 Hz) oscillation of neocortical neurons in vivo: depolarizing and hyperpolarizing components. *The Journal of Neuroscience*, 13(8), 3252–3265. doi:10.1523/JNEUROSCI.13-08-03252.1993
- Steriade, M., Timofeev, I. & Grenier, F. (2001). Natural waking and sleep states: A view from inside neocortical neurons. *Journal of Neurophysiology*, 85(5), 1969–1985. doi:10.1152/jn.2001.85.5.1969
- Sterkin, A., Lampl, I., Ferster, D., Grinvald, A. & Arieli, A. (1998). Real time optical imaging in cat visual cortex exhibits high similarity to intracellular activity. *Neurosci. Lett.*, 51, S41.
- Stetter, M., Greve, H., Galizia, C. & Obermayer, K. (2001). Analysis of calcium imaging signals from the honeybee brain by nonlinear models. *NeuroImage*, 13(1), 119–128. doi:10.1006/nimg.2000.0679

- 
- Stevenson, I. H. (2018). Omitted variable bias in GLMs of neural spiking activity. *Neural Computation*, 30(12), 3227–3258. doi:10.1162/neco\_a\_01138
- Stevenson, I. H. & Kording, K. P. (2011). How advances in neural recording affect data analysis. *Nature Neuroscience*, 14(2), 139–142. doi:10.1038/nn.2731
- Stroh, A., Adelsberger, H., Groh, A., Rühlmann, C., Fischer, S., Schierloh, A., ... Konnerth, A. (2013). Making waves: Initiation and propagation of corticothalamic Ca<sup>2+</sup> waves in vivo. *Neuron*, 77(6), 1136–1150. doi:10.1016/j.neuron.2013.01.031
- Swokowski, W. (1979). *Calculus with Analytic Geometry* (2nd ed.). Boston, MA: Prindle, Weber & Schmidt.
- Symonds, L. & Rosenquist, A. (1984). Corticocortical connections among visual areas in the cat. *The Journal of Comparative Neurology*, 229(1), 1–38. doi:10.1002/cne.902290103PM-6490972M4-Citavi
- Takagaki, K., Lippert, M., Dann, B., Wanger, T. & Ohl, F. (2008). Normalization of voltage-sensitive dye signal with functional activity measures. *PLoS ONE*, 3(12), e4041. doi:http://dx.doi.org/10.1371/journal.pone.0004041
- Tan, M. & Le, Q. V. (2019). MixConv: Mixed depthwise convolutional kernels. arXiv: 1907.09595
- Tandel, G. S., Biswas, M., Kakde, O. G., Tiwari, A., Suri, H. S., Turk, M., ... Suri, J. S. (2019). A review on a deep learning perspective in brain cancer classification. *Cancers*, 11(1), 111. doi:10.3390/cancers11010111
- Tasaki, I., Watanabe, A., Sandlin, R. & Carnay, L. (1968). Changes in fluorescence, turbidity, and birefringence associated with nerve excitation. *Proceedings of the National Academy of Sciences of the United States of America*, 61(3), 883–8. doi:10.1073/pnas.61.3.883
- Thoreson, W. B., Babai, N. & Bartoletti, T. M. (2008). Feedback from horizontal cells to rod photoreceptors in vertebrate retina. *Journal of Neuroscience*, 28(22), 5691–5695. doi:10.1523/JNEUROSCI.0403-08.2008
- Thorpe, S., Fize, D. & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582), 520–522. doi:10.1038/381520a0
- Tishby, N., Pereira, F. C. & Bialek, W. (2000). The information bottleneck method. arXiv: 0004057 [physics]
- Tishby, N. & Zaslavsky, N. (2015). Deep learning and the information bottleneck principle. arXiv: 1503.02406
- Townsend, R. G. & Gong, P. (2018). Detection and analysis of spatiotemporal patterns in brain activity. *PLoS Computational Biology*, 1–29. doi:10.1371/journal.pcbi.1006643
- Townsend, R. G., Solomon, S. S., Chen, S. C., Pietersen, A. N. J., Martin, P. R., Solomon, S. G. & Gong, P. (2015). Emergence of complex wave patterns in primate cerebral



