



TECHNISCHE
UNIVERSITÄT
WIEN
Vienna | Austria



MEDIZINISCHE
UNIVERSITÄT WIEN

DIPLOMARBEIT

Assessing Deep Learning Methods for pRF Mapping of the Visual Cortex

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Technische Physik

eingereicht von

Florian Gisinger, B.Sc.

Matrikelnummer 01616234

ausgeführt am Atominstitut
der Fakultät für Physik der Technischen Universität Wien
in Zusammenarbeit mit dem Center for Medical Physics and Biomedical
Engineering der Medizinischen Universität Wien

Betreuung:

Em.Univ.-Prof. Dipl.-Ing. Dr.techn. Gerald Badurek

Assoc.-Prof. Priv.-Doz. Dipl.-Ing. Dr.techn. Christian Windischberger (MedUni
Wien)

Mitwirkung:

Dipl.-Ing. Dr. Michael Woletz (MedUni Wien)

Ass.Prof.in Dipl.-Ing.in Dr.in techn. Karin Poljanc

Wien, 11.02.2025

(Unterschrift Verfasser)

(Unterschrift Betreuer)

Acknowledgements

This thesis represents the conclusion of my studies in Technical Physics at TU Wien. As such, I want to express my appreciation to the people who supported me in successfully navigating this journey and especially this thesis.

I first want to express my deep gratitude towards Em.-Univ.Prof. Dipl.-Ing. Dr.techn. Gerald Badurek for agreeing to supervise this thesis. Without his support, I would not have had the chance to gain first-hand experience in medical physics and fMRI research, which I value greatly. I further want to extend my gratitude to Assoc.-Prof. Priv.-Doz. Dipl.-Ing. Dr.techn. Christian Windischberger for welcoming me into his lab and for offering corrections and suggestions to the text that have improved its quality immeasurably. To Dipl.-Ing. Michael Woletz, PhD I am indebted for his unwavering support during my time conducting this thesis as well as for offering corrections and suggestions to the text. I further owe thanks to Dipl.-Ing. David Linhardt, PhD for supporting me in every way possible while conducting this thesis. I am also grateful to Ass.Prof.in Dipl.-Ing.in Dr.in techn. Karin Poljanc for introducing me into the exciting world of medical physics through her seminar on the subject and for proofreading of the final text.

To Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Shuhei Yoshida and Assoc.-Prof. Dipl.-Ing.in Dr.in techn. Ille C. Gebeshuber, I am grateful for their supervision and mentorship during the two project works I conducted for my Master's degree, which have largely influenced my research interests.

To Dipl.-Ing. Philipp Maier, Dipl.-Ing.in Sarah Tobisch and Aida Mashaal, MSc, I am grateful for our continued friendship since the start of our studies at TU Wien. I also want to thank Johanna Bogner, BSc for her camaraderie and bright spirits in our shared workspace during the work on our respective theses.

To my parents, I am eternally grateful for supporting me throughout my years of study and for allowing me to follow my own interests in life. I am also immensely grateful to my partner for the loving support and goodwill offered during the more stressful times of conducting this thesis and for motivating me to give my best with it.

Finally, I am grateful to you, dear reader, for picking up this thesis and I hope it may prove useful and above all interesting to you.

English Abstract

Functional magnetic resonance imaging (fMRI) is a non-invasive method for mapping brain activation patterns. For investigating the representation of visual space on the visual cortex, population receptive field (pRF) mapping is a state-of-the-art fMRI approach. However, the fitting procedure employed by that technique is computationally expensive because of the typically large number of parameters that need to be fitted. In 2019, DeepRF, a method to accelerate the pRF fitting procedure using deep learning algorithms, was introduced, and results showed it to be able to achieve benchmark results at speeds more than three orders of magnitude faster than the conventional method. In this thesis, the machine learning model devised by the original authors of DeepRF was rebuilt as SnapRF to overcome limitations in the original approach and enable fair comparisons with a reference method. In doing so, improvements were introduced to the model in regards to user control and training procedures. Results of thorough tests and comparisons of SnapRF with standard analysis software have shown SnapRF to be capable of comparable performance at speeds two magnitudes faster than the reference method, albeit with results more prone to error. Possible pitfalls such as suboptimal visual stimulus design were identified and discussed. Additionally, a new noise model based on a principal component analysis (PCA) of resting-state fMRI data was introduced and shown to hold promise for future simulation and empirical studies. Importantly, this thesis shows that no "one-size-fits-all" approach is recommended in regards to DeepRF training parameters and instead, careful analysis and comparison of results is emphasized as the most important part of employing the method. Additionally, the thesis includes suggestions for further experiments and research directions that could facilitate the establishment of the DeepRF approach as a standard method in pRF mapping.

Deutsche Kurzfassung

Die funktionelle Magnetresonanztomographie (fMRT) ist eine nicht-invasive Methode zur Untersuchung von Hirnaktivitätsmustern. Population receptive field (pRF) mapping ist eine standardmäßig angewandte fMRT Methode, um die Repräsentation des visuellen Feldes im visuellen Kortex zu untersuchen. Die konventionelle Methode zur Optimierung des Mappings ist jedoch aufgrund der großen Zahl der Parameter rechnerisch aufwändig und langsam. DeepRF ist eine Methode zur Beschleunigung der Optimierung unter Zuhilfenahme von Deep Learning Algorithmen, die 2019 veröffentlicht wurde. Erste Ergebnisse zeigten, dass diese Methode den Prozess um mehr als drei Größenordnungen beschleunigen kann. In dieser Diplomarbeit wurde das DeepRF Modell als SnapRF nachgebildet, um Limitationen des ursprünglichen Zugangs zu überwinden und um faire Vergleiche mit einer Referenzmethode zu ermöglichen. Im Zuge dessen wurde das Modell in Bezug auf Kontrollmöglichkeiten und Training-Prozeduren erweitert. Die Resultate gründlicher Tests sowie Vergleiche mit der Referenzmethode zeigen, dass SnapRF vergleichbare Ergebnisse bei mehr als zwei Größenordnungen kürzeren Rechenzeiten erbringen kann. Allerdings zeigen die mit SnapRF gewonnenen Parameterabschätzungen höhere Abweichungen als die mittels Standardanalyse erhaltenen Ergebnisse. Mögliche Fallstricke an denen die Methode scheitern kann, wie etwa suboptimalem Design der visuellen Stimuli, wurden identifiziert und hervorgehoben. Weiters wurde ein neues Rausch-Modell eingeführt, das auf einer Hauptkomponentenanalyse (PCA) von resting-state fMRT Daten basiert. In dieser Arbeit wird gezeigt, dass dieses Modell hohes Potential für den Einsatz in zukünftigen Simulationen und Studien an empirischen Datensätzen hat. Ein besonders nennenswertes Ergebnis dieser Arbeit zeigt, wie wichtig ein auf spezifische Anwendungen maßgeschneiderter Zugang beim DeepRF-Zugang ist, da nur dann die Vorteile der Methode voll zu Tragen kommen. Um DeepRF als neue Standardmethode für pRF Mapping zu etablieren, müssten weitere Experimente und Untersuchungen angestellt werden, für die Empfehlungen in dieser Arbeit gegeben werden.

Contents

1	Introduction	1
1.1	Nuclear Magnetic Resonance (NMR)	2
1.2	NMR in Bulk Matter	6
1.3	Relaxation Mechanisms	8
1.4	Magnetic Resonance Imaging (MRI)	12
1.5	BOLD Contrast	15
1.6	Visual Processing in the Human Brain	16
1.7	Population Receptive Field (pRF) Mapping	19
1.8	Neural Networks and Deep Learning	21
2	Materials and Methods	27
2.1	mrVista	27
2.2	DeepRF and SnapRF	27
2.3	Simulated Data	32
2.4	Stimsim24 Dataset	32
2.5	NYU Retinotopy Dataset	34
3	Results	39
3.1	SnapRF Training Parameters and Testing Performance	39
3.2	Results on Simulated Data	41
3.3	Results on Stimsim24 Dataset	44
3.4	Results on NYU Retinotopy Dataset	48
4	Discussion	51
4.1	Results on Simulated Data	51
4.2	Results on Stimsim24 Dataset	53
4.3	Results on NYU Retinotopy Dataset	54
5	Conclusion and Outlook	57

1 Introduction

This thesis is based on the work conducted at the Center for Medical Physics and Biomedical Engineering of the Medical University Vienna. The aim was to test and expand the limits of currently available machine learning methods for analyzing functional magnetic resonance (fMRI) data, in particular the representation of the visual space in the human visual cortex. To that end, the **DeepRF** method developed by Thielen et al. in 2019 [1] was reimplemented as **SnapRF** and expanded with new functionalities and features.

DeepRF is a deep learning framework that is trained with simulated functional magnetic resonance imaging (fMRI) data to analyse responses of the visual cortex to visual stimuli. The original authors of the method showed an increase in analysis speed by over three orders of magnitude compared to standard analysis software. Such an increase represents a significant reduction in necessary computing power and thus might be able to save large amounts of energy. In addition, the approach holds promise for resolving two important issues the standard approach harbours that are related to the fitting of cortical responses to visual stimuli.

In particular, the following research questions were investigated:

- How does **DeepRF** compare to standard analysis software on simulated data with known ground truth?
- How reproducible are **DeepRF** results across multiple measurements of single subjects compared to standard analysis results?
- How does **DeepRF** compare to standard analysis software on empirical data from multiple subjects?

These research questions were tackled by comparing different implementations of **SnapRF** to the standard analysis software package **mrVista** on simulated data, in-house empirical datasets, and a large external empirical dataset. In these experiments, a clear focus was put on fairness of comparison and applicability under different experimental conditions.

This first chapter of the thesis aims to familiarize readers with the history and most important concepts underlying the thesis, and thus provide the necessary tools for appreciating its content and implications. The specifics of **DeepRF** and the extensions to it implemented in this thesis as **SnapRF** are detailed in chapter 2, together with descriptions of other software and the datasets used in the thesis. The results of the comparative studies conducted over the course of this thesis can be found in chapter 3, and their discussion in chapter 4. Finally, conclusions from the work performed in this thesis including recommendations for further directions of research are given in chapter 5.

1.1 Nuclear Magnetic Resonance (NMR)

The phenomenon of nuclear magnetic resonance (NMR) is the manifestation of a fundamental process that occurs in the interaction of nuclear magnetic dipole moments with external magnetic fields. As this section is meant to provide readers with a quick review of the most important topics for understanding this thesis, we shall only recount the basic facts and results here. Detailed derivations and precise descriptions of the processes involved in NMR can be found in references [2], [3] and [4].

Every elementary particle possesses spin. This vector-like quantity can be thought of as an intrinsic angular momentum. As described by quantum theory, spin values are not arbitrary, but instead quantized and their magnitudes may only take on multiples of the reduced Planck constant, $\hbar = \frac{h}{2\pi}$, where $h \approx 6.626 \cdot 10^{-34}$ Js. The absolute spin values of particles of the same type are constant and cannot be changed. The direction of the spin, however, is random until measured, a defining feature of quantum theory.

Importantly, the magnetic dipole moment of a particle is intimately related to its spin. For a spin- $\frac{1}{2}$ particle, such as a proton or an electron, the relationship between the two is given by

$$\boldsymbol{\mu} = g \frac{q}{2m} \boldsymbol{S}, \quad (1.1)$$

where g is a particle-specific constant, q the particle's charge, m its mass, and \boldsymbol{S} its spin. From here on out, bold letters shall represent vector-like quantities. Thus, it follows from Eq. (1.1) that the magnetic moment of a particle always points in the direction of its spin. Therefore, when describing the interaction of magnetic fields and particles, spin and magnetic moment go hand in hand and the respective terms are used interchangeably in the following discussions.

The peculiar behaviour of particle spin described above was first uncovered experimentally by Otto Stern and Walther Gerlach in the seminal Stern-Gerlach Experiment (SGE) of 1922 [5] via the deflection of a beam of silver atoms in an inhomogeneous magnetic field (Fig. 1). The inhomogeneous field used in the SGE was produced by an electromagnet with poles shaped as an edge and a groove. In

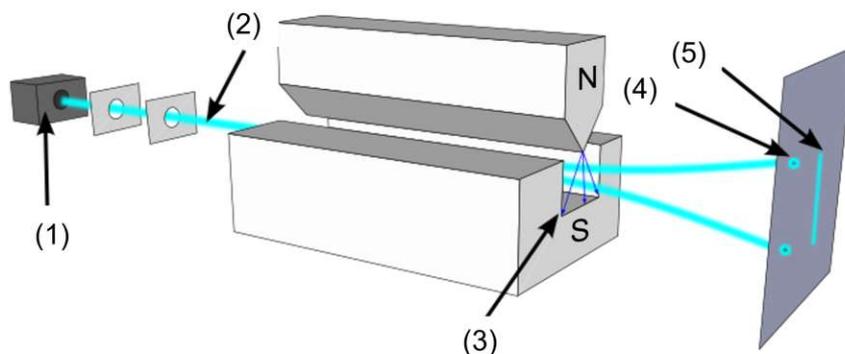


Figure 1: Experimental setup for the Stern-Gerlach Experiment (adapted from [6]). Silver atoms are emitted from a furnace (1) and are collimated into a beam of atoms (2) before entering an inhomogeneous magnetic field (3). Quantum mechanics predicts a splitting of the beam into two separate components (4) while classically a broadening of the beam on the detector screen is expected (5).

classical physics, where the atoms are thought of as small dipoles with random orientation, a broadening of the incoming beam of atoms by the inhomogeneous magnetic field would have been expected. What was shown, however, was the splitting of the incoming beam into two discrete bands, as predicted by quantum theory (Fig. 2).

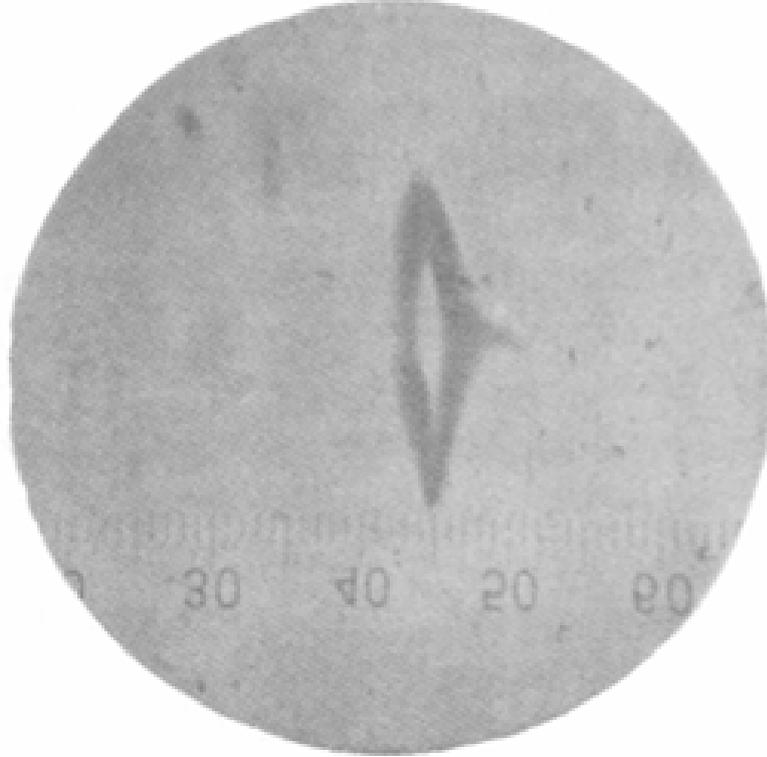


Figure 2: Experimental evidence for spin quantization of silver atoms in an inhomogeneous magnetic field [5]. In the scale below, 20 units are equal to 1 mm.

The asymmetry between the two bands observed in the result stems from the configuration of the poles, where atoms passing close to the edge of the upper magnet experienced a stronger force than those farther away. The meeting points of the bands likewise stem from the pole configuration, as the inhomogeneity is strongest in the center of such a field. The finite width of the bands is due to the thermal distribution of velocities of the atoms, as faster atoms spent less time in the field and were thereby deflected by smaller amounts than slower ones, leading to a broadening of the bands on the detector screen. At the time the SGE was conducted, spin was not known yet, and physicists at the time believed that the experiment had proven quantization of orbital angular momentum. However, as Otto Stern realized magnetic birefringence should have been observed as well if this was the case [7]. Therefore, this interpretation could not fully describe the features seen in the results, a problem which was resolved by the introduction of spin in 1925 [8].

The introduction of particle spin opened up a plethora of new and exciting things to probe: To study spin flips, variations of the SGE were employed with three consecutive magnetic fields [9, 10]. In these, the first and third field would split incoming beams, while the second would flip the spin of particles going through it after being deflected by the first field. Theoretical descriptions of these experiments [11][12] could not account for all of the observed spin flips, however. Isidor Isaac Rabi recognized the error lay in the fact that nuclear spin had been disregarded up

until then [13].

In composite particles, such as nucleons or atomic nuclei, the spins of the constituent particles add up. Even though Eq. (1.1) suggests no magnetic moment to be present for neutral particles (where $q = 0$), neutrons do in fact exhibit one, owing to their composite nature of three charged quarks. The Pauli exclusion principle states that two spin- $\frac{1}{2}$ particles of the same kind may not occupy the same state, i.e. may not be equal in all quantum numbers. This is the reason for the stability of atomic matter: In an atom, for a given energy level and orbital momentum, two bound electrons may only occupy the same state if their spins are antiparallel along a given axis. Such combinations of paired electrons do not exhibit magnetic dipole moments. That's why macroscopic paramagnetism, which results from coupling of electron magnetic dipole moments in a solid, only exists in materials with unpaired electrons in their constituent atoms. In the shell model of atomic nuclei, protons and neutrons pair off similarly to electrons in atoms, and so fill their respective energy shells. It thus follows that nuclear magnetic moments can only exist in nuclei with an odd number of protons, and/or neutrons.

In Rabi's insight not to disregard nuclear spin, he further realized that experiments could be used to directly measure not only the magnitude of nuclear magnetic moments, but also their relative orientation to the nuclear spin. He set out to do so and refined the three-stage SGE by considering the effects of an intermediate field oscillating in time with constant frequency. This line of research was undertaken on instigation by Cornelius Jacobus Gorter [14–16], who tried to discover NMR by a change of heat capacity of a substance in resonance with an oscillating field.

Earlier, we said that the absolute spin value of a particle is random along any axis until measured. What exactly constitutes a measurement has been the subject of much debate since the inception of quantum mechanics. What is undoubtedly understood, though, is that the presence of an external magnetic that is much stronger than the magnetic moment of a particle constitutes a measurement of the particle's magnetic moment along the external field axis. Therefore, the absolute spin value along the axis of the field is constant. The value along any other axis, however, remains random. Let the external field point along the z axis and be of the form $\mathbf{B}_0 = B_0 \mathbf{z}$, where \mathbf{z} is the unit vector in z direction. This external field may either be homogeneous or inhomogeneous (spatially varying). Before the presence of the external field, all spin orientations relative to \mathbf{z} are equal in energy. In an external field, however, this so-called degeneracy is resolved, so that the energy of a given particle is determined by the direction of its spin. Spins parallel to the external field have lower energy than those antiparallel to the field. This is known as the Zeeman effect. Magnetic dipole moments experience a torque in such a field, causing them to precess around the field axis in a phenomenon called Larmor precession (see Fig. 3).

A frequently invoked analogon for this behaviour is that of a spinning top in the earth's gravitational field, though a more intuitive one is a bicycle wheel gyroscope suspended from a string. Demonstrations of such a configuration can be found aplenty online, and readers are encouraged to search for them. The frequency of precession is called the Larmor frequency and is proportional to the acting external field. For our particle in the magnetic field it is given by

$$\omega_0 = -\gamma B_0. \quad (1.2)$$

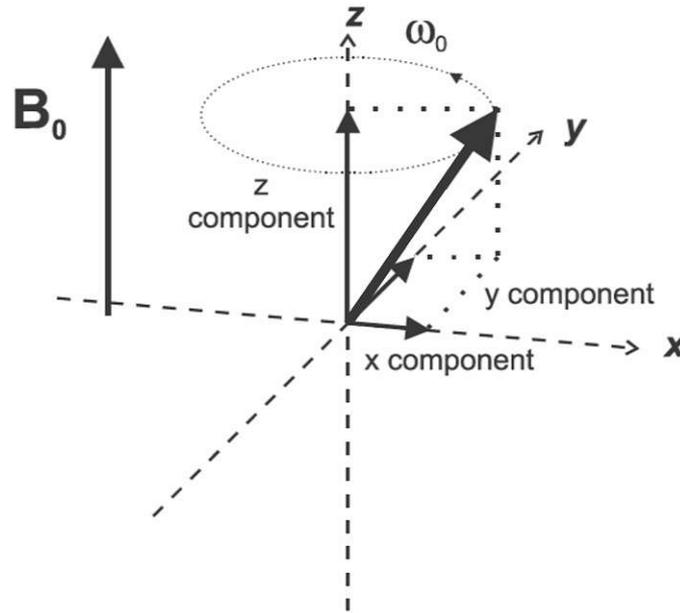


Figure 3: Larmor precession of a nuclear spin around an external field with precession frequency ω_0 (adapted from [17]).