- 
- cortex. *Journal of Neuroscience*, 35(11), 4657–4662. doi:10.1523/JNEUROSCI.4509-14.2015
- Tretter, F., Cynader, M. & Singer, W. (1975). Cat parastriate cortex: a primary or secondary visual area. *Journal of Neurophysiology*, 38(5), 1099–1113. doi:10.1152/jn.1975.38.5.1099
- Ts'o, D., Gilbert, C. & Wiesel, T. (1986). Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *The Journal of Neuroscience*, 6(4), 1160–1170. doi:10.1523/JNEUROSCI.06-04-01160.1986
- Tsodyks, M., Kenet, T., Grinvald, A. & Arieli, A. (1999). Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science*, 286(5446), 1943–1946. doi:10.1126/science.286.5446.1943
- Tusa, R. J., Palmer, L. A. & Rosenquist, A. C. (1978). The retinotopic organization of area 17 (striate cortex) in the cat. *Journal of Comparative Neurology*, 177(2), 213–235. doi:10.1002/cne.901770204
- Tusa, R. J., Rosenquist, A. C. & Palmer, L. A. (1979). Retinotopic organization of areas 18 and 19 in the cat. *The Journal of Comparative Neurology*, 185(4), 657–678. doi:10.1002/cne.901850405
- van der Maaten, L. & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(86), 2579–2605.
- Van Rossum, G. & Drake, F. L. (2009). *Python 3 Reference Manual*. doi:10.5555/1593511
- Vogelstein, J. T., Packer, A. M., Machado, T. A., Sippy, T., Babadi, B., Yuste, R. & Paninski, L. (2010). Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology*, 104(6), 3691–3704. doi:10.1152/jn.01073.2009
- Wang, Z. & Bovik, A. (2009). Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1), 98–117. doi:10.1109/MSP.2008.930649
- Wang, Z., Bovik, A., Sheikh, H. & Simoncelli, E. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612. doi:10.1109/TIP.2003.819861
- Warrant, E. J. & Johnsen, S. (2013). Vision and the light environment. *Current Biology*, 23(22), R990–R994. doi:10.1016/j.cub.2013.10.019
- Waschke, L., Tune, S. & Obleser, J. (2019). Local cortical desynchronization and pupil-linked arousal differentially shape brain states for optimal sensory performance. *eLife*, 8. doi:10.7554/eLife.51501
- Webster, R. & Oliver, M. A. (2007, January). *Geostatistics for Environmental Scientists* (2nd ed.). doi:10.1002/9780470517277

- 
- Werblin, F. S. (1974). Control of retinal sensitivity. *Journal of General Physiology*, 63(1), 62–87. doi:10.1085/jgp.63.1.62
- Whiteway, M. R. & Butts, D. A. (2017). Revealing unobserved factors underlying cortical activity with a rectified latent variable model applied to neural population recordings. *Journal of Neurophysiology*, 117(3), 919–936. doi:10.1152/jn.00698.2016
- Wiesel, T. N., Hubel, D. H. & Lam, D. M. K. (1974). Autoradiographic demonstration of ocular-dominance columns in the monkey striate cortex by means of transneuronal transport. *Brain Research*, 79(2), 273–279. doi:10.1016/0006-8993(74)90416-8
- Woolrich, M. W., Behrens, T. E. & Smith, S. M. (2004). Constrained linear basis sets for HRF modelling using Variational Bayes. *NeuroImage*, 21(4), 1748–1761. doi:10.1016/j.neuroimage.2003.12.024
- Wu, W., Nagarajan, S. & Chen, Z. (2016). Bayesian machine learning: EEG/MEG signal processing measurements. *IEEE Signal Processing Magazine*, 33(1), 14–36. doi:10.1109/MSP.2015.2481559
- Yamanakkanavar, N., Choi, J. Y. & Lee, B. (2020). MRI segmentation and classification of human brain using deep learning for diagnosis of Alzheimer’s disease: A survey. *Sensors*, 20(11), 3243. doi:10.3390/s20113243
- Yang, Y., Pesavento, M., Chatzinotas, S. & Ottersten, B. (2018). Successive convex approximation algorithms for sparse signal estimation with nonconvex regularizations. *IEEE Journal of Selected Topics in Signal Processing*, 12(6), 1286–1302. doi:10.1109/JSTSP.2018.2877584
- Yavuz, E. (2012). *Source separation analysis of visual cortical dynamics revealed by voltage sensitive dye imaging* (Doctoral dissertation, Pierre and Marie Curie University, Paris).
- Zheng, L. & Yao, H. (2012). Stimulus-entrained oscillatory activity propagates as waves from area 18 to 17 in cat visual cortex. *PLoS ONE*, 7(7), e41960. doi:10.1371/journal.pone.0041960
- Zhu, G., Jiang, B., Tong, L., Xie, Y., Zaharchuk, G. & Wintermark, M. (2019). Applications of deep learning to neuro-imaging techniques. *Frontiers in Neurology*, 10. doi:10.3389/fneur.2019.00869



---

---

## Curriculum Vitae

Name	Kilian Leonard Heck
Date of Birth	14.01.1989
Place of Birth	Aschaffenburg

## Professional Experience

### Research Associate

2020 - 2023	Technische Universität Darmstadt, Dept. of Biology, Research Group <i>Systems Neurophysiology</i>
-------------	------------------------------------------------------------------------------------------------------

### Scientific Coordinator

2018 - 2020	Rhein-Main Universities (RMU), DFG GRK 2715/0: <i>Robust and Fast Data Processing for Real-Time Interrogation of Neuronal Circuits</i> , Speaker: Prof. Dr. Heinz Koepl, Technische Universität Darmstadt
-------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

### Research Assistant

2017 - 2018	Technische Universität Darmstadt, Dept. of Biology, Research Group <i>Systems Neurophysiology</i> .
2015	Otto-Friedrich Universität Bamberg, Chair for Economics, esp. Empirical Microeconomics.
2015	Leibniz Institute for Educational Trajectories, Bamberg, Working Unit <i>Methods, Weighting, and Imputation</i> .
2014 - 2015	Leibniz Institute for Educational Trajectories Bamberg, Pillar 5: Working Unit <i>Returns to Education</i> .
2013	National Educational Panel Study Bamberg, Pillar 5: Working Unit <i>Returns to Education</i> .
2012 - 2014	Otto-Friedrich Universität Bamberg, Chair of Sociology, esp. Methods of Empirical Social Research.

## Education

2018 - 2023	PhD Student, Supervision: Prof. Dr. Ralf Galuske, Dept. of Biology, Co-Supervision: Prof. Dr. Heinz Koepl, Dept. of Electrical Engineering and Information Technology, Technische Universität Darmstadt.
2014 - 2017	M.Sc. Survey Statistics, Otto-Friedrich Universität Bamberg.
2010 - 2014	B.A. Sociology, Otto-Friedrich Universität Bamberg.
2009 - 2010	B.Sc. Business Information Systems, Technische Universität Darmstadt.

---

## Relevant Publications

- 2021 | Korn, U., Krylova, M., Heck, K. L., Häußinger, F. B., Stark, R. S., Alizadeh, S., Jamalabadi, H., Walter, M., Galuske, R. A. W., Munk, M. H. J. (2021). EEG-microstates reflect auditory distraction after attentive audiovisual perception recruitment of cognitive control networks. *Frontiers in Systems Neuroscience*, 15.

---

## Acknowledgement

---

I would like to express my sincere gratitude to all those who have supported and contributed to the completion of this thesis.

First and foremost, I would like to extend my deepest appreciation to my supervisor, Prof. Dr. Ralf Galuske, for his unwavering trust, patience and advise throughout my journey into the field of neuroscience. I could always rely on his support, even during difficult life circumstances.

I am also grateful to Prof. Dr. Heinz Koepl for the co-supervision of this thesis, for the opportunity to have worked as his scientific coordinator, and the time he spent on many advisory meetings. My thanks also go to all the members of his Self-Organizing Systems lab, who have given me lots of inspirations and ideas in numerous lab meetings.

I would like to thank Dr. Daniel Hofmann for his contributions to a wide range of methodological aspects of this work, including his experiences about the experimental setup and data processing in context of optical imaging. Furthermore, our countless discussions on all kinds of techy stuff were greatly appreciated.

Also I am grateful to Dr. Mathias Peter for his input on various neurophysiological topics and for the many opportunities we had to chat about guitar equipment and music.

My heartfelt thanks go to my master student Philip Groß for his contributions to the simulation paradigm and for our time together as office buddies.

I would like to acknowledge all my dear colleagues from Prof. Dr. Galuske's working group, including Nico Grieser, Ute Korn, PD Dr. Matthias Munk, and Kirsten Wehner, as well as all bachelor and master students accompanying the group over the years. The familial environment and strong cohesion within the group undoubtedly set a high bar for all my future working environments. All the coffee and cake sessions as well as summer and christmas parties will be fondly remembered.

Special thanks go to Dr. Daniel Hofmann, Dr. Mathias Peter and Ute Korn for proof-reading this thesis.

Finally, I would like to express my gratitude to my family and friends for their encouragement and support throughout my academic journey.

---

## Erklärungen laut Promotionsordnung

### § 8 Abs. 1 lit. c PromO

Ich versichere hiermit, dass die elektronische Version meiner Dissertation mit der schriftlichen Version übereinstimmt.

### § 8 Abs. 1 lit. d PromO

Ich versichere hiermit, dass zu einem vorherigen Zeitpunkt noch keine Promotion versucht wurde. In diesem Fall sind nähere Angaben über Zeitpunkt, Hochschule, Dissertationsthema und Ergebnis dieses Versuchs mitzuteilen.

### § 9 Abs. 1 PromO

Ich versichere hiermit, dass die vorliegende Dissertation selbstständig und nur unter Verwendung der angegebenen Quellen verfasst wurde.

### § 9 Abs. 2 PromO

Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Darmstadt, 17.08.2023

---

K. L. Heck