The proportionality factor γ is called the gyromagnetic ratio of the particle carrying the dipole moment, and is given by

$$\gamma = g \frac{q}{2m}. \quad (1.3)$$

Recall that we already encountered this ratio in Eq. (1.1). Every isotope has a unique gyromagnetic ratio. It can be positive or negative, determining the sense of rotation in the precession. For external magnetic field strengths of the order of 1 T, the Larmor frequencies of NMR-active nuclei lie in the radiofrequency (RF) range (MHz).

Rabi first conducted theoretical investigations into oscillating fields applied to molecular beams in 1937 [18], and then experiments together with Jerrold Zacharias, Sidney Millman and Polykarp Kusch in 1938 and onwards [14, 19] (Fig. 4). When the frequency of the oscillating field is close to that of the Larmor precession of the nuclear magnetic moment of an atom in an external field, the resonance condition for absorption of energy from the oscillating field is fulfilled and the magnetic moment of the nucleus is changed, so that the atom is not refocused onto the detector screen. As the resonance condition is dependent on the Larmor frequency of spin precession in the inhomogeneous fields, there are two possibilities for achieving it: One can either keep the oscillating field at a constant frequency and vary the static fields slowly by changing the current through the electromagnets, or equivalently keep the static field strengths constant and vary the frequency of oscillation. Rabi and collaborators used the first method, while today it is common practice to use static external and varying oscillating fields. The first direct experimental evidence for NMR can be seen in Fig. 5. At the right combination of field strength and oscillation frequency, a sharp drop in refocused particles is visible.

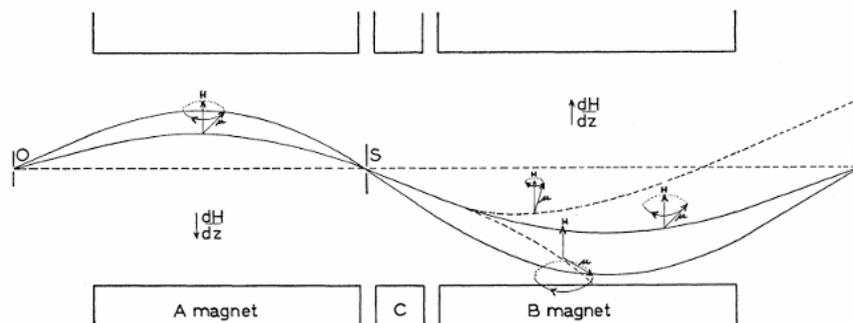


Figure 4: Schematic of Rabi's magnetic resonance method setup [19]. Magnets *A* and *B* create inhomogeneous magnetic fields with anti-parallel gradients, magnet *C* creates the oscillating field. The solid curves represent paths of molecules without change of magnetic moments in the apparatus. The dashed curves represent paths of molecules whose magnetic moments were changed in the excitatory field of the *C* magnet. Such molecules are not refocused onto the detector.

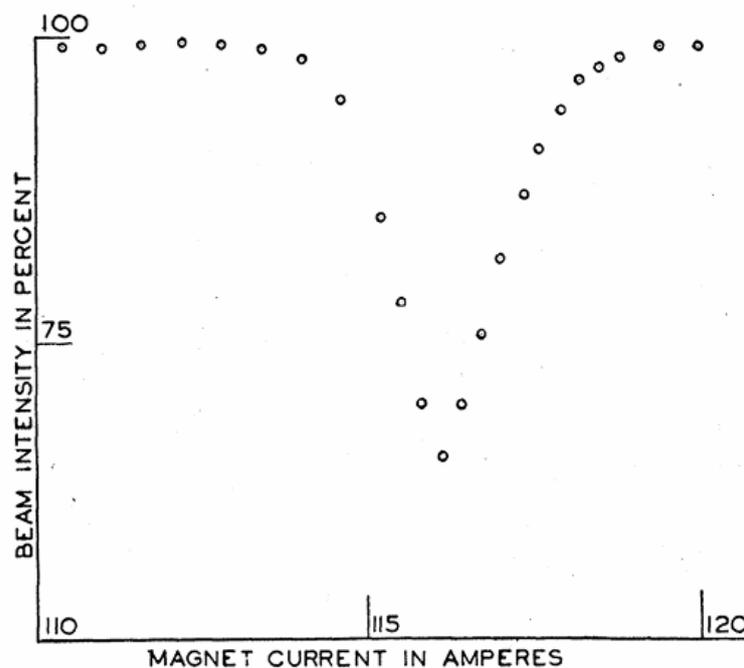


Figure 5: First observation of nuclear magnetic resonance [14]. At the resonance condition, a sharp drop in beam intensity at the detector is visible, as molecules are not refocused onto the detector screen.

1.2 NMR in Bulk Matter

The observation of NMR in non-gaseous matter was first reported in 1946 independently by Edward Purcell [20] and Felix Bloch [2, 21]. To appreciate this effect, let us consider a sample of 1 g of water in an external magnetic field $\mathbf{B}_0 = B_0 \mathbf{z}$ pointing in z direction. Water molecules contain two hydrogen nuclei, which each contain one proton, resulting in two nuclear magnetic moments per molecule. In contrast, the most common isotope of Oxygen, ^{16}O , does not feature a nuclear magnetic moment. Let S_z be the z -component of a nuclear spin. If S_z is positive, i.e. parallel to the field, the nuclei's energy is lower than it would be with the opposite,

antiparallel orientation. We shall denote the former state as $|1\rangle$, the latter as $|2\rangle$. Their energies, E_1 and E_2 thus obey $E_2 > E_1$. In the system's ground state, which is defined as a state with minimum energy, all nuclear spins occupy state $|1\rangle$. In thermal equilibrium at room temperature, however, a significant portion of hydrogen nuclei will occupy state $|2\rangle$. The ratio between their occupancies is related to their energy difference. This difference, ΔE , is given by

$$\Delta E = E_2 - E_1 = \frac{\hbar\omega_0}{2} - \left(-\frac{\hbar\omega_0}{2}\right) = \hbar\omega_0. \quad (1.4)$$

According to the Boltzmann distribution, the ratio of occupancy is

$$\frac{n_1}{n_2} = \exp\left(\frac{\Delta E}{k_B T}\right), \quad (1.5)$$

where n_i is the number of nuclei in state $|i\rangle$, k_B is the Boltzmann constant and T the system's temperature. For protons at room temperature ($T = 300$ K) in an external field of 1 T, the difference in occupancy is only about seven in a million. Still, as there are around 6.7×10^{22} hydrogen nuclei in 1 g of water, this small majority of nuclei in state $|1\rangle$ is enough to result in an appreciable net magnetization \mathbf{M}_0 of the sample in direction of the external field. The net magnetization is static, however, and is therefore not directly measurable.

To measure the magnetization, and thereby the nuclear magnetic moment of the hydrogen nuclei, the system must be brought out of equilibrium. This is achieved by applying an oscillating RF magnetic field, \mathbf{B}_1 along a direction perpendicular to \mathbf{B}_0 . Recall that if the oscillation frequency is close to the Larmor frequency of precession of the nuclear spins, energy from \mathbf{B}_1 can be absorbed by the nuclei, thereby bringing the system out of thermal equilibrium.

In the experiments by Rabi et al. described in the previous chapter, the oscillating field acted only briefly on the particles, whereas now we shall investigate the effects of longer application. Through absorption, nuclei from state $|1\rangle$ will be excited to state $|2\rangle$. In addition, nuclei in state $|2\rangle$ will be stimulated to emission of radiation at the same frequency, leading to more absorption. Therefore, the net magnetization along z will decrease. Another effect induced by the RF field in the nuclear spin system is that of phase-coherence. Originally in the external static field, all spins precess around the field direction with the same frequency. Their phases, however, are randomly distributed. The field \mathbf{B}_1 , oscillating with the Larmor frequency of precession, forces the spins to precess in phase. This can be understood by analogon of children on swings: Suppose two children of equal weight are swinging side by side on swings of the same length, and therefore with the same frequency, but different phases. If simultaneous pushed were supplied to both children at the resonance frequency, both will soon swing in phase (Fig. 6). In NMR, this phenomenon is called forced precession. The consequence of these effects is a rotation of the axis of precession for all phase-coherent spins, which leads to a tipping of the net magnetization vector towards the traverse plane. The changing orientation of the net magnetization can be measured, as a varying magnetic field induces a voltage in a suitable receiver coil.

In a frame of reference that is rotating with the Larmor frequency around the z axis, the individual spins do not precess, but are stationary. In this picture, the RF field is static, so that the nuclear spins align along an axis that is a superposition

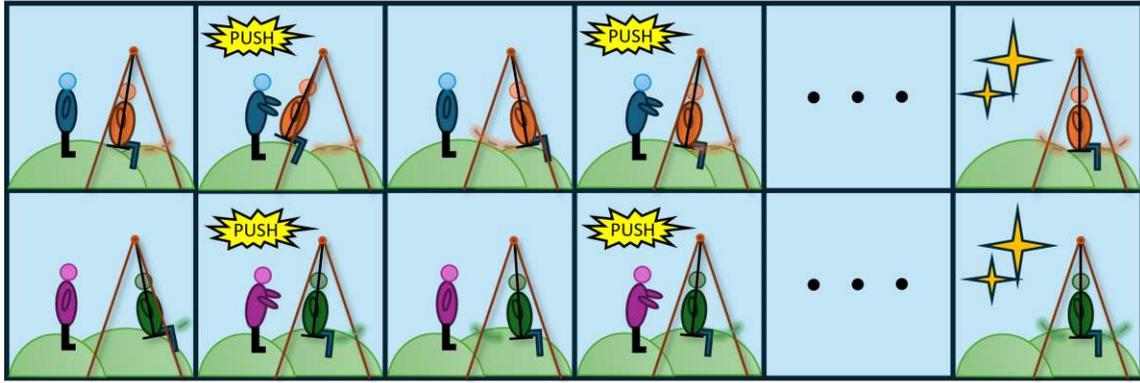


Figure 6: Establishment of phase coherence between children of swings through pushes applied at the resonance frequency.

of the two external static field axes. Thus, the tipping of the magnetization can be understood in a straightforward way (Fig. 7).

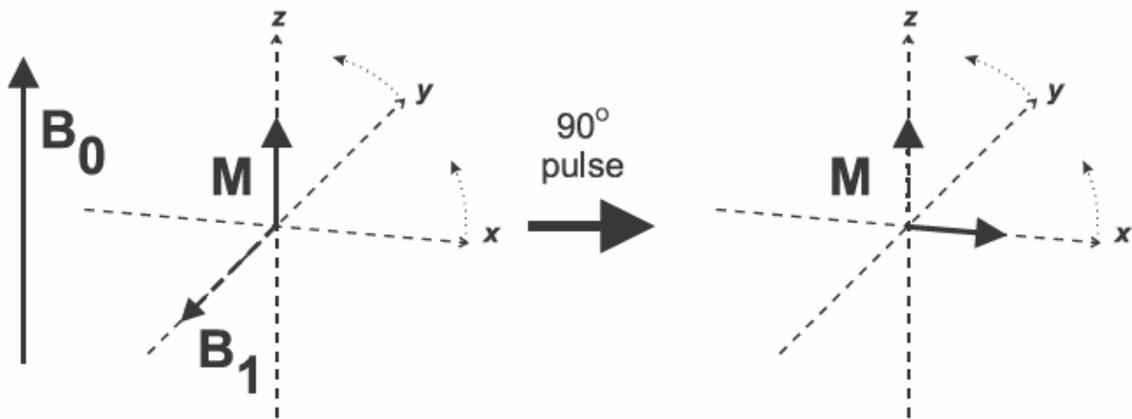


Figure 7: Tipping of the net magnetization vector in a coordinate system rotating with the Larmor frequency [17].

Mathematically, the behaviour of the magnetization vector can be described with the Bloch equation first introduced in [2]. The angle by which the magnetization tips is dependent on the time the RF field is active and its strength, and is called the flip angle, defined by

$$\varphi = \gamma \cdot B_1 \cdot \Delta T. \quad (1.6)$$

It is therefore possible to define pulse lengths that correspond to specific flip angles for specific nuclei, such as $\frac{\pi}{2}$ -pulses that flip the magnetization 90° , or π -pulses that flip the magnetization 180° .

1.3 Relaxation Mechanisms

In the previous discussions, a few assumptions were made without explicitly mentioning them. Let us now make these explicit: First we assumed uniform external fields throughout our system. In practice, magnetic field uniformity within a sample of inhomogeneous and anisotropic matter is very difficult to achieve, and perfect uniformity an impossibility. Therefore, the Larmor frequency is not constant throughout the system, but varies in space. Second, we assumed external fields as

the only sources changing the orientation of nuclear spins. For this to hold, the assumptions of negligible contributions from atomic electrons, neighbouring nuclei and thermal effects must be made. In practice, all of these are invalid, which manifests in relaxation processes.

After deviation from equilibrium and cease of the excitatory RF field, the nuclear spin system will naturally regain thermal equilibrium. Thus, the net magnetization will realign itself in the direction of \mathbf{B}_0 . Various processes contribute to this relaxation, and they are most commonly described by the longitudinal relaxation rate T_1 and the transverse relaxation rate T_2 .

The establishment of the magnetization in z direction (the direction of the static external field) is described by an exponential growth (Fig. 8). This growth occurs both when a sample is placed in a static external magnetic field and when a nuclear spin system in an external field is brought out of equilibrium by a resonant RF field. The growth of net magnetization in direction of \mathbf{B}_0 can be written as

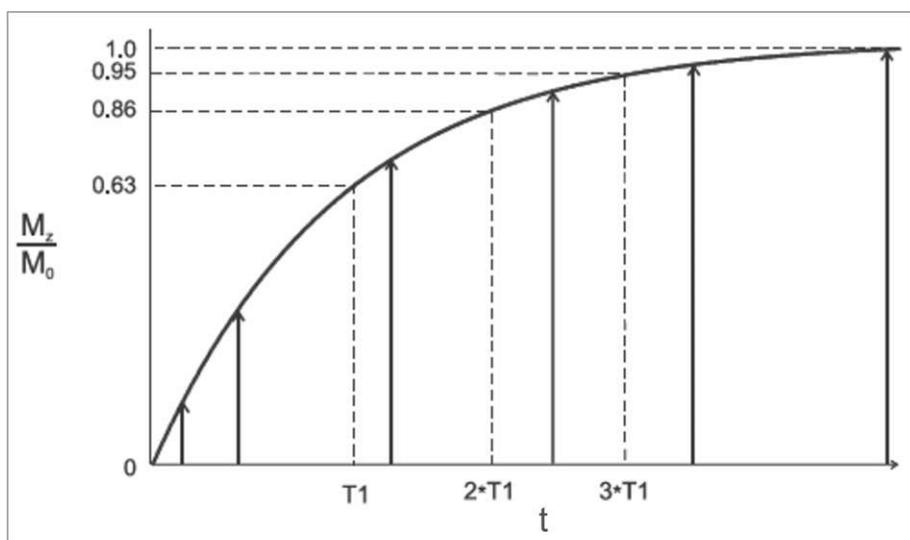


Figure 8: Relaxation or recovery of longitudinal magnetization M_z after a $\frac{\pi}{2}$ excitation pulse of the resonant RF field or after exposure to a static external magnetic field, respectively [17].

$$M_z(t) = M_0 \left(1 - \exp\left(-\frac{t}{T_1}\right) \right), \quad (1.7)$$

where \mathbf{M}_0 points in direction of \mathbf{B}_0 as described earlier. From Eq. (1.7) it can be seen that T_1 is defined as the time after which the longitudinal magnetization has regained $1 - \frac{1}{e} \approx 63\%$ of its equilibrium value.

T_1 can then be measured by a method called inversion recovery (IR): Suppose the excitatory RF field is deactivated at a time $t = 0$ after a π -pulse that flipped the net magnetization into $-z$ direction. Applying a $\frac{\pi}{2}$ -pulse some time after the initial π -pulse results in a signal that is proportional to the strength of magnetization in z direction at that time. Repeating the experiment multiple times with different intervals between the pulses allows calculation of T_1 . The associated microscopic processes that enable the observed relaxation are collectively called T_1 relaxation, longitudinal relaxation or spin-lattice relaxation. The latter term indicates that the energy of excited nuclei is dissipated to their surroundings (the "lattice"). Processes

that contribute to this relaxation include collisions and electromagnetic interactions between molecules [2].

Let us next consider the decay of transverse magnetization M_{xy} after a $\frac{\pi}{2}$ -pulse, which can be described by an exponential decay of the form

$$M_{xy}(t) = M_{xy}(0) \exp\left(-\frac{t}{T_2^*}\right). \quad (1.8)$$

It follows that T_2^* is defined as the time it takes for the transverse component of the magnetization to decay to $\frac{1}{e} \approx 37\%$ of its value $M_{xy}(0)$ just after disturbance from equilibrium (excitation). This decay is called the free induction decay (FID). Transverse relaxation occurs through fundamentally different processes than longitudinal relaxation. T_2^* can be decomposed into different contributions:

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_2'}. \quad (1.9)$$

Here, T_2 describes contributions from spin-spin interactions that lead to dephasing after the forced precession, while T_2' describes effects from external field inhomogeneities, which also lead to a loss of phase coherence, as locally varying field strengths lead to spatial variation of the Larmor frequency. T_2 can be measured independently of T_2^* by making use of a phenomenon termed spin-echo, which was first described in 1950 [22]. After excitation, the transverse magnetization decays over time due to statistical relaxation effects (T_2) and local field inhomogeneities as described above (T_2'). In the spin-echo approach a second pulse is applied, this time however with a flip angle of π . This second pulse flips all spins along the transversal axis and thus compensates for any phase changes from field inhomogeneities. This leads to a resurgence of transverse magnetization which is referred to as a "spin echo" (Fig. 9). Improvements to the method can be achieved if the

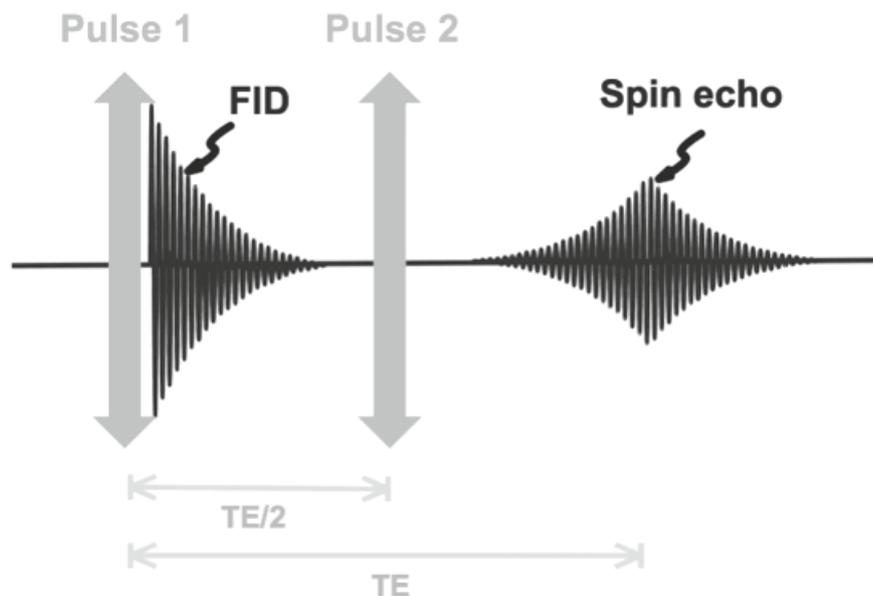


Figure 9: Occurrence of free induction decay after one $\frac{\pi}{2}$ pulse and a spin echo after a second rephasing pulse that reverses dephasing due to magnetic field inhomogeneities (adapted from [23]).

pulses are applied along perpendicular transverse axes, as this reduces the influence

of diffusion effects on the measurements of T_2 [24]. The so-avoided effects of local B_0 inhomogeneities mainly stem from field distortions near interfaces between tissues of different magnetic susceptibilities and the presence of paramagnetic atoms or molecules.

A similar technique to that of spin echoes described above is that of gradient echoes, which were first reported on in 1960 [25]. These can be achieved by applying a dephasing magnetic field gradient after an initial RF pulse, which is then reversed so that the decoherence due to the applied inhomogeneous field is reversed, resulting in an echo signal in a pickup coil (Fig. 10).

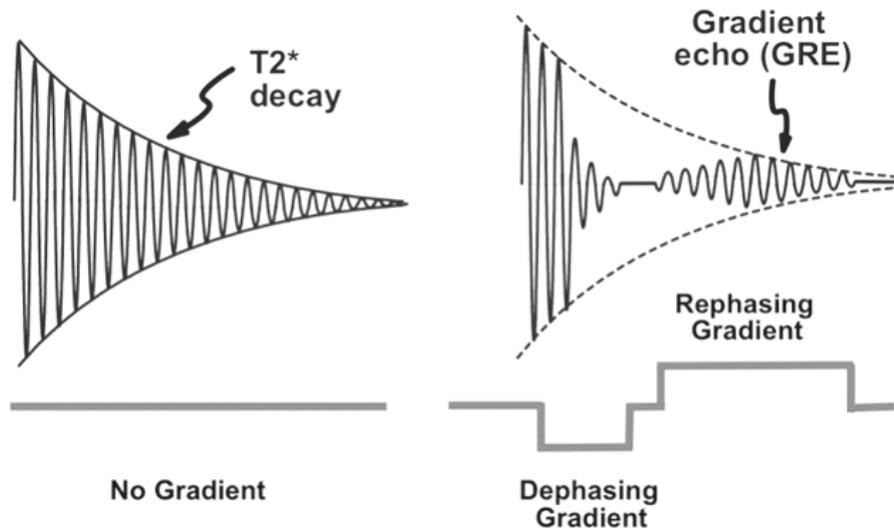


Figure 10: Emergence of a gradient echo after reversing a dephasing magnetic field gradient (adapted from [26]).

By creating multiple echoes (spin or gradient) in a row, T_2^* can be measured accurately as the later echo intensities get progressively smaller solely due to T_2^* relaxation effects (Fig. 11).

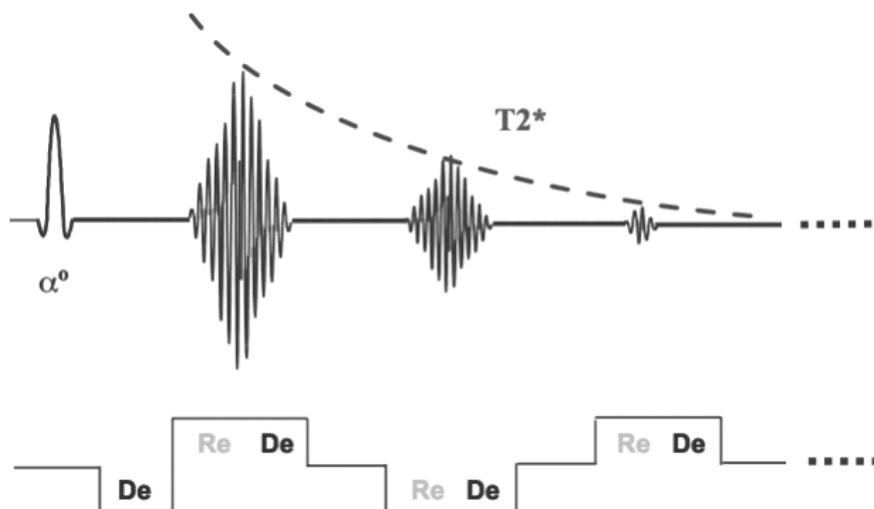


Figure 11: Measurement of T_2^* with a train of gradient echoes created through application of dephasing (De) and rephasing (Re) gradients after excitation with an RF pulse with a flip angle of α (adapted from [27]).

In general, T_1 and T_2 of a given sample of matter are not the same, and $T_2 \leq T_1$. Importantly, T_1 and T_2 cannot be given for a single spin, but always refer to an ensemble of many spins, as they are time constants of statistical processes. In human biological tissues, typical values for T_1 are between 0.3 and 2 seconds, while typical T_2 values are below 0.3 seconds [28].

1.4 Magnetic Resonance Imaging (MRI)

The concept of magnetic resonance imaging (MRI) was first described in 1973, and already then the potential for imaging biological systems was recognized [29, 30]. Today, the technique is widely used in basic brain research and diagnostic medicine. Because of the abundance of water in our bodies, the nucleus most widely used for imaging purposes is that of the hydrogen atoms in water molecules. To produce an image instead of a measurement of bulk magnetization, the signal from all hydrogen nuclei in the body has to be collected in such a way as to be sensitive to position. The most straightforward way of doing so is introducing a linear magnetic field gradient to the external static field, so that the external field \mathbf{B} spatially varies in strength from one end of the MRI scanner to the other. Thereby, the Larmor frequency of the nuclei of interest varies in that direction as well, so that nuclei inside specific slices of the body in the scanner perpendicular to the field axis can be excited selectively. This ensures that the signal acquired in the imaging process can only originate from this slice. To achieve a second dimension of imaging, another magnetic field gradient is applied during measurement of a specific slice perpendicular to the slice-encoding gradient. This additional gradient is called the frequency encoding gradient and alters the Larmor precession of nuclei depending on their position in the selected slice along the direction of the gradient. Finally, to achieve three-dimensional localization, a third gradient is applied perpendicular to the previous two during measurement. This gradient is called phase-encoding and features an amplitude varying in time. While the gradient is active, it affects the phase of the precessing nuclei within the band delineated by slice and frequency encoding gradients. The phase shifts acquired by the nuclei persist after the phase encoding gradient is deactivated, which allows localization after the nuclei return to equilibrium, where they precess with the Larmor frequency of the other two gradient fields, but locally varying phase shifts (Fig. 12). Measuring the same frequency-encoded band of voxels for multiple phase shifts allows the delineation of a three dimensional region of space, which is commonly called a voxel. Typical MRI systems are able to achieve voxel sizes of below 1 mm^3 . Additional cues for the localization of voxels can be extracted from the signal strength, as receiver coils close to a given voxel receive a higher signal intensity from that voxel than coils with a greater distance to the voxel [31].

MR images of a given slice are not recorded in three dimensional real space, but instead in a space spanned by spatial frequencies, which is called k-space [33]. There exists no point-to-point mapping between the real space and k-space, as every point in k-space contains information on all voxels in the real space. Transformation of signals from one space to the other is possible via Fourier transformations. To understand this relationship, consider the decomposition of a complex sound wave into a spectrum of amplitudes at different frequencies via a Fourier transform. In the same way, a complex image can be decomposed into a spectrum of planar wave amplitudes at different spatial frequencies. Central parts of k-space correspond to

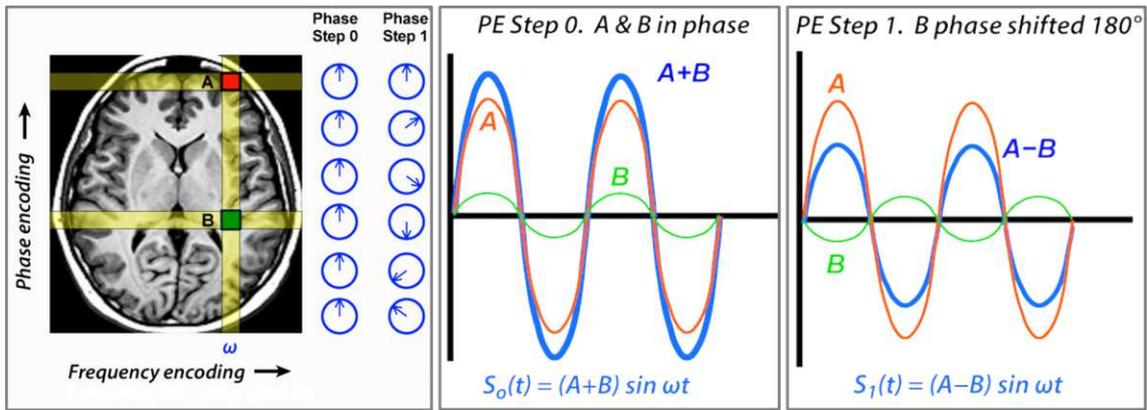


Figure 12: Schematic representation of how overlapping frequency and phase encoding magnetic field gradients in a selected slice define three-dimensional voxels of space that can be selectively measured (adapted from [32]). At step 0, the phase encoding gradient is not active, and the resulting measured signal $S_0(t)$ is a linear combination of the signals from voxels A and B . At step 1, the phase of voxel B was shifted by 180° relative to voxel A by the phase encoding gradient, and the measured signal $S_1(t)$ is the difference between the two individual signals. The combination of measurements allows the calculation of individual voxel signals.

low spatial frequencies. These contain information about image contrast in the real space. High spatial frequencies, on the other hand, correspond to edges and finer features in the image and are associated with points in the periphery of k-space (Fig. 13).

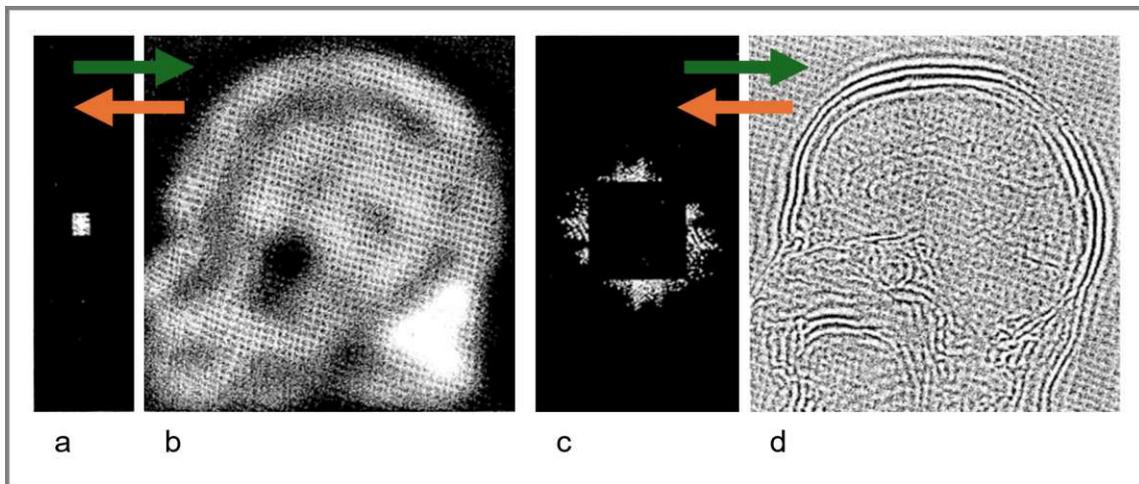


Figure 13: Central parts of k-space (a) encode general features like contrast in images (b), while peripheral parts (c) encodes details but little contrast (d) (adapted from [33]). Arrows represent Fourier transforms between k-space and real space and vice-versa.

The order of when and how RF pulses, magnetic field gradients and readouts are applied or performed is called an imaging sequence. In 1977, echo planar imaging (EPI) was introduced [34] as a measurement sequence for achieving short acquisition times. It is still in use today and works by first exciting a given slice and then periodically switching the polarity of a strong frequency encoding gradient. Between readouts, a weak phase encoding gradient is activated for a short time. The combined

effect of these modulations is the quick traversal of k-space, where the frequency encoding repeatedly sweeps from one edge to the other and the phase encoding "jumps" the readout up another line of the slice each time it is active, so that the readout is performed in a rectangular motion (Fig. 14). Measurement of one slice therefore requires only one RF pulse, as the signals measured are formed by gradient echoes resulting from the traversal process. This enables acquisition times of ≤ 100 ms per slice. Today, it is possible to excite multiple slices at the same time with specially formed RF pulses which reduces the overall measurement time even further. The slices so measured in parallel can be decomposed using the receiver characteristics of different coil elements in phased array coils [35].

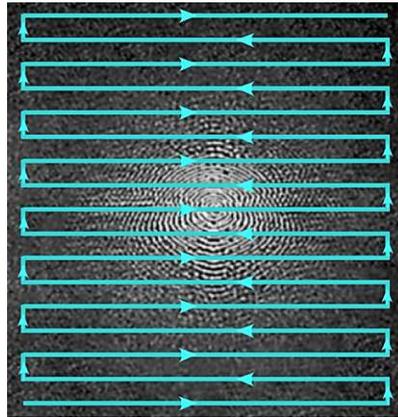


Figure 14: Rectangular motion through k-space in the echo planar imaging (EPI) technique [36].

In chapter 1.3, we encountered T_1 and T_2 relaxation mechanisms that govern the decay of forced precession and the return to thermal equilibrium after disturbance by an RF pulse. We discussed T_1 and T_2 differ from one tissue to another and how they can be measured. This has proven useful for producing images by making the image contrast dependent on relaxation time, to enable differentiation of different tissues and discover abnormal regions for diagnostic purposes (Fig. 15).

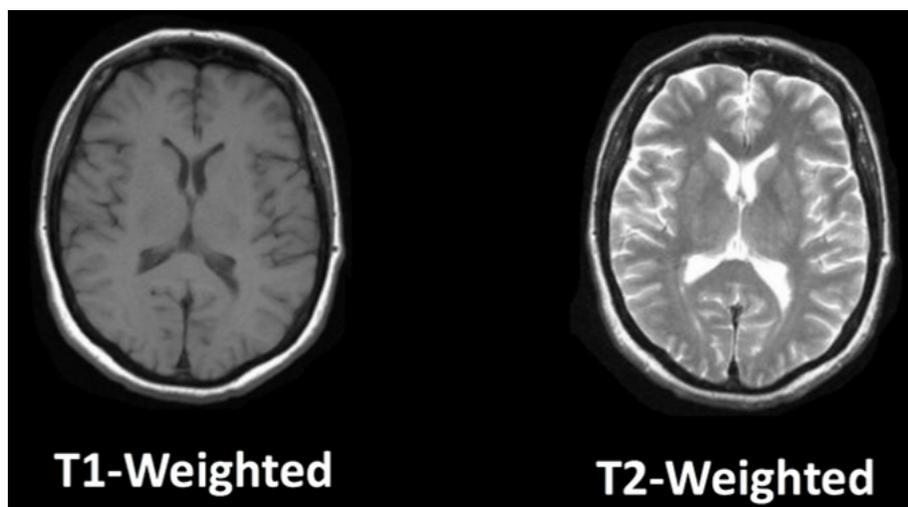


Figure 15: T_1 and T_2 weighted images of the same brain slice [37].

1.5 BOLD Contrast

Functional magnetic resonance imaging (fMRI) has enabled researchers to investigate the processes occurring in the brain when subjects are performing a given task, such as grasping something. It has further allowed the identification of connections between different areas within the brain from a functional point of view, instead of a purely anatomical one.

Brain activity is accompanied by changes in cerebral blood flow (CBF) and cerebral blood volume (CBV). When a subject is presented with a stimulus, such as a picture or a sound, the stimulus will be processed in certain areas of the brain. At sites on the cortex where information about the stimulus is processed, neuronal activity is evoked and arterial blood flow is increased. This relationship was investigated by means of positron emission tomography (PET) in the 1980s, where it could be shown that more arterial blood is supplied to an active region than needed, even though the spatial resolution of the method was inherently limited [38–40].

To make the changes in CBF visible with MRI, the magnetic properties of blood itself had to be leveraged. Recall that we encountered T_2^* in chapter 1.3 and saw it is influenced by magnetic inhomogeneities. These inhomogeneities may arise from external fields, interfaces between tissues with different susceptibilities, or local atomic or molecular magnetic fields. Paramagnetic molecules are a major source of the latter. One approach of visualizing CBF with MRI was to make use of a paramagnetic MRI contrast agent that would dissolve in blood, thereby enabling the study of the distribution and flow of blood via a series of EPI images [41–46]. In such measurements, accumulation of the contrast agent in a tissue (through an increase in blood flow) would enhance relaxation of nearby protons, thereby increasing relaxation-weighted MRI signals. Another idea was using the intrinsic magnetic properties of red blood cells. Already in 1936, the different magnetic properties of oxygenated and deoxygenated hemoglobin were shown [47]: While the former is diamagnetic, the latter is paramagnetic. Venous blood, which has a higher concentration of deoxygenated hemoglobin than arterial blood, therefore shortens T_2^* relaxation. As we have just discussed, however, arterial blood flow to a region on the cortex in response to its activation is increased by more than necessary to replenish the site with oxygen, so that the concentration of oxygenated hemoglobin increases. Thus, the signal in a T_2^* weighted image increases after neuronal activation. The first evidence of this effects in NMR was published in 1982 [48]. Eight years later, it was shown that these effects, which were named blood oxygenation level dependent (BOLD) contrast, could be used to make the change of CBF following neural activation directly visible [49]. A conclusive interpretation of the BOLD signal as stemming from neuronal activation was achieved in 2001 [50]. BOLD contrast is to this day the standard method of investigating task-related brain activity and functional brain connectivity. Using EPI, sections of the brain can be repeatedly imaged with a short repetition time (TR), making it possible to record a three-dimensional "movie" of changes in CBF.

The increase in blood flow does not happen instantaneously, but instead follows the so-called hemodynamic response function (HRF), which is characterised by a maximum about five seconds after stimulation and a dip below baseline after the signal decrease. This can be understood as an oversupply of nutrients to an active region and a reduction of supply after the need of the region has been met. It can be modelled in a number of ways, though the most common is the so-called canonical

HRF given by a difference between two gamma distributions (Fig. 16):

$$HRF_{canonical} = g(x, p_1, w_1) - a \cdot g(x, p_2, w_2), \quad (1.10)$$

where p_i refers to the peak position and w_i to the width of the respective gamma distribution

$$g(x, p, w) = \frac{x^{p-1} e^{-x/w}}{\Gamma(p) \cdot w^p}. \quad (1.11)$$

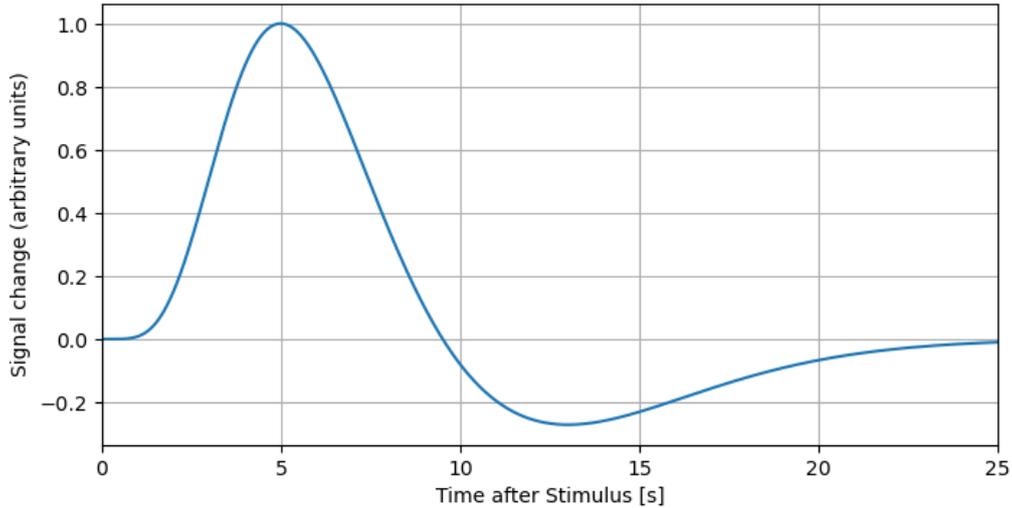


Figure 16: Normalized canonical hemodynamic response function (HRF) that models the increase in bloodflow to an active brain region following a short stimulus.

Standard analysis of fMRI data always requires a model of the response evoked by the task at hand. Usually, assumption of linear relationships between input stimulus and neuronal activity and between neuronal activity and the measured response is valid [51–53]. In such a model, the neuronal activity can be inferred via deconvolution of the input stimulus and the output response. fMRI measurements are not absolute, so instead the brain activity evoked by a task is compared to a baseline measurement, so that the difference between the two states can be given in percent signal change or arbitrary units.

1.6 Visual Processing in the Human Brain

The human visual cortex is located on the back of our brain, in the occipital lobe. Visual input travels from the retina through the optic nerve and the lateral geniculate nucleus (LGN) to the primary visual cortex (V1) (Fig. 17). This area is situated in and around the calcarine sulcus, one of the deepest inward folds of the brain. There, a direct mapping from visual space to neurons exists, in which the input from the left hemifield is processed on the right hemisphere, and input from the right hemifield is processed on the left hemisphere. Additionally, the neurons on the cortex are organised in such a way that neighbouring neuronal populations respond to neighbouring sections of the visual field. This property, called retinotopy, was first uncovered in the late 1870s from experiments on dogs and monkeys [55]. It was found that removal or injury of certain regions of the cerebral cortex of those

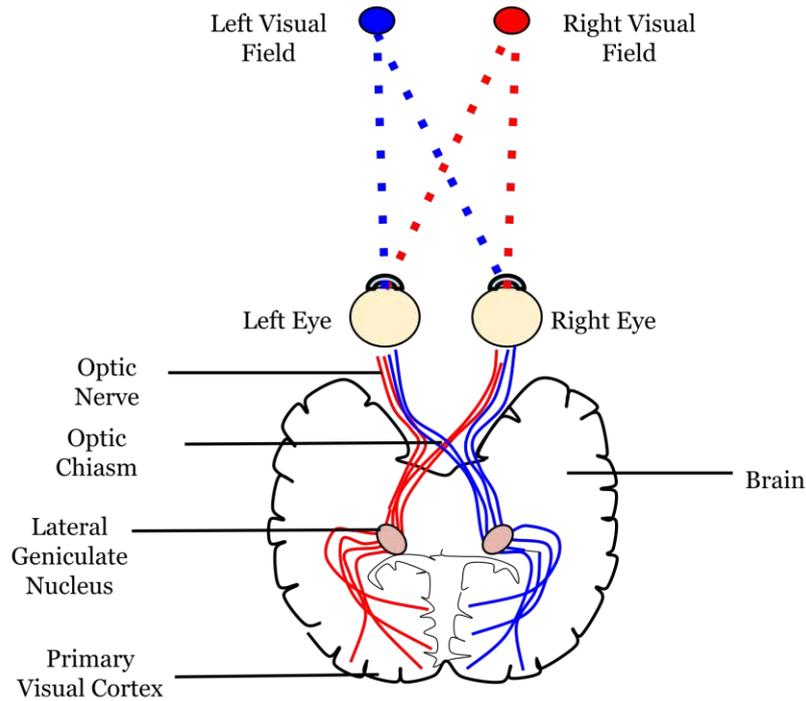


Figure 17: The visual pathway in humans, shown from the retina to the visual cortex [54]. Note the contralateral endpoints for left and right hemifields, respectively.

animals resulted in the loss of processing of visual input. By studying the effects of incomplete removal of these areas, which we now know encompassed the visual cortex, it was shown for the first time that a direct mapping from visual field to the cortex exists [56] (Fig. 18). Further, evidence for cortical magnification was found in the fact that more of the cortex area encoded for the central part of the visual field than for similarly sized areas of the peripheral visual field [57].

The first evidence for retinotopic organisation of the visual cortex in humans came from studies of brain lesions suffered by soldiers in the Boxer uprising of 1900 and the Russo-Japanese war of 1904-1905 [58], as well as from those fighting in World War I [59, 60]. From these investigations, the first maps of the representation of visual field on the human primary visual cortex were constructed (Fig. 19).

Further understanding of the visual system was gained from animal studies conducted on primates, most notably macaques. In these studies, responses could be measured at the single neuron level via implanted electrodes, enabling the delineation of receptive fields in visual space to which single neurons respond to. The receptive field of a neuron is defined as a region in visual space in which a visual stimulus evokes a response in that neuron.

The applicability of fMRI to investigating the cortical representation of visual space was first demonstrated with so-called travelling wave experiments [61, 62]. In these, participants viewed a periodic stimulus of expanding concentric rings on a screen and brain activity in the visual cortex was continuously recorded. The results showed a periodic wave of neural activity, corresponding to the periodic activation pattern with a constant phase-shift in time (Fig. 20). Over the years since, fMRI became a state-of-the-art method for mapping the neuronal organization of the visual cortex.

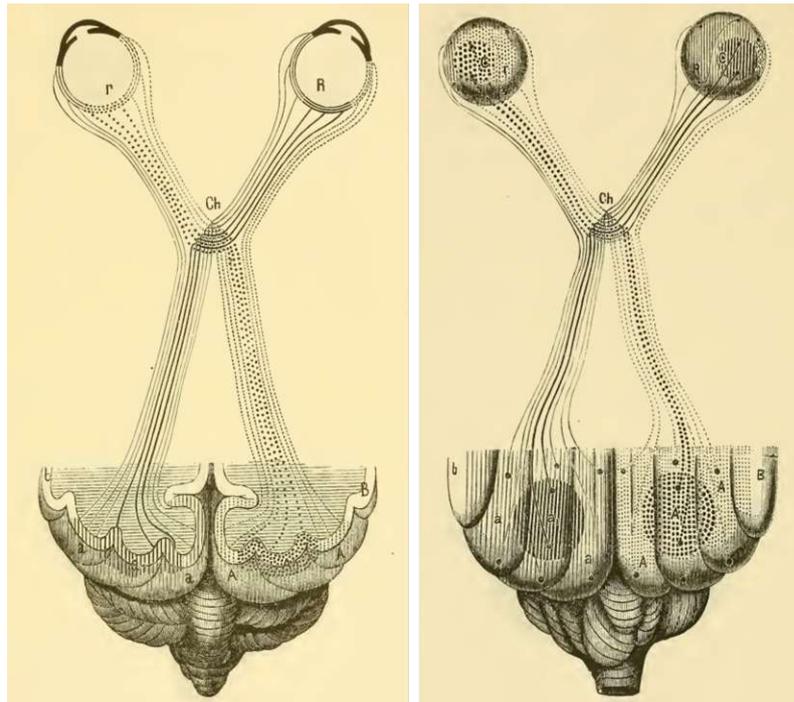


Figure 18: Retinotopic maps of a dog created by Munk in the late 1870s [56]. The maps show the reorganization of the optic nerves at the optic chiasm (Ch) and the contralateral endpoints for input from the left and right hemifields, respectively. Additionally, the retinotopic organization of visual space on the visual cortex is shown.

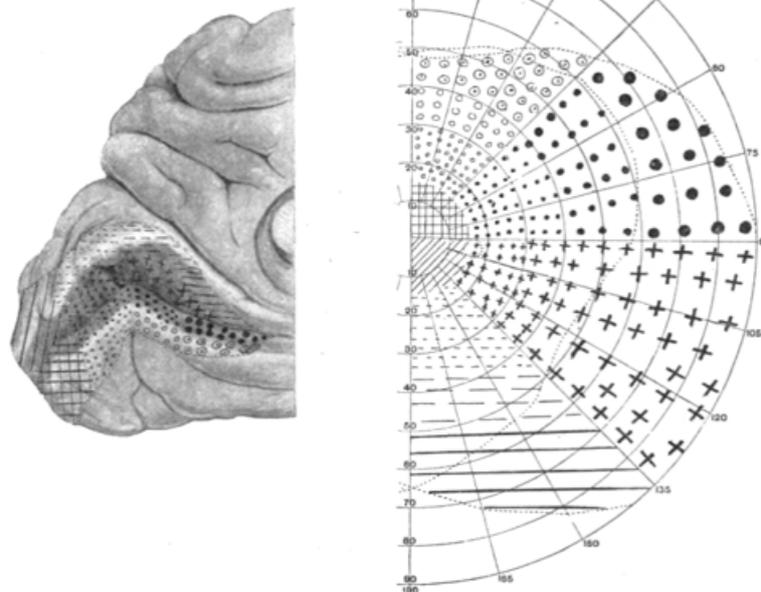


Figure 19: One of the first retinotopic maps in humans, created by Holmes in 1944 from lesion studies of soldiers [60]. The retinotopic organization of the visual cortex is shown both in eccentricity and polar angle for the right visual hemifield.

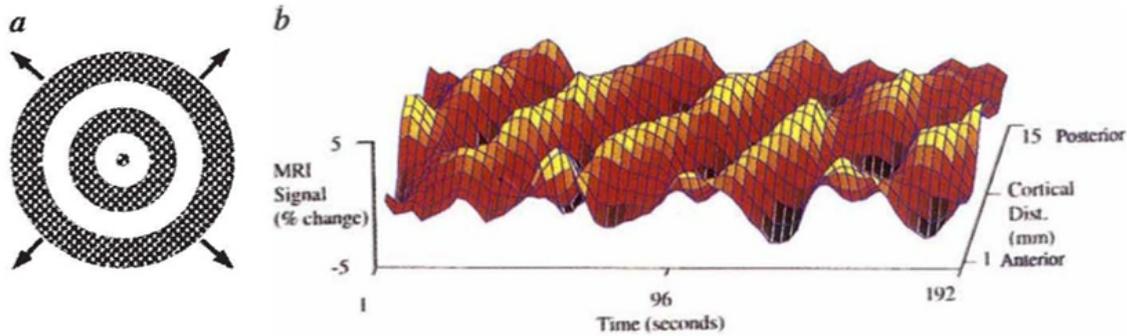


Figure 20: Periodic stimulus (a) evokes observed wave of neural activity (b) [63].

When combining such travelling wave experiments with a second set of measurements obtained with a rotating wedge stimulus, the measured responses could be analysed to extract those stimulus positions in the visual field that produced the greatest signal per voxel [64, 65]. However, as each voxel contains a multitude of neurons that each have their own receptive field, the position in visual space that elicits the greatest overall signal for a voxel does not describe the extent of the cortical representation of visual space fully. To take this into account, a population receptive field (pRF) must be introduced that corresponds to the extended area of visual space that a population of neurons in the visual cortex encodes [66].

1.7 Population Receptive Field (pRF) Mapping

In 2008, pRF mapping was introduced as a method for estimating the retinotopic organization in the visual cortex and has since become the standard approach for investigating this property [65].

The standard assumption for the shape of a pRF is a two dimensional Gaussian with two parameters for center position (x , y) and one for its width (σ), or equivalently two parameters for eccentricity (ecc) and polar angle (φ) and one for width (σ). The standard method of pRF mapping works by assuming a general linear model (GLM) to describe the fMRI measurement signal of a voxel as a sum of a predictable response to a stimulus and measurement noise. The original mapping procedure consists of a two-step fitting process, where at first a rough grid search is performed for every second voxel, in which its timecourse is compared to a large host of computer-generated timecourses. In the second step, a finer fit is performed by applying an optimization algorithm to voxels where more than some threshold (usually 10-15 %) of variance could be explained in the first step, with the previously found rough parameter estimates as starting points. In these fitting procedures, the residual sum of squares (RSS) between prediction and measured fMRI timecourse is minimized. Since its first publication, the method has been extended in numerous ways, either to include more complex pRF shapes or other fitting procedures [67–69].

In pRF mapping experiments, subjects are instructed to fixate a point on a projector screen in the middle of their visual field. To track attention levels, the point changes colour periodically and subjects are instructed to push a button each time they recognize the colour changed. While subjects are fixating their gaze, high-contrast flickering stimuli move across the screen. The stimuli can be any shape, though the two forms most often used are (a) a combination of a rotating wedge and an expanding ring and (b) sweeping bars that move from side to side with some

rotation applied to their direction after each pass (Fig. 21).

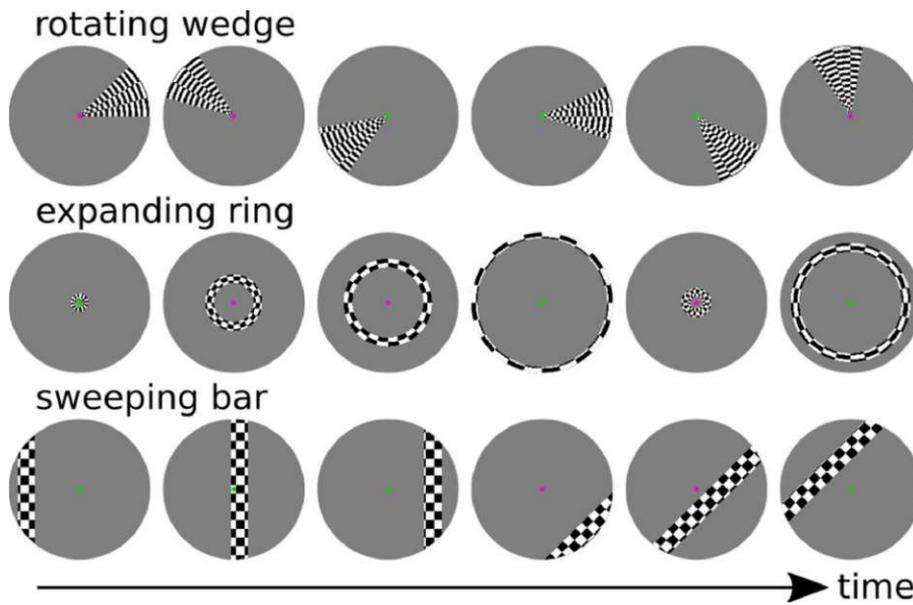


Figure 21: Typical stimuli used in pRF mapping experiments [70].

In the original design, the bar stimulus is tilted by 45° after each pass of the visual field and its direction reversed: First it sweeps from left to right (L-R), then from down right to up left (DL-UR), followed by a sweep from up to down (U-D), and so on (Fig. 22). After each diagonal sweep, a blank period is inserted. After eight sweeps and a final blank period, a full run is complete. Thus, each quadrant of the visual field is stimulated eight times in this design. The blank periods help in the analysis, as they reduce overlap of two consecutive stimulations. The location and shape of pRFs can then be inferred from the acquired fMRI timecourses by knowledge of the stimulation paradigm, i.e. the temporal and spatial order of the stimulation events.

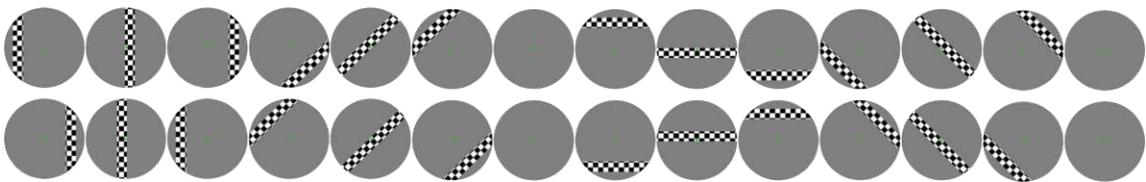


Figure 22: Sweeping bar stimulus used in pRF mapping experiments (adapted from [70]).

An important criticism of traditional pRF mapping is the fact that the two-step fitting procedure may miss local RSS minima for voxels by first discarding every second timecourse. Further, in measuring the goodness-of-fit as the RSS between prediction and measurement, a bias towards fitting noise rather than signal may be introduced in measurements with low contrast-to-noise ratios (CNRs), as the method is highly sensitive to outliers in the data (Fig. 23).

In 2019, DeepRF was proposed as a new technique for pRF mapping without the limitations of the original method, based on a deep learning procedure that greatly improves the speed of pRF mapping [1]. DeepRF is the focus of this thesis and is described in more detail in chapter 2.2.

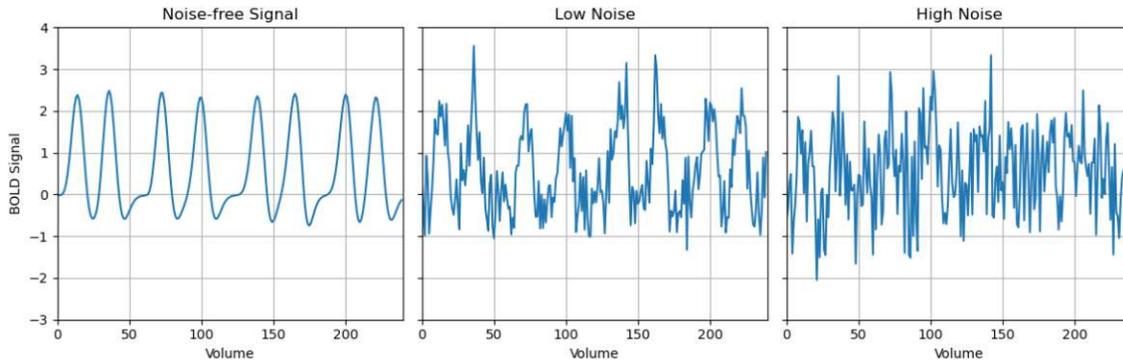


Figure 23: Simulated timecourses for the same voxel with different noise levels. Note the peaks introduced by the higher noise that may be fitted as signal.

1.8 Neural Networks and Deep Learning

Today, machine learning, neural networks, deep learning and artificial intelligence are widely known and recognized terms. This chapter aims to call the most important aspects of these interconnected topics into mind, so as to enable readers a comprehensive understanding of the methods used in this thesis.

Machine learning (ML) describes any procedure in which a computer program is trained to perform a given task on some unseen data. Such a task may be as conceptually simple as distinguishing pictures of cats from pictures of houses or as complex as safely steering a car through traffic. If a program is able to successfully perform the task on new datasets it was not trained upon, it is said to be able to generalize. It is useful to think of the output of such a program in terms of a model with free parameters that need to be fitted to a given task via a training procedure. Commonly, data in ML is described in terms of features and labels: when we see a cat, we assign it the label "cat" based on a collection of features - a certain size, certain facial features, the presence of a tail, specific sounds it produces, and so on. Correct categorization is typically the goal of ML. The fitting process for a ML model can be formulated as the minimization of an objective function, often called the loss function. This point of view - with models, data and loss - naturally leads to the notions of underfitted and overfitted models: An underfitted model may not have enough free parameters to accurately capture the details of all necessary features to make good predictions for labels, or it may instead not have had enough datapoints in training. Conversely, an overfitted model may have too many free parameters, so that it eagerly learns all features of the training data, but cannot generalize to other data with different features. The aim of training a model is to strike a balance between these two extremes. Strategies that are employed to prevent over- and underfitting are generally known as regularization techniques.

For training a model, there are a number of training paradigms to choose from, and none is optimal for every end. They can broadly be categorized into supervised and unsupervised learning strategies. In supervised learning, the data used for training has known labels which the model should learn to predict based on the features of the datapoints. In unsupervised learning, no labels are supplied to the model, so instead the model must learn to find some underlying structure in the features to sort the data. Even though there exist a multitude of other strategies in ML, these two represent the most widely used and important ones for classification

tasks.

All models used in this thesis are trained with supervised learning, and make use of so-called training and validation datasets in training: Models learn relationships between features and labels from training data, and are subsequently tested on unseen validation data. This procedure is repeated until the predictive error on the validation data converges. The performance on training and validation data over time can be used as a way to monitor overfitting: If the predictive error on the training set decreases, but increases on the validation set, overfitting has occurred. A technique to prevent this is called early stopping, which halts the training process if a given condition is met, for example if the validation error starts rising. A complete step of training and validation stages is called a training epoch.

Artificial neural networks (ANNs) are a type of model used in ML applications. They are constructed from layers of artificial neurons that are interconnected. If connections are only present in forward direction, i.e. from layers closer to the input to layers closer to the output, we speak of a feedforward network. Layers that are not directly connected to either the input or the output of the model are called hidden layers. The architecture of ANNs has been heavily inspired by the neuronal organization of the brain, and there exist many similarities between them.

At its most basic implementation, an artificial neuron performs a weighted summation of inputs and applies a nonlinear function to the result (Fig. 24). In bio-

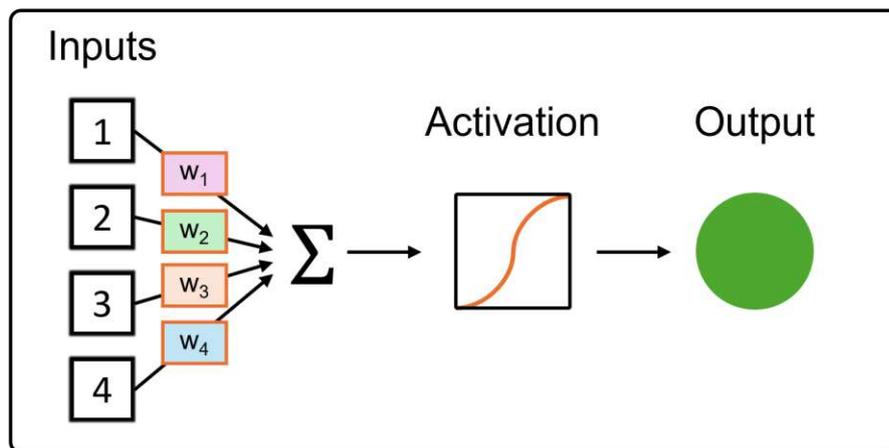


Figure 24: Artificial neuron that applies a nonlinear function to a weighted sum of inputs.

logical neurons, dendrites conduct electrical signals to cell nuclei, from where action potentials may be activated to send a signal via an axon to a synapse, where the signal can be transmitted to further neurons (Fig. 25). In artificial neurons, the dendrites can be identified with the weighted connections. In terms of the ML model from before, the weights applied to inputs by individual artificial neurons are the free parameters to be adjusted in training. The activation function of artificial neurons is the nonlinearity, and its output is forwarded to any other neuron it is connected to, just as in the brain. Artificial and biological neurons may also include a negative bias or activation threshold that must be overcome for a neuron to “fire”, i.e. to send an output signal. It was shown in 1989 that feedforward ANNs with a nonlinear activation function and as few as a single hidden layer (Fig. 26) act as universal approximators for piecewise linear functions [71, 72]. How well a given function can be approximated depends on the number of artificial neurons.

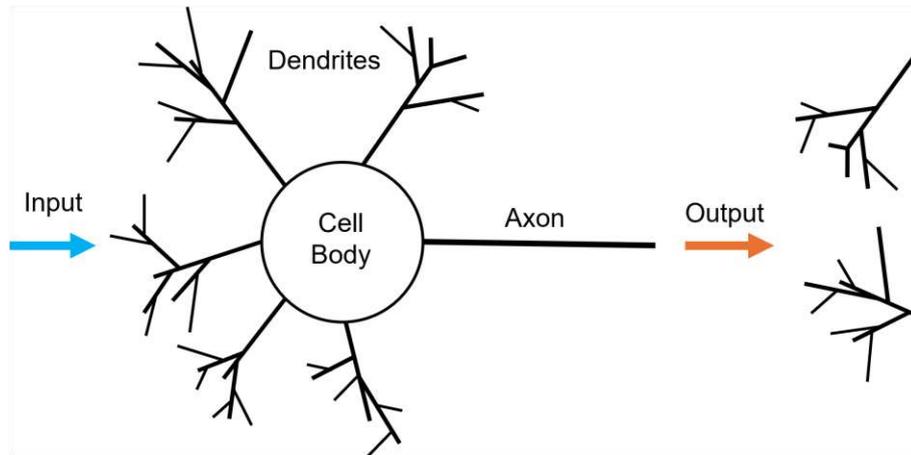


Figure 25: Schematic representation of a biological neuron in the brain.

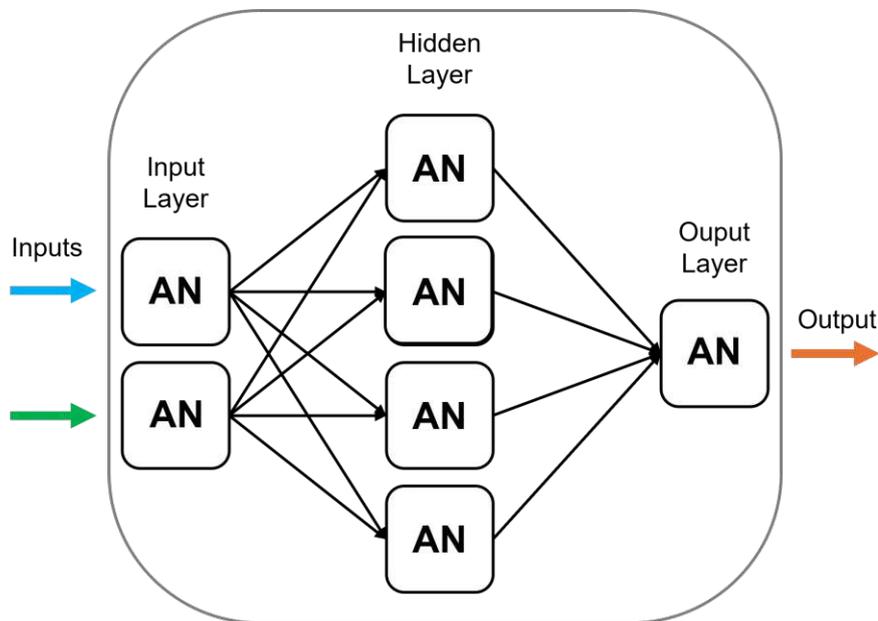


Figure 26: Artificial neural network (ANN) constructed from layers of artificial neurons (ANs), with one input layer, a single hidden layer, and one output layer.

Convolutional neural networks (CNNs) are a class of ANNs that have traditionally found application in image recognition tasks. Their most important feature is the presence of convolutional layers, where multi-dimensional inputs are convolved with a filter kernel that moves over the input (Fig. 27). For regions on the edge of an image, where the filter would not have enough values for the convolution process, values are inserted in a procedure known as padding (Fig. 28). From the plethora of padding options, the most commonly used are repeating the outermost values of the array or introducing constant values along the edges. Since the filter moves over the whole input, neighbouring artificial neurons in the next layer receive input from neighbouring sections of the input, akin to the property of retinotopy in the visual cortex (chapter 1.6). The inputs of neighbouring neurons overlap because of the design of the filter, which enables the recognition of connected features. The overlap is determined by the filter width and the stride length employed for traversing the input. In this way, the convolution kernel acts akin to a population receptive field when the input is two-dimensional (chapter 1.7). The second core feature of CNNs

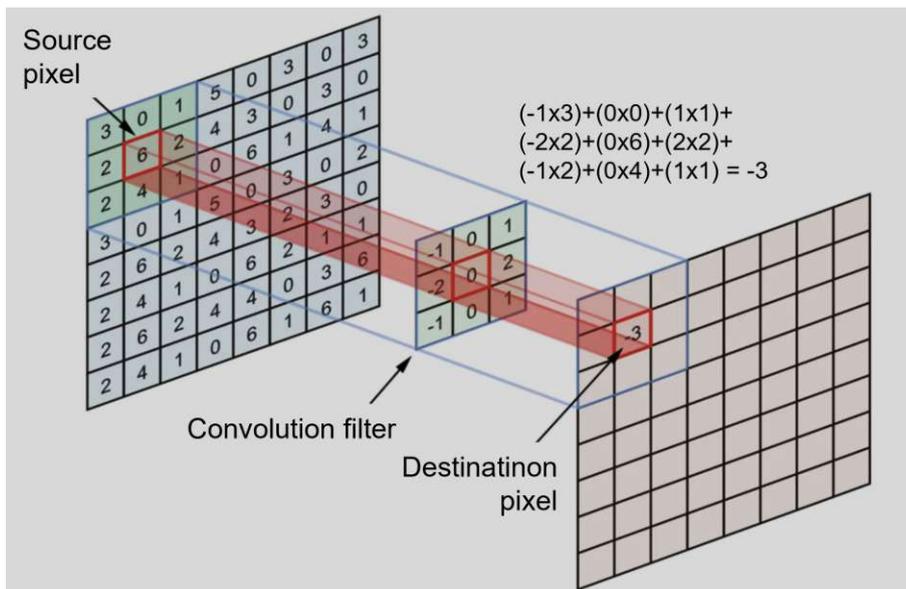


Figure 27: Schematic representation of a layer with a convolutional kernel filter in a convolutional neural network (CNN) (adapted from [73]).

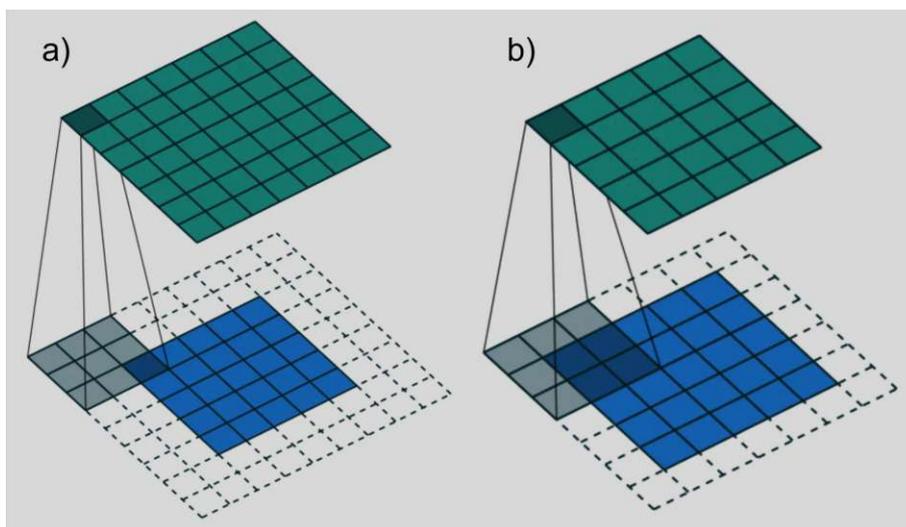


Figure 28: Schematic representation of padding before convolution in a CNN (adapted from [73]). Padding can be used to increase the image size (a) or to keep it the same (b) after convolution with a filter kernel.

is the existence of pooling layers, which reduce the input size. In this way, redundancy in the input is reduced and the model is made more resilient to variations in the input. The most commonly used pooling strategies are max pooling, where the maximum value of a subsection of the output of the previous layer is projected onto a smaller area of the next layer, and average pooling, where instead the average of the subsection is used (Fig 29).

In CNNs, the filter kernels themselves are optimized in training. Therefore, the layer weights and the values of the individual convolutional filter kernels are the model parameters on which the loss function depends.

Up until relatively recently, the training of deep CNNs (DCNNs), where the number of hidden layers is greater than about ten, has been notoriously difficult [74]. In such network architectures, a number of non-trivial problems may arise,

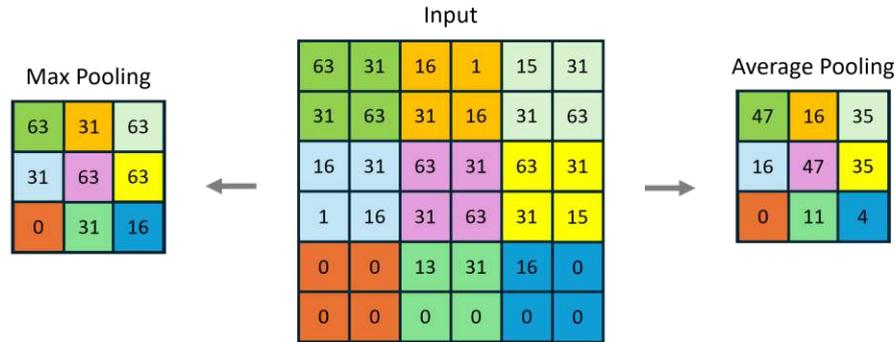


Figure 29: Schematic representation of pooling operations in a convolutional neural network (CNN).

the most pressing of which are connected with the calculation of gradients in training. Gradient-based approaches in training are commonly used as they allow the incremental adjustment of parameters: Every parameter is adjusted by an amount proportional to the derivative of the loss function with respect to that parameter. The proportionality is called the learning rate of a training procedure, which is one of several so-called hyperparameters in training.

The process of adjusting parameters in ANNs is most commonly done via backpropagation of error, which is equivalent to the chain rule of derivation: The output of later layers depends on the output of earlier layers. Therefore, the partial derivative of the loss function with regards to the output of any layer depends on how later layers affect the output. In backpropagation, the partial derivatives of the loss function with regards to individual parameters are thus calculated from the latest layer backwards to avoid calculating single derivatives multiple times. Due to the process of multiplying derivatives, early layer weights may receive smaller and smaller updates for classical sigmoid activation functions like $\tanh(x)$, where the gradient is in the interval of $[-1, 1]$. If the resulting gradients get too small, learning halts and no meaningful progress can be made. This is referred to as the vanishing gradient problem. If some parameters are too large instead, their updates may exponentially increase, leading to the so-called exploding gradient problem. Today, both of these problems can be largely avoided by using other activation functions like the rectified linear unit (ReLU), defined as $f(x) = \max(0, x)$ (Fig. 30), normalized weight initialization schemes and batch normalization of input data [75–78]. DCNNs with these improvements have shown to reach or even surpass human level performance on some image recognition tasks [79].

As these problems were tackled, another problem emerged: Through various experiments, it became clear that adding more and more layers led to saturation effects in model accuracy, with worse performance of deeper models than shallower ones [80]. This behaviour, known as the degradation problem, is not obvious from a theoretical standpoint, as additional layers could in principle just act as identity mappings if the model already found a local minimum of the loss function with fewer layers. This, however, is not observed.

One of the first successful approaches to solving the degradation problem was the introduction of residual neural networks (ResNets) in 2016, which feature feedforward identity mapping connections between individual layers (Fig. 31) [81]. Therefore, only a residual mapping must be fitted in training these models, and vanishing

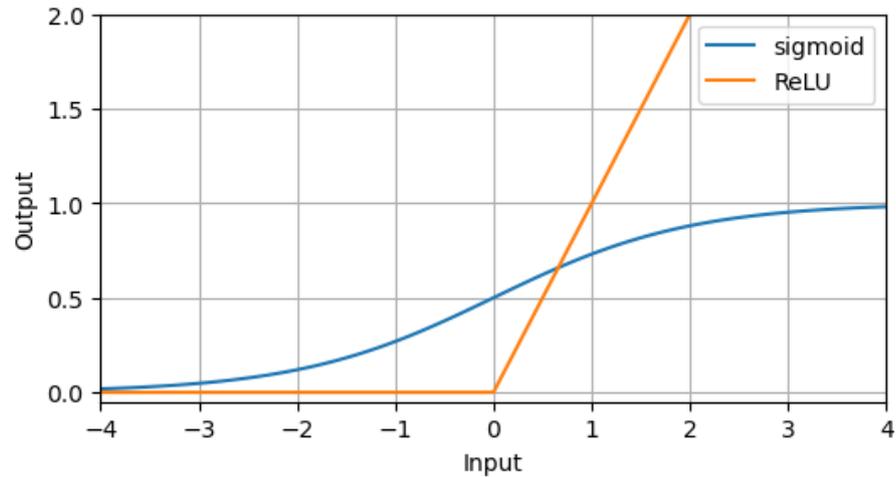


Figure 30: Different nonlinear activation functions.

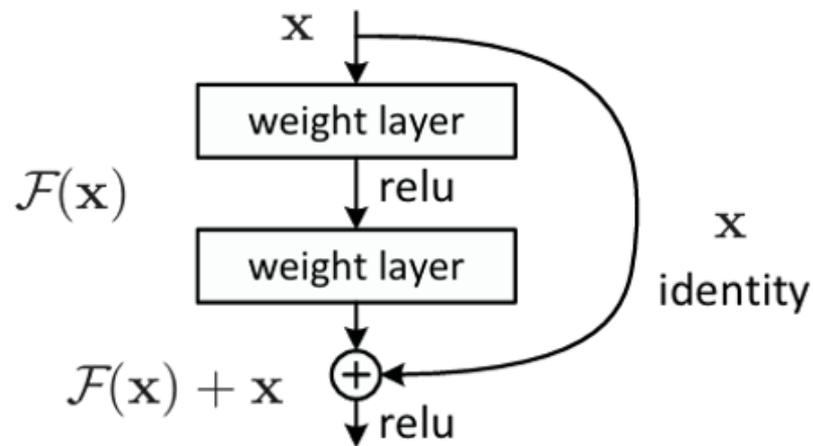


Figure 31: Schematic representation of a block of layers in a residual neural network (ResNet) with identity skip connection [81]. The output of such a building block is given by the sum of the functions applied by the layers and the input to the layers itself.

layer weights do not lead to a degradation of performance, as flow of information is not unrestricted through the inherent identity skip connections. Theoretical investigations into this architecture revealed them to behave more like ensembles of shallow networks than a single deep network, so that only relatively short paths contribute to the model performance, and removal of single layers after training does not impact the results strongly [82].

2 Materials and Methods

2.1 mrVista

The Matlab toolbox **mrVista** (<https://github.com/vistalab/vistasoft>) is developed and maintained by the Vista (vision, image science and technology applications) lab at Stanford University, California, USA. It includes functionalities for analyzing anatomical as well as functional MRI data. The toolbox is the de-facto gold standard in the field of pRF mapping, and thus was used as a reference pRF analysis tool for this thesis. All datasets were analyzed with the `prfanalyze-vista` docker container of the pRF validation framework [83], version 2.3.1 with default settings. This means that a coarse to fine fit was performed by **mrVista** as described in section 1.7. As a search radius, **mrVista** considers pRF center locations within twice the stimulus extent.

The fitting process employed by **mrVista** is the multi-stage minimization of the RSS between a timecourse and the model prediction to extract optimal pRF parameters described in chapter 1.7. This standard pRF mapping approach is notably only valid for symmetric noise sources, such as gaussian or white noise, and outliers in the data such as those due to subject motion might greatly influence results based on least squares optimization.

2.2 DeepRF and SnapRF

DeepRF is a deep learning framework originally proposed in 2019 [1]. Its goal is to enable faster and possibly more accurate pRF mapping compared to standard analysis software. To that end, a deep neural network is trained on artificially generated fMRI data with known ground truth. In keeping the terminology from chapter 1.8, the true pRF parameters of the synthetic data can be identified as the labels the network should learn to predict from the features of fMRI timecourses. As input, noisy artificial fMRI data generated with the stimulus used in the pRF mapping experiments of interest is given to the network. While the training procedure may take up to several hours on a graphical processing unit (GPU), the application of a trained model on a set of empirically gathered fMRI data containing several thousand voxels can be achieved within seconds. Importantly, only one network needs to be trained per stimulus.

The original authors of the method tested it on a single empirical dataset as well as on simulated data and compared the pRF mapping results from **DeepRF** with those of **popeye**, a python implementation of the standard approach. However, the simulated data in these investigations was generated in the exact same way as the artificial data with which **DeepRF** was trained, thereby introducing a significant bias towards **DeepRF** that was not corrected for. In addition, the simulated data used for training and performance evaluation was not realistic in the sense that it featured unrealistically low noise levels and non-physiologically large pRF sizes. For the empirical dataset, the **popeye** results were taken as reference, and reproducibility of results on empirical data was not examined. Further, only one stimulus was used for all experiments, so that the influence of stimulus design on **DeepRF** performance was not investigated.

In this thesis, the entire **DeepRF** pipeline was rebuilt in order to enable fair comparisons of the method with a reference method and to investigate its merits

on multiple datasets. This process also helped in gaining a deeper understanding of the approach and to introduce minor bugfixes and include various extensions and improvements to the method. To distinguish the reimplementations from the original **DeepRF**, the new version of the method written for this thesis shall be called **SnapRF** in all following descriptions in reference to the high speed afforded by the approach.

In **SnapRF**, the training and testing procedures were made entirely BIDS (Brain Imaging Data Structure) compliant. Further, a comprehensive logging procedure was introduced to document every step and output in the extensive model pipeline for documentation and debugging purposes. For training and testing the model, new call procedures were implemented in the form of **json** (Javascript object notation) scripts where all relevant parameters can be easily and flexibly set. For the network architecture, a 50 layer ResNet was used, as the original authors found such an architecture to perform best in comparison with others. CNNs are mostly used in image recognition tasks, and therefore the convolutional and pooling layers, as well as the kernel filters, are two-dimensional. Since fMRI timecourses are one-dimensional, the ResNet used in **DeepRF** and **SnapRF** employs one-dimensional convolutions and pooling operations. These can be thought of as smoothing operations or extractions of relevant features from the timecourses, such as peaks and their relative positions. For adaptive adjustments of learning rates during training, the popular Adam [84] optimization algorithm was used, as in the original implementation. In training, early stopping was introduced as a regularization technique and to reduce excessive GPU usage. Control over whether to make use of the technique was given to users in the **json** call script.

For training **DeepRF** and **SnapRF** models, artificial fMRI data must be generated with known ground truth. Models for the pRF shape, the HRF, and the noise sources present during measurement are necessary in order to calculate fMRI responses to a visual stimulus. In **DeepRF**, the pRF shape is modelled as a two-dimensional Gaussian, in the same way as in the original pRF mapping approach. This shape was also chosen for **SnapRF**.

The original **DeepRF** can only generate pRF center position within the extent of the input stimulus. In **SnapRF**, the model was modified to allow the generation of pRF center positions outside of that area. This was done in order to enable fair comparisons with the standard pRF mapping analysis software **mrVista**, which has a search radius of twice the stimulus radius. Control over the maximum eccentricity allowed for center locations was given to users in the **json** call script.

All pRF parameters for the signal in **SnapRF** were drawn from uniform distributions. The values for x and y could range from anywhere in the specified area within and around the stimulus. Combinations where the resulting center would lie outside of the stimulus were rejected and redrawn. The pRF size parameter could take on values from the smallest value possible within the stimulus resolution to a given maximum value.

In the original **DeepRF**, the upper limit for pRF sizes is given by the stimulus radius. Not only would a voxel with such a large pRF size receive input from almost anywhere in the area of visual field covered by the stimulus, thereby making it difficult to fit, but such a voxel would also be non-physiological: Mean pRF sizes in V1 have been shown to follow a linear relationship with eccentricity in various datasets [65, 85]. These findings suggest an upper limit of mean sizes below 6° when extrapolated for the eccentricities considered in this study. Therefore, pRF sizes

were restricted to a maximum of 6° in SnapRF, independently of the stimulus, as this size should be adequate to model even the largest pRF sizes found in V1.

The hemodynamic response of a voxel to optical stimulation is modelled by convolution between an HRF and the result of a multiplication between the Gaussian pRF and the stimulus for every generated voxel. Before this operation, the stimulus is downsampled in SnapRF to a resolution of 101×101 pixels to reduce the necessary computing power. While a classical pRF is described by three parameters, DeepRF and SnapRF allow for a fourth: An HRF delay. In training, delay values in seconds within a given symmetrical range around zero would be randomly generated from a uniform distribution and added to the modelled HRFs. This adds another parameter in optimization.

In addition to the HRF used in the original program, another HRF used by mrVista was implemented in this thesis to extend the capabilities of the method (Fig. 32).

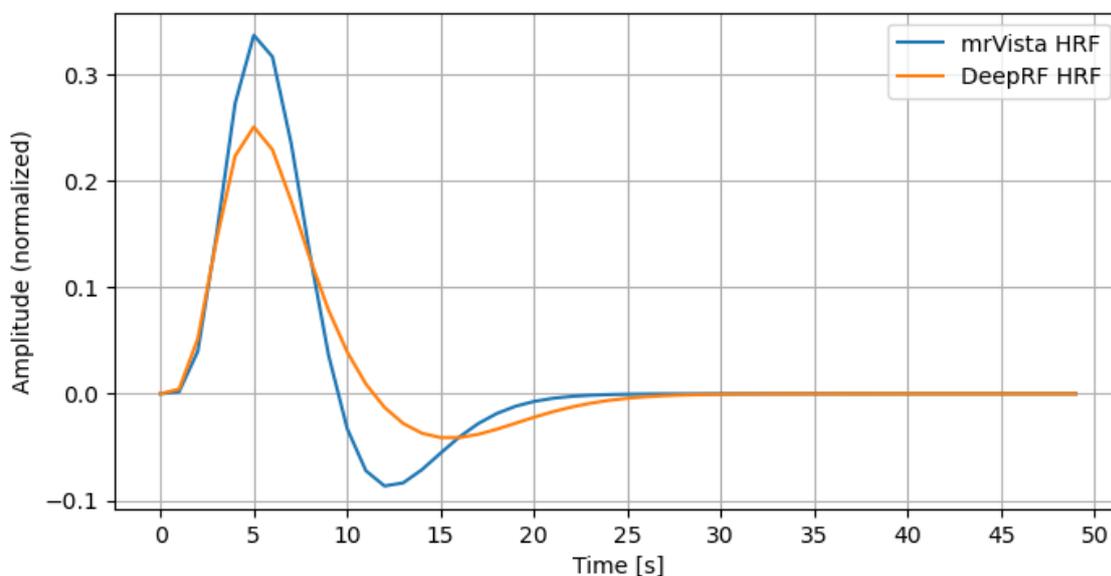


Figure 32: Comparison of HRFs used by DeepRF and mrVista to model stimulation at $Time = 0$. Both HRFs have been used for this thesis.

For adding noise to the signal, the CNR parameter is used to set the level of noise. In SnapRF, this parameter is sampled from a uniform distribution, while DeepRF samples from an exponential distribution of values.

DeepRF uses noise models from the neuRosim toolbox [86], which was developed to facilitate comparability between fMRI simulation studies in an effort to establish a standard approach for the field. The toolbox includes various functions for generating noise one might encounter in fMRI measurement data.

In the original DeepRF, noise is modelled as follows:

- System noise is modelled by white noise with a Gaussian distribution
- Physiological noise, such as the one induced by respiratory and cardiac activity, is implemented as a sum of two cosine waves of constant frequencies 0.2 Hz and 1.17 Hz
- Task-dependent noise is modelled as additional white noise during times when the stimulus is within a simulated voxel's pRF

- To account for signal drift over time due to low-frequency fluctuations in the scanner, a sum of cosine functions with frequencies in the range between the repetition frequency (TR) of the measurement and 128 seconds is used (the latter being inherited from the SPM toolbox, where 128 seconds is the default high-pass filter cutoff)
- To model temporal correlation between different points in time, an autoregressive process of first order is implemented

In **SnapRF**, the physiological noise described above was diversified to capture a more realistic distribution of cardiac and respiratory processes. To that end, frequencies for heart and breathing rate are drawn from uniform distributions between the intervals of [1.0, 2.0) Hz and [0.25, 0.4) Hz, respectively, for every simulated voxel. These values are chosen to represent average resting heart and breathing rates. Additionally, randomisation was introduced in the phases of the modelled cosine waves. All other noise functions were left as they were, except for minor bugfixes in the low-frequency noise component.

In addition to these changes to the **neuRosim** noise models, a completely new noise model was developed in this thesis for use in data generation to enable fair comparisons between **SnapRF** and **mrVista**. This new model is based on a principal component analysis (PCA) of resting-state (RS) fMRI data, and thus called the RS noise model. RS fMRI is a field of neuroscience concerned with investigating the processes happening in the brain during rest. By focussing on a small seed region and examining correlations between activations there and in other regions, the technique can be used to study connectivity in the brain. We assume that RS data can be treated as noise for the purpose of training **SnapRF** models, as RS data heavily features the noise components present in the brain during scanning and notably no task-evoked activations. The RS noise model used a normal distribution of weights for the combination of RS principal components (PCs), and white noise added on top. The parameters of the normal distribution of weights are taken from the PCA as well. Since the noise components and their relative contributions to the overall noise present in fMRI measurements are dependent on scanner, field strength, repetition time (TR), measurement sequence and site, the RS data was gathered from different studies using different scanners, sequences and field strengths, conducted in different MRI centers, so as to average out any effects specific to certain studies.

Table 2 at the end of this chapter gives an overview of the RS datasets used for the RS noise model.

Training performance of **DeepRF** and **SnapRF** models is evaluated based on the mean squared error (MSE) between predicted and true pRF parameters during validation phases in training. In terms of the generic model introduced in chapter 1.8, the MSE can be interpreted as the loss to be minimized in training. As such, in all following descriptions the model training performance is given in terms of the MSE validation loss.

For every empirical dataset used in the thesis, eight different **SnapRF** models were trained. This was done to investigate the influence of training parameters on model testing performance. The parameter combinations used in training per dataset can be seen in Table 1. To distinguish the models from one another, they are given names based on the training parameters. As names, a combination of three letters is used to identify the models: The first letter indicates the noise type

('N' for neuRosim, 'R' for RS noise), the second the HRF ('D' for DeepRF, 'V' for mrVista), and the third whether an HRF delay was trained ('y' for delay, 'n' for zero delay).

Table 1: SnapRF models trained per dataset.

Model	Noise Type	HRF	Delay Range [s]
NDn	neuRosim	DeepRF	0
NVn	neuRosim	mrVista	0
RDn	RS noise	DeepRF	0
RVn	RS noise	mrVista	0
NDy	neuRosim	DeepRF	$[-2, 2]$
NVy	neuRosim	mrVista	$[-2, 2]$
RDy	RS noise	DeepRF	$[-2, 2]$
RVy	RS noise	mrVista	$[-2, 2]$

DeepRF and SnapRF use z-score normalization for the generated artificial data in training, so that a baseline signal without stimulation is zero. Therefore, empirically gathered data with non-zero baseline values need to be normalized in the same way before application of DeepRF. This functionality is included in the newly implemented testing procedure, so that no manual normalization of input data is necessary before application of SnapRF.

The following table details the RS datasets used for the RS noise model. In the column titles, "subs" refers to the number of subjects, "runs" to the number of measurements gathered per subject, "voxels" to the number of voxels acquired, "volumes" to the length of each acquisition, "TR" to the repetition time of measurement between two consecutive scans of the same voxel, "TE" to the time between initial excitation and measured echo signal, "FA" to the flip angle of magnetization, "B" to the external static field strength, "R" to the isotropic resolution, "Scanner" to the scanner used for acquisition, "Center" to the institution where the scans were performed, and "Sequence" to the measurement sequence used. All datasets were gathered from openneuro.org.

Table 2: Openneuro datasets used for RS noise models.

Dataset	subs	runs	voxels	volumes	TR [s]	TE [ms]	FA	B [T]
ds001168	1	4	143360	300	4.0	26.0	70	7
ds001168	21	84	1146880	300	3.0	17.0	70	7
ds001454	24	48	147456	195	2.0	30.0	71	3
ds001566	1	1	778752	430	1.55	N/A	N/A	N/A
ds001728	1	1	102400	277	1.5	25.0	70	3
ds002156	1	16	131072	196	2.5	15.1-55.9	80	3
ds002766	1	47	147456	818	2.2	27.0	90	3
ds002766	3	153	453600	1636	1.1	33.0	84	3
ds004787	5	128	405000	240	2.5	12.9-70.9	77	3
ds004787	1	3	819200	300	2.5	30.0	77	3
ds004787	4	17	1228800	300	2.5	30.0	77	3
ds005069	1	1	159744	225	2.0	30.0	90	3
ds005072	1	1	159744	225	2.0	30.0	90	3

Table 2: Openneuro datasets used for RS noise models (*Continuation*).

Dataset	R [mm]	Scanner	Center	Sequence
ds001168	0.75	Siemens Magnetom	MPG, Leibniz	EPI31 2D
ds001168	1.5	Siemens Magnetom	MPG, Leibniz	2D
ds001454	3.0	Siemens Skyra	Princeton	GE EPI
ds001566	N/A	N/A	N/A	N/A
ds001728	N/A	Siemens TrioTim	Penn_State_SLEIC	EP-SK
ds002156	N/A	GE Discovery MR750	NIH FMRI	EP-RM
ds002766	4.0	Siemens TrioTim	Washington University	EP-SK
ds002766	N/A	Siemens PrismaFit	Washington University	EP-SK
ds004787	N/A	GE Discovery MR750	NIH FMRI	EP-RM
ds004787	N/A	GE Discovery MR750	NIH FMRI	EP-RM
ds004787	N/A	GE Discovery MR750	NIH FMRI	EP-RM
ds005069	N/A	GE Discovery MR750	UESTC	EP-GR-SS
ds005072	N/A	GE Discovery MR750	UESTC	EP-GR-SS

2.3 Simulated Data

For comparisons between **SnapRF** and **mrVista** with known ground truth, a dataset was created with artificially generated data. The data was generated using the fMRI timecourse models and parameter distributions described in chapter 2.2. As **SnapRF** is trained with data created in the same fashion, this introduces an inherent bias in comparisons. To reduce this asymmetry, the dataset was created with the RS noise model described earlier, and only **SnapRF** models trained with **neuRosim** noise models were tested on the data. Therefore, the **SnapRF** models applied on the artificially generated data had no prior knowledge of the noise structure in the data.

The stimulus used for data creation and model training will be described in chapter 2.4. For data generation, center locations with eccentricities up to twice the stimulus radius were allowed, the **mrVista** HRF was used and the HRF delay was set to zero for all voxels. In total, 10,000 voxels were generated.

2.4 Stimsim24 Dataset

The **Stimsim24** dataset was acquired at the Center of Excellence for High-field MRI at the Medical University Vienna by David Linhardt and Luna Müller-Hartburg. It features anatomical and functional data from two healthy subjects (25f, 23m), with 30 pRF mapping runs over five sessions each. For this study, only functional data from the primary visual cortex (V1) was used. The data was acquired on a 3 T Siemens **PrismaFit** MRI scanner (Siemens Healthineers, Erlangen, Germany) with the lower part of a 64 channel head coil and the CMRR multiband EPI sequence (Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, USA). The sequence parameters were TE = 38 ms, TR = 1 s, voxel size = 1.5 mm isotropic, 30 slices parallel to the calcarine sulcus, 10 percent slice spacing, multiband factor = 3, flip angle = 55°, 240 volumes per run. The stimulus used for the pRF mapping experiments had a width of 1.6° and occupied the central 18° of visual field. It jumped by 0.8° after each TR. The temporal pattern of stimulation is a modification of the original sweeping bar design described in chapter 1.7. Instead of a 45° rotation after each sweep, the sweeping axis is rotated 135°, so that it sweeps

first from left to right (L-R), then from down left to up right (DL-UR), followed by a sweep from down to up (D-U), then from down right to up left (DR-UL), and so on (Fig. 33). As in the original configuration, blank periods are inserted after each diagonal sweep. In this design, the time between two consecutive stimulations of any quadrant is increased, so that the responses will show less overlap than in the original design. After eight sweeps and a last blank period, a single run is complete.

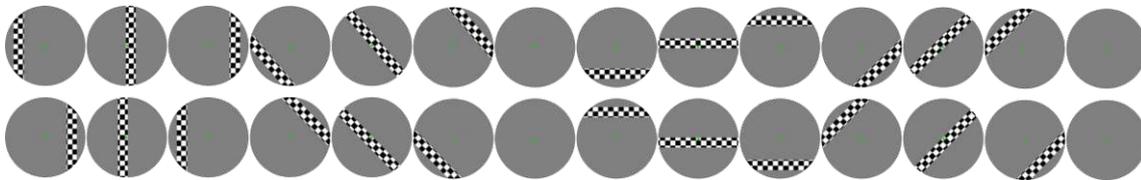


Figure 33: Sweeping bar stimulus used in the Stimsim24 dataset (adapted from [70]).

For denoising, NORDIC [87] (noise reduction with distribution corrected PCA) was applied to the functional data. The following paragraph was generated by *fMRIPrep* and details the preprocessing steps taken after denoising.

Anatomical and functional data was preprocessed using *fMRIPrep* 23.1.4 [88, 89], which is based on *Nipype* 1.8.6 [90, 91]. A total of 12 fieldmaps were found available per subject. A B_0 -nonuniformity map (or *fieldmap*) was estimated based on two (or more) echo-planar imaging (EPI) references with *topup* [92]. A total of 12 T_1 -weighted (T_1w) images were found per subject. Anatomical preprocessing was reused from previously existing derivative objects. For each of the 72 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed: First, a reference volume and its skull-stripped version were generated by aligning and averaging 3 single-band references (SBRefs). Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using *mcfirt* [93]. The estimated *fieldmap* was then aligned with rigid-registration to the target EPI reference run. The field coefficients were mapped on to the reference EPI using the transform. BOLD runs were slice-time corrected to 0.441s (0.5 of slice acquisition range 0s-0.882s) using *3dTshift* from AFNI [94]. The BOLD reference was then co-registered to the T_1w reference using *bbregister* (FreeSurfer) which implements boundary-based registration [95]. Co-registration was configured with six degrees of freedom. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, [96]) and Jenkinson (relative root mean square displacement between affines, [93]). FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* [96]. The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*) [97]. Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor,

three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, a mask of pixels that likely contain a volume fraction of GM is subtracted from the aCompCor masks. This mask is obtained by dilating a GM mask extracted from the FreeSurfer’s *aseg* segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components’ time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each [98]. Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. Additional nuisance timeseries are calculated by means of principal components analysis of the signal found within a thin band (*crown*) of voxels around the edge of the brain, as proposed by [99]. The BOLD time-series were resampled into standard space, generating a *preprocessed BOLD run in MNI152NLin2009cAsym space*. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. The BOLD time-series were resampled onto the following surfaces (FreeSurfer reconstruction nomenclature): *fsnative*, *fsaverage*. All resamplings can be performed with a *single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using *antsApplyTransforms* (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels [100]. Non-gridded (surface) resamplings were performed using *mri_vol2surf* (FreeSurfer).

2.5 NYU Retinotopy Dataset

The NYU retinotopy dataset was acquired at the Center for Brain Imaging of the New York University [85]. It contains anatomical and functional data of 44 healthy subjects and was acquired on a 3 T Siemens Prisma scanner (Siemens Medical Solutions, Erlangen, Germany) with a Siemens 64 channel head coil. For this study, only functional data from the primary visual cortex (V1) was used. For each subject, between four and 12 pRF mapping runs were acquired using a multiband EPI sequence with TE = 37 ms, TR = 1 s, voxel size = 2 mm³, flip angle = 68°, 192 volumes per run. The sweeping bar stimulus of 3.1° width occupied the central 24.8° of visual field and is similar to the original design used in [65], except for the diagonal sweeps, which only sweep over half of the screen (Fig. 34). After every TR, the bar jumped 1.033°. Therefore, every quadrant of visual field receives six stimulations in this design, as opposed to eight in the original. Additionally, the background revealed by the bar aperture did not consist of a high contrast checkerboard pattern, but instead of an assortment of natural images (e.g. faces, plants, animals) at multiple scales on

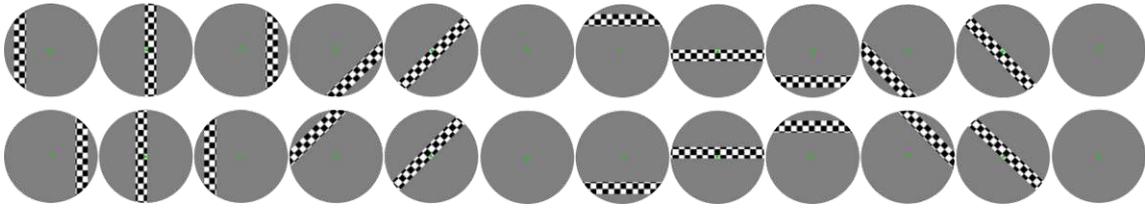


Figure 34: Sweeping bar stimulus used in the NYU retinotopy dataset (adapted from [70]).

top of a pink noise background (Fig. 35). The images were changed three times per second. This background was chosen for comparability with the HCP retinotopy dataset, which uses the same images and noise background as such textures were produce high signal-to-noise ratios (SNRs) in higher visual areas [101].

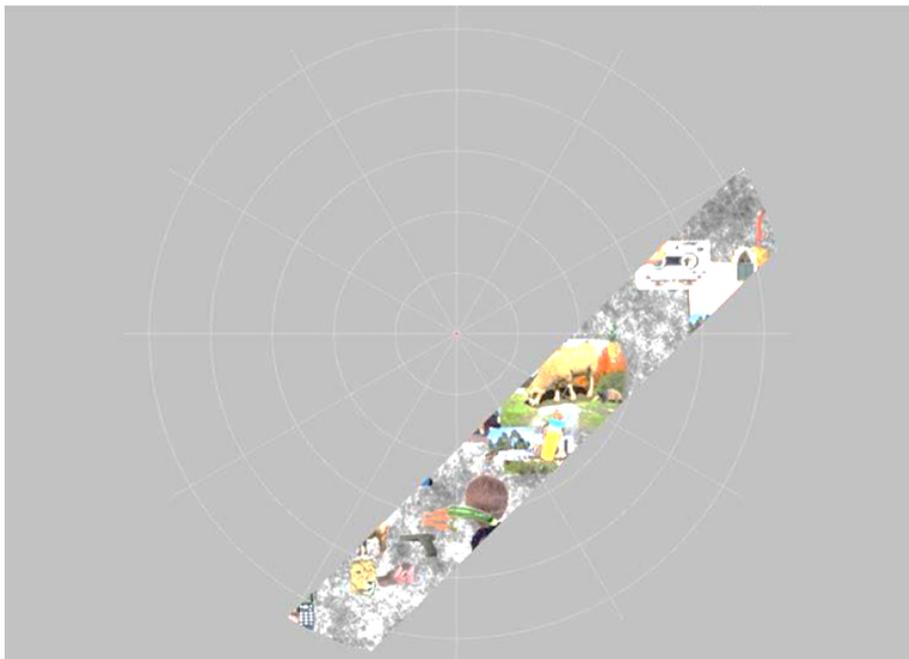


Figure 35: Background consisting of natural images over pink noise revealed by the stimulus aperture in the NYU retinotopy dataset.

The following paragraph was generated by *fMRIPrep* and details the preprocessing steps taken by the original authors.

Anatomical and functional data was preprocessed using *fMRIPrep* 20.0.1 [88, 89], which is based on *Nipype* 1.4.2 [90, 91]. T_1 -weighted (T_1w) anatomical images were corrected for intensity non-uniformity (INU) with *N4BiasFieldCorrection* [102], distributed with ANTs 2.2.0 [103], and used as T_1w -reference throughout the workflow. The T_1w -reference was then skull-stripped with a *Nipype* implementation of the *antsBrainExtraction.sh* workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T_1w using *fast* [104]. Brain surfaces were reconstructed using *recon-all* [105], and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle [106]. Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with

antsRegistration (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: *ICBM 152 Nonlinear Asymmetrical template version 2009c* [107]. For each of the 12 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. A B0-nonuniformity map (or *fieldmap*) was estimated based on two (or more) echo-planar imaging (EPI) references with opposing phase-encoding directions, with *3dQwarp* [94] (AFNI 20160207). Based on the estimated susceptibility distortion, a corrected EPI reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using *bbregister* (FreeSurfer) which implements boundary-based registration [95]. Co-registration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using *mcflirt* [93]. BOLD runs were slice-time corrected using *3dTshift* from AFNI 20160207 [94]. The BOLD time-series were resampled onto the following surfaces (FreeSurfer reconstruction nomenclature): *fsnative*, *fsaverage*. The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. The BOLD time-series were resampled into standard space, generating a *preprocessed BOLD run in MNI152NLin2009cAsym space*. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* [96]. The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction *CompCor* [97]. Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 5% variable voxels within a mask covering the subcortical regions. This subcortical mask is obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, components are calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms

for each [98]. Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using *antsApplyTransforms* (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels [100]. Non-gridded (surface) resamplings were performed using *mri_vol2surf* (FreeSurfer).

3 Results

In order to answer the questions posed in chapter 1 concerning fair comparisons with standard analysis tools and the assessment of the reproducibility of results, **SnapRF** and **mrVista** were applied on the three different datasets described in chapter 2.3-2.5: (1) Simulated data, (2) **Stimsim24** data, (3) **NYU** retinotopy data. The datasets were chosen specifically because their properties made them well suited for investigating different aspects of the research questions.

In this chapter, the **SnapRF** training performance and comparisons between **SnapRF** and **mrVista** pRF mapping results are presented for all datasets. Even though **SnapRF** is agnostic to variance-explained during training and testing, this metric was calculated for every voxel of every dataset after application of **SnapRF**. In pRF mapping experiments, voxels with variance-explained below a certain threshold are usually discarded before analysis of results. This is done because pRF fitting models will always give an estimate, no matter the signal content of a voxel. Therefore, the same was done for **SnapRF** results.

As described in chapter 2.2, eight **SnapRF** models were trained per empirical dataset with different training parameters (see Table 1). All of those models were then tested on all pRF measurement files of the empirical datasets and variance-explained was calculated for all voxels of those files. In a next step, rank-sums were calculated independently for the mean and median variance-explained values per measurement file, and for the number of voxels above a 0.1 variance-explained threshold. Finally, the resulting rank-sums were again ranked per measure (mean variance-explained, median variance-explained, number of voxels) and a last rank-sum was calculated from the ranks of the models in that assessment (see Tables 5 and 6). The **SnapRF** model so chosen as best for an empirical dataset was then compared to **mrVista**.

3.1 SnapRF Training Parameters and Testing Performance

Before method comparison on the datasets, the **SnapRF** training and testing performance was evaluated to guide the choice of training parameters and to examine the model's speed depending on dataset size.

To set a sensible learning rate for training the model, three different values were compared by means of MSE loss in training (Fig. 36). The results of this comparison showed a learning rate of 0.001 to converge faster and to lower minimum loss values than the others. Therefore, this value was chosen for all experiments reported herein.

To adjust the level of noise in training, the mean variance-explained values between the pure and noisy timecourses of 1000 voxels for different CNR values was calculated (Fig 37). From these calculations, it is apparent that the newly implemented RS noise model is comparable to the **neuRosim** noise model in noise level, though the average variance-explained values are a little smaller for the RS noise model. In **DeepRF**, an exponential distribution of CNR values between 0.5 and 2.0 were used. This range in CNR corresponds to average variance-explained values between 50 and 95 percent. In **SnapRF**, a uniform distribution of CNR values was used and control over the CNR range was given to users in the **json** call script. For the experiments reported herein, CNR values between 0.15 and 1.0 were used, which corresponds to average variance-explained values between 10 and 80 percent. This

was done so as to make the training timecourses more realistic, as in actual pRF mapping measurements very few if any voxels at all have variance-explained values as high as the ones used in the original DeepRF training data. Further, the lower bound chosen for this thesis of 10 percent variance-explained corresponds to a cutoff often employed in pRF mapping analyses.

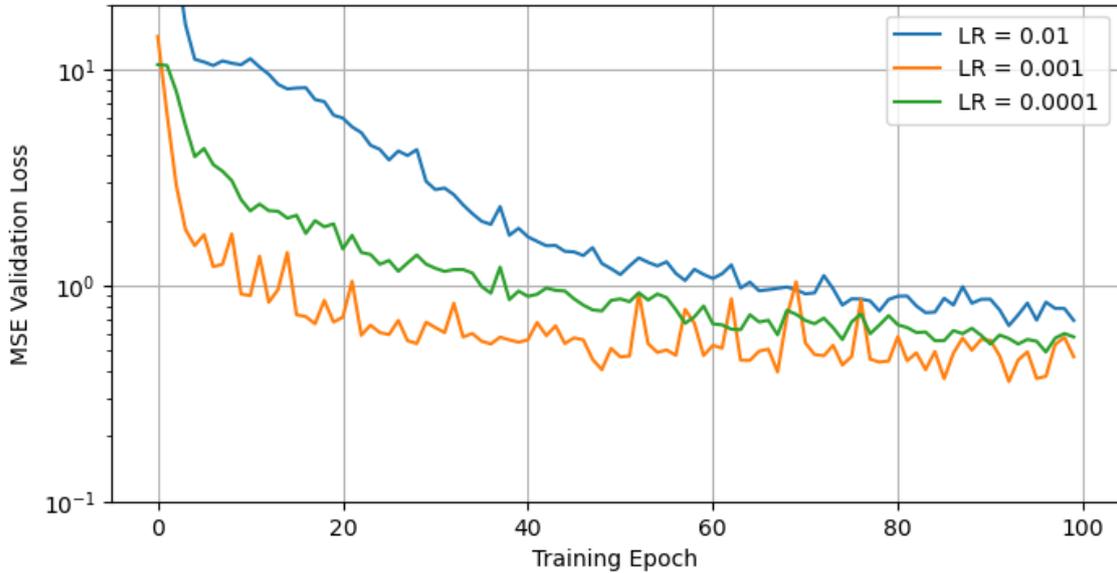


Figure 36: SnapRF MSE validation loss for different learning rates (LR) showing a model with a learning rate of 0.001 converging faster and to lower loss values than with other learning rates.

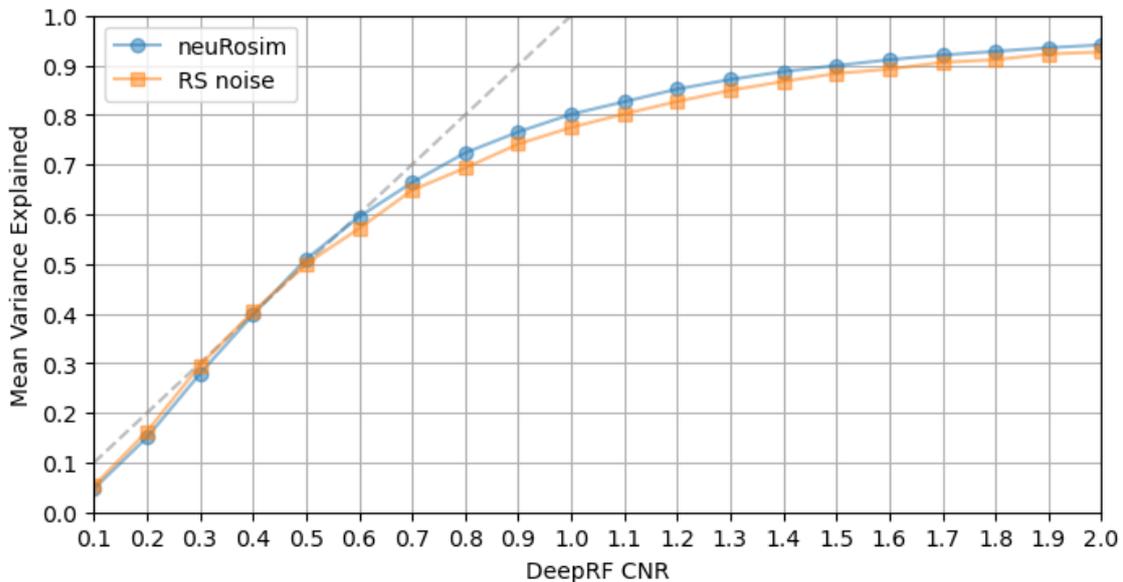


Figure 37: Mean variance-explained values of 1000 voxels for different SnapRF CNR values and noise models showing comparable.

When testing a trained model on a dataset to obtain the SnapRF parameter estimates, an almost linear relationship between the number of voxels to analyze and the time taken for testing can be observed (Fig. 38). The measured pure

testing times (without writing results to file) for two different simulated datasets are listed in Table 3.

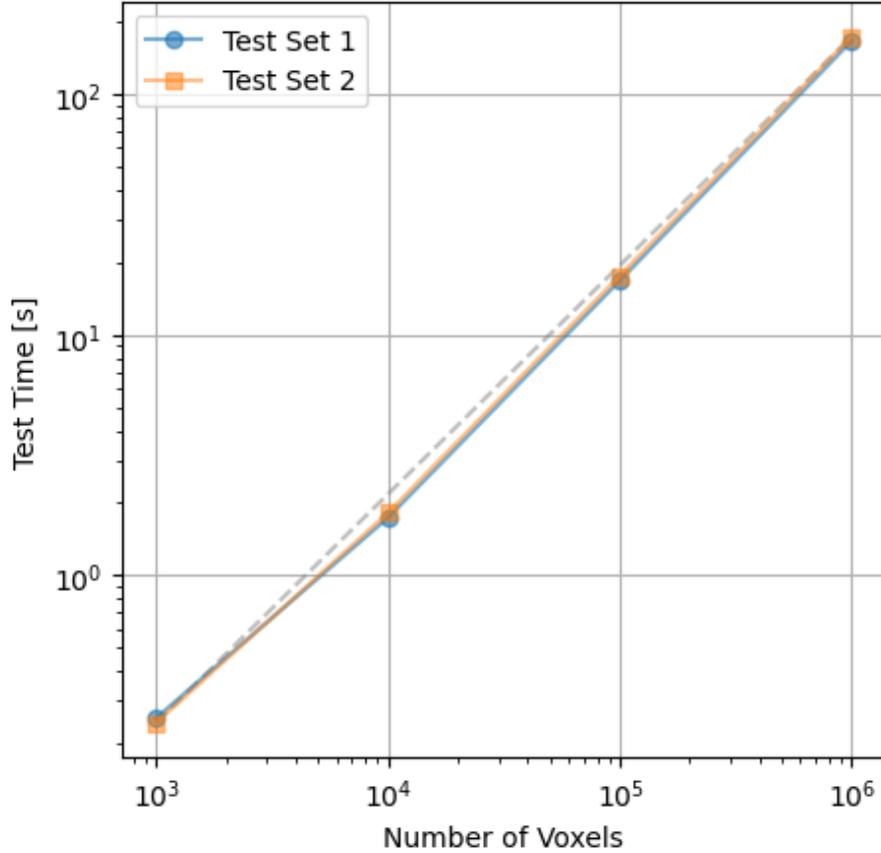


Figure 38: Almost linear relationship between number of voxels in a test set and the time taken for applying SnapRF in seconds.

Table 3: SnapRF pure testing times on simulated data.

Number of Voxels	Set 1 Pure Test Time [s]	Set 2 Pure Test Time [s]
10^3	0.254	0.243
10^4	1.737	1.819
10^5	16.862	17.653
10^6	167.903	174.725

3.2 Results on Simulated Data

For testing SnapRF on the simulated dataset, a model was trained with neuRosim noise models and the Stimsim24 stimulus (Fig. 39). In training, data with eccentricities of up to twice the stimulus radius was generated, the mrVista HRF was used and the HRF delay was set to zero, in an effort to establish maximum comparability between SnapRF and mrVista.

In contrast to the empirical datasets, this dataset provides the unique opportunity of direct assessment of pRF mapping errors. After creation of the dataset, the SnapRF-NVn model and mrVista were applied and the time necessary for application was taken: Testing mrVista on the 10,000 simulated voxels took about 2296 seconds,

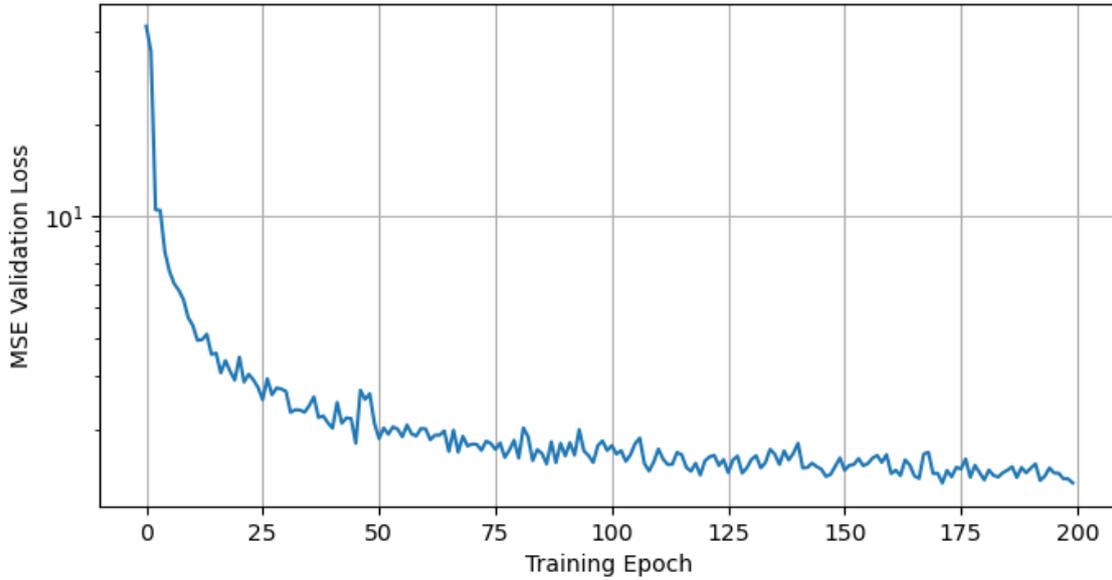


Figure 39: MSE validation loss of model used for evaluation of **SnapRF** on simulated data converging towards a value close to 1. This is a higher loss than in the original **DeepRF** results but expectedly so due to higher noise and larger search radius in training **SnapRF**.

while **SnapRF** took only about 12 seconds for the same task (Table 4). In both of these times, writing the mapping results to output files is included, which increases the total testing time almost ten-fold for **SnapRF** (see pure testing times in Table 3). Still, this represents a 191-fold increase in testing speed for getting usable results in comparison to **mrVista**.

Table 4: **SnapRF** and **mrVista** testing times on simulated dataset.

Number of Voxels	SnapRF Test Time [s]	mrVista Test Time [s]
10^4	12	2296

After testing the models, the errors between predictions and true parameters were calculated and plotted in dependence of eccentricity, pRF size, and variance-explained (Fig. 40). For every calculated error, moving percentiles were calculated with a Gaussian filter kernel as a window.

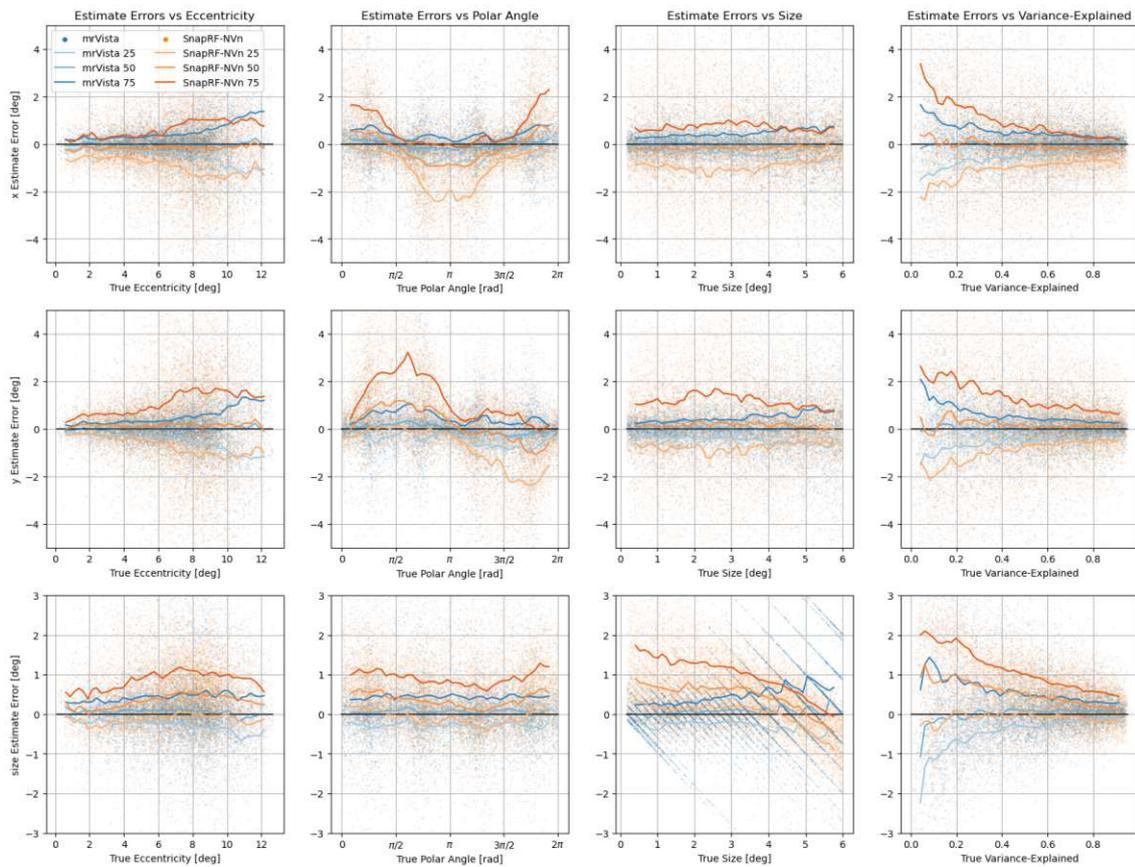


Figure 40: PRF estimate errors for mrVista and SnapRF-NVn on 10,000 simulated voxels in dependence of eccentricity, pRF size, and variance-explained. Moving percentiles are also plotted.

3.3 Results on Stimsim24 Dataset

For the **Stimsim24** dataset, the **SnapRF** model **RVn** proved best by metric of rank-sums described in the introduction to this chapter: First, all models were compared in terms of mean and median variance-explained values as well as in terms of the number of voxels above a variance-explained threshold of 0.1 per measurement file. From these comparisons, rankings were performed for each file and the ranks over all files and comparisons added. This left three total values per model which were again ranked individually and these ranks were then added up to give the final ranking of a model (see Table 5). In training, the **SnapRF-RVn** model reached an MSE Loss close to two, which is higher than in the original **DeepRF** results but not alarmingly so (Fig. 41).

The **Stimsim24** dataset provided the unique opportunity to test reproducibility of empirical pRF mapping results because it contains a large number of measurements for single subjects. To that end, all 30 measurement runs were averaged to create a single average run per subject. This has the effect of greatly increasing the signal-to-noise ratio (SNR) in comparison to single runs. To verify whether **SnapRF** is capable of reliably reproducing retinotopy, eccentricity estimates on the average run were plotted onto the inflated brain surface (Fig. 42 and 43). To compare the inter-method reproducibility of estimates, direct voxel-wise comparisons of pRF estimates between **mrVista** and **SnapRF** were conducted on the averaged run (Fig. 44). To assess inter-measurement reproducibility, the mean deviations of pRF estimates across the thirty runs from the pRF estimates on the single average run were then calculated and the estimates on the average run were taken as "ground truth" values with respect to each method (Fig. 45).

Table 5: Variance-explained Rank-Sums for **SnapRF** Models on the **Stimsim24** Dataset.

Model	Rank-Sum Mean	Rank-Sum Median	Rank-Sum Cutoff	Rank-Sum Total
RVn	393 (1)	381 (1)	558 (4)	6
NDy	468 (3)	463 (3)	220 (2)	8
NVy	438 (2)	397 (2)	749 (8)	12
RVy	531 (4)	557 (6)	543 (3)	13
RDy	574 (6)	633 (7)	208 (1)	14
NVn	533 (5)	507 (4)	616 (6)	15
NDn	591 (7)	555 (5)	690 (7)	19
RDn	618 (8)	653 (8)	562 (5)	21

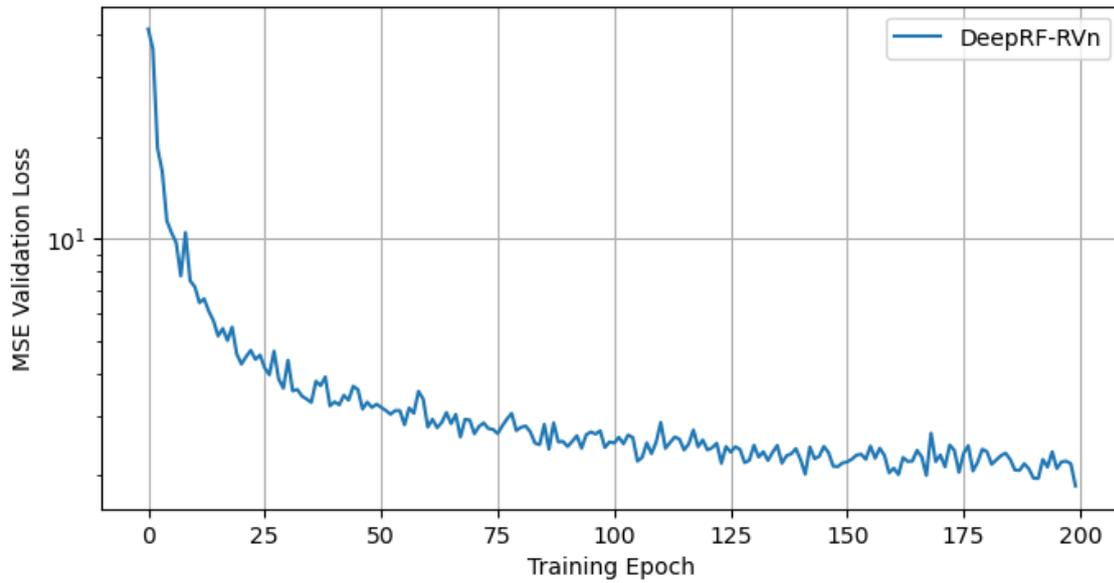


Figure 41: MSE validation loss of model used for evaluation of SnapRF on the Stimsim24 dataset showing convergence to a value below 2. This is higher than in the simulation study, likely owing to the different type of noise used in training.

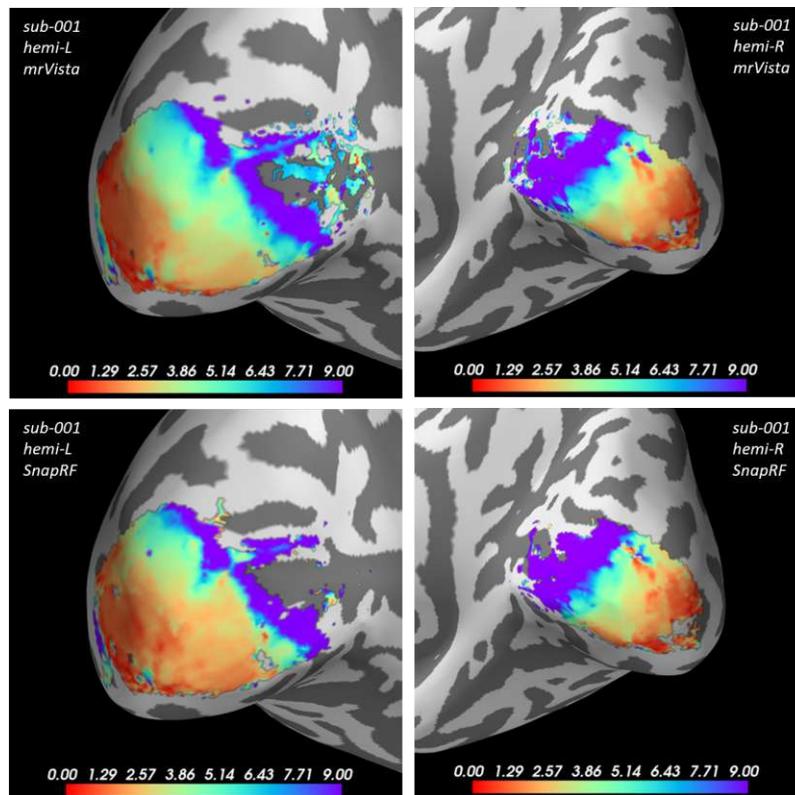


Figure 42: Comparison of eccentricity estimates of mrVista (top) and SnapRF-RVn (bottom) on the inflated brain surface of subject 001 for the averaged run.

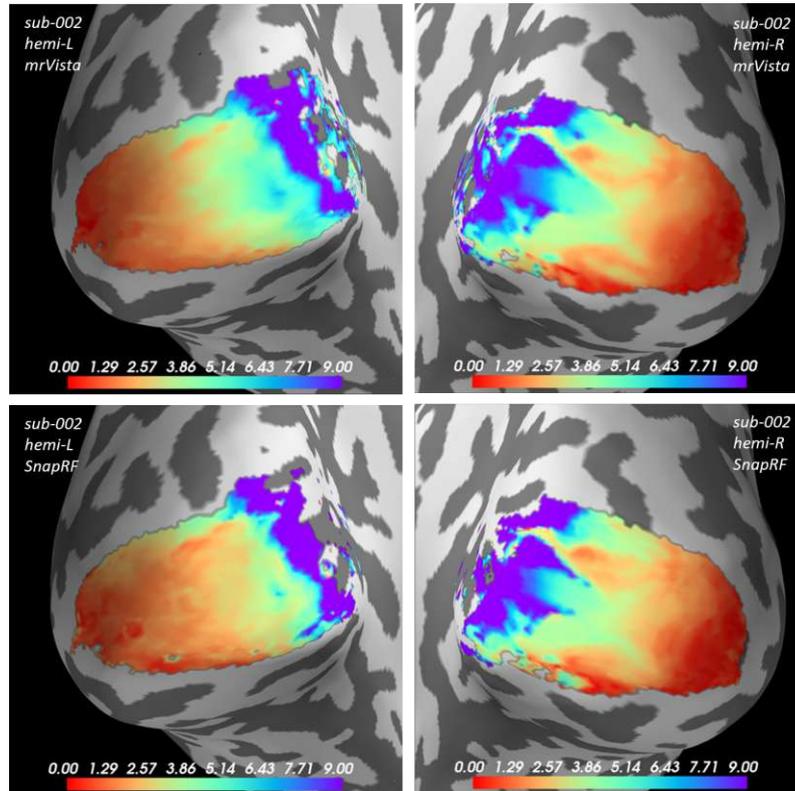


Figure 43: Comparison of eccentricity estimates of mrVista (top) and SnapRF-RVn (bottom) on the inflated brain surface of subject 002 for the averaged run.

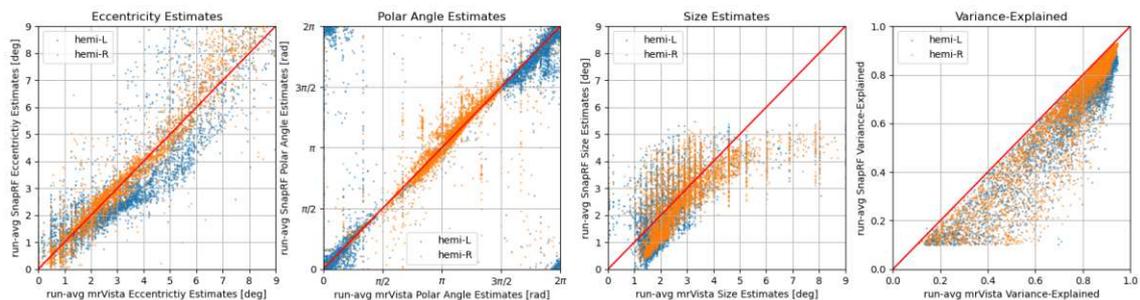


Figure 44: Voxel-wise comparison of pRF estimates of mrVista and SnapRF-RVn on averaged run for both Stimsim24 subjects. Only voxels with variance-explained > 0.1 for both methods are plotted. Blue voxels are from the left hemisphere, and orange voxels from the right hemisphere.

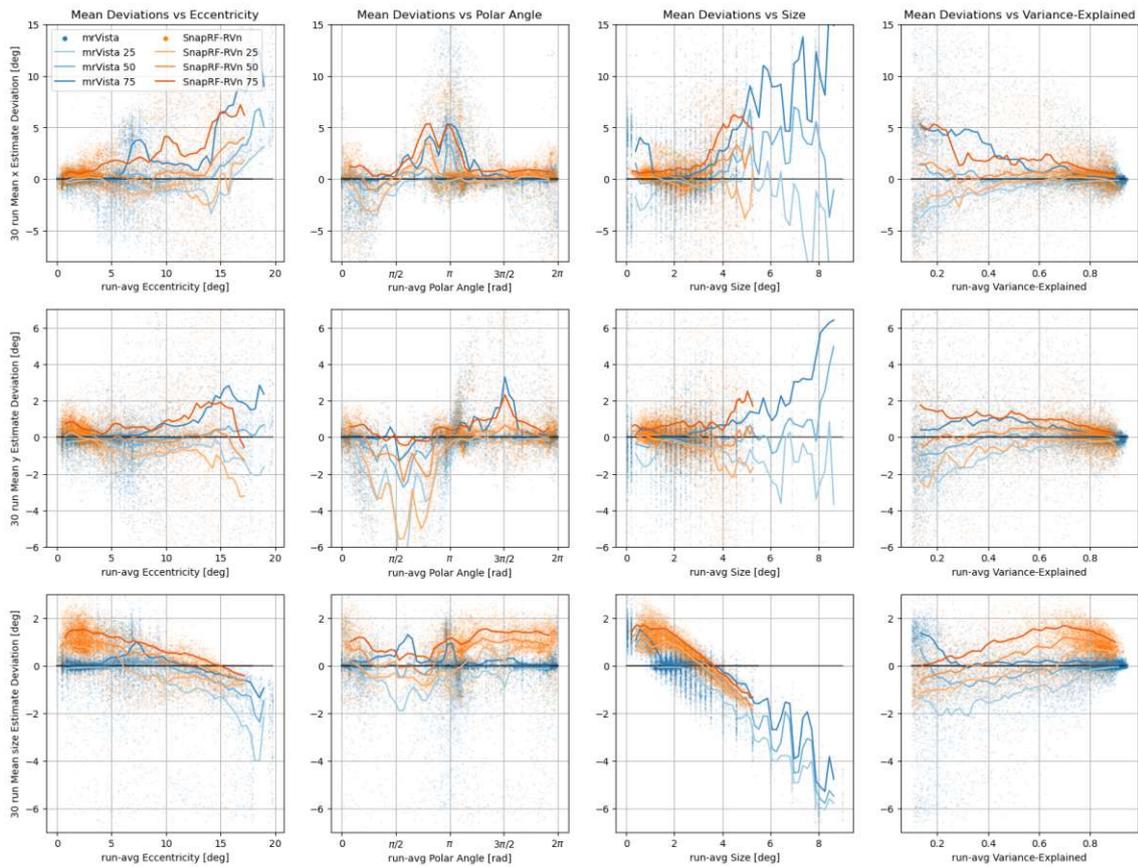


Figure 45: Mean pRF estimate deviations over 30 runs from estimates on averaged run in dependence of eccentricity, pRF size, and variance-explained of averaged run estimates. Data for both subjects of the Stimsim24 dataset were pooled. Moving percentiles are plotted additionally.

3.4 Results on NYU Retinotopy Dataset

As before, the model used for analysis of the NYU dataset was chosen by means of rank-sums of variance-explained mean, median and quantity above a threshold values per measurement file. These rank-sums were then themselves ranked and these final ranks added together, which showed the **SnapRF-RDy** model to perform best (see Table 6). In training, the model reached an MSE Loss close to seven, which is significantly higher than models trained with the **Stimsim24** stimulus, indicating less predictive power of this trained model than in the other datasets (Fig. 46).

The NYU retinotopy dataset is well suited to assess the inter-method reproducibility across multiple subjects, as it features data from 44 participants. To facilitate the comparison, the measurement files were first averaged per subject, then the pRF mapping methods were applied on the native brain surfaces. Afterwards, these results were interpolated onto the fsaverage surface with *freesurfer mri_surf2surf*. From these, median subject estimate maps were calculated and these were compared voxel-wise between **mrVista** and **SnapRF** (Fig. 47 top panels). Additionally, group-average timeseries were calculated from the measurement files averaged per subject. To do so, these were interpolated onto the fsaverage surface and then averaged over all subjects. Then, **mrVista** and **SnapRF** were applied to the group-average timeseries. On these results, too, the voxel-wise reproducibility between the methods was compared (Fig. 47 bottom panels).

Table 6: Variance-explained Rank-Sums for **SnapRF** Models on the NYU Retinotopy Dataset.

Model	Rank-Sum Mean	Rank-Sum Median	Rank-Sum Cutoff	Rank-Sum Total
RDy	1933 (1)	2019 (1)	1589 (1)	3
RVy	2223 (2)	2030 (2)	2285 (2)	6
NVn	2621 (3)	2437 (3)	3335 (6)	12
NDy	2996 (5)	3076 (5)	2342 (3)	13
RVn	2942 (4)	2985 (4)	3419 (7)	15
NDn	3125 (6)	3238 (6)	2637 (5)	17
RDn	3500 (8)	3497 (8)	2390 (4)	20
NVy	3204 (7)	3262 (7)	4511 (8)	22

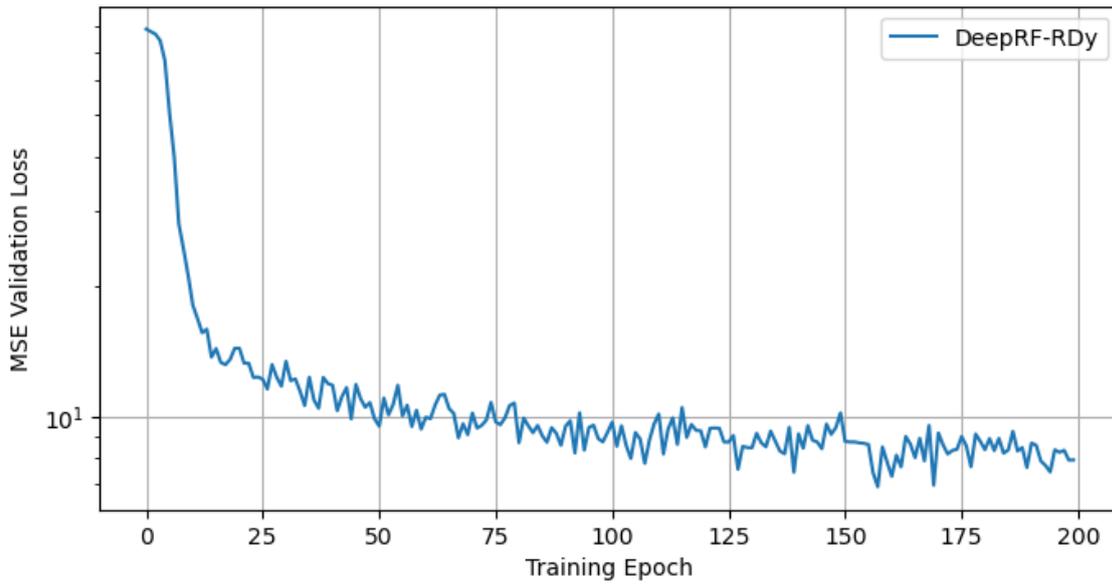


Figure 46: MSE validation loss of model used for evaluation of SnapRF on the NYU retinotopy dataset reaching lowest values of about 7, which is much higher than in previous datasets, indicating less predictive power of the trained model in this dataset.

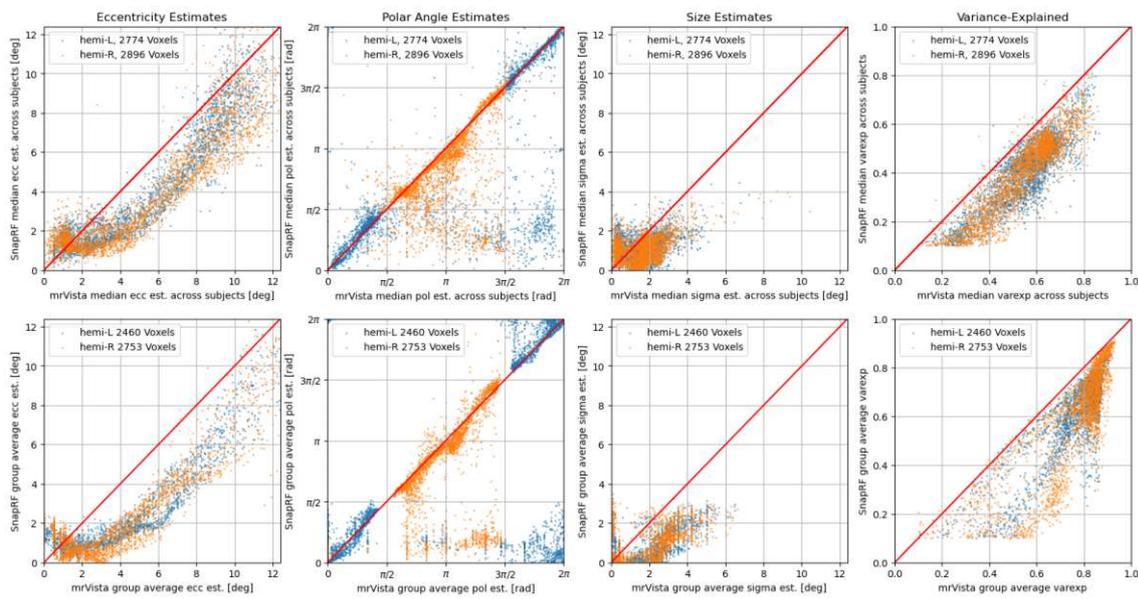


Figure 47: Comparison of SnapRF-RDy and mrVista voxel-wise median subject pRF mapping results (top) and group-average pRF mapping results (bottom) on the NYU retinotopy dataset. Only voxels with variance-explained > 0.1 for both methods are plotted. Blue points are voxels from the left hemisphere, and orange ones from the right hemisphere.

4 Discussion

Comparing the DeepRF approach with conventional pRF mapping software poses a problem that cannot be easily resolved. This arises from the different approaches used by the tools to estimate the parameters of the pRF model on real data. While DeepRF is trained to capture parameters directly from a voxel’s timecourse, standard analysis methods perform a multi-step optimization with the goal of maximizing variance-explained by their choice of parameters. It is central to note here that DeepRF has no knowledge of variance-explained at any point, neither in training nor in testing. On the other hand, mrVista relies on variance-explained to quantify the goodness of parameter fits.

Since DeepRF is trained with artificially generated data with known ground truth, testing it on data that is generated in the same way will always lead to a bias compared to a software agnostic to the input data. Testing it on empirical data, however, makes comparison difficult because DeepRF does not optimize for the parameter used to quantify goodness-of-fit. DeepRF could of course be implemented to optimize for variance-explained. However, this would result in a model that fits noise rather than ground truth, exemplifying Goodhart’s law that a measure ceases to be useful when it becomes a target to optimize for [108]. It is conceivable that in high noise settings, a DeepRF model may fare worse in regards to variance-explained, while actually extracting more correct parameters than conventional methods. On empirical data, however, variance-explained is a useful metric useful for discriminating good from bad predictions. Thus, we have arrived at a circular argument which cannot be easily resolved.

In this thesis, new approaches were taken to break the cycle: Through the introduction of the new RS noise model, a simulation study was conducted with artificial data containing noise components neither model knew beforehand, effectively removing a large bias towards the reimplemented SnapRF in comparison to mrVista. Further, the reproducibility of pRF mapping results on empirical data was investigated through calculation of deviations from estimates on a multi-run average timeseries. Additionally, the inter-method reproducibility of pRF mapping results was compared on an external dataset featuring a large number of subjects with a different stimulus design from the other datasets.

4.1 Results on Simulated Data

When comparing the MSE training loss in Fig. 39 with the training losses the original authors presented for DeepRF [1], it is apparent that the losses presented herein are magnitudes larger than those of the original implementation. This effect stems from the fact that the model was trained with much higher noise than in the original implementation as well as with timecourses generated from pRF center locations outside of the stimulus (Fig. 48).

In Fig. 40, it is apparent that x and y errors show no strong bias when plotted against eccentricity for either method, which is highly desirable. However, it is notable that SnapRF errors stray further than mrVista errors. The same behaviour can be observed in dependence of size and noise (variance-explained). From the x and y errors in dependence of polar angle it is evident that SnapRF results compared to those of mrVista show a strong influence of the angular distance to the converse

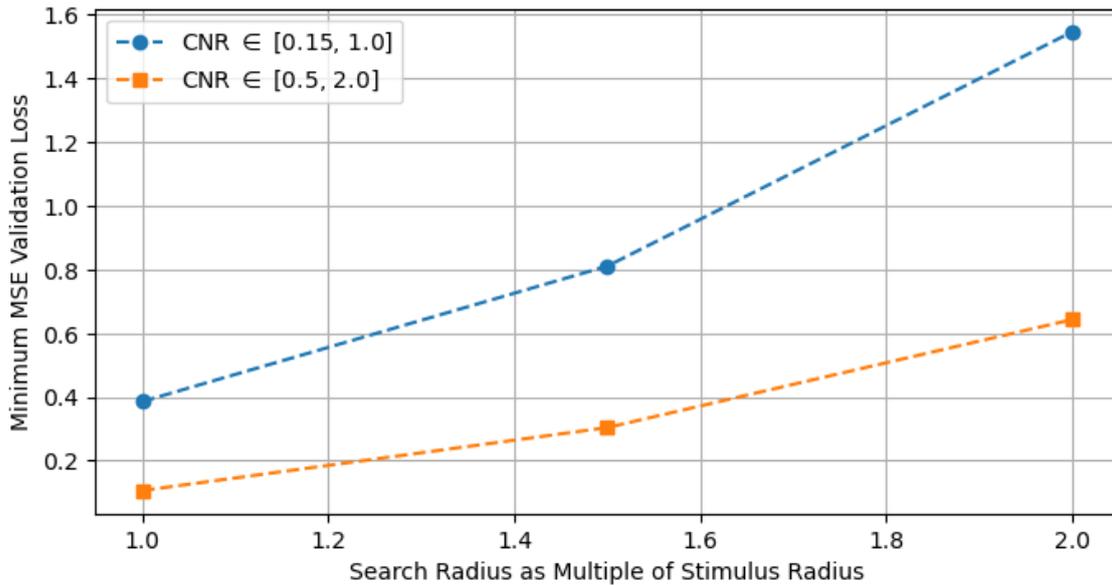


Figure 48: Minimum MSE validation loss of **SnapRF** models trained with different combinations of search radius and CNR values for the **Stimsim24** stimulus, showing how lower CNR values and larger search radii in training decrease model training performance.

axis, respectively. This effect is expected and also visible in the results of **mrVista**, though not as strongly.

In the size errors, there exists a clear bias towards overestimation of size for **SnapRF**, which is especially pronounced for small sizes. This effect is likely due to the pRF parameter space in **SnapRF** training and not an inherent feature of the method. However, it is interesting because the artificial data was drawn from the same distributions of x , y , pRF size and CNR values that were used in training the **SnapRF** model. Therefore, these effects might indicate the need for a more robust training of the size parameter. Splitting the loss function into the individual pRF parameters during model training, however, shows that the mean error in pRF size is minimized better in training than those of x and y positions (Fig 49). Thus, the bias of **SnapRF** in size estimation might also stem from the different type of noise in the dataset than the model was trained with. For **mrVista**, the effect of the grid fit is clearly visible in the size error dependent on actual size, leading to a larger spread of errors for larger sizes than **SnapRF**.

Notably, the MSE loss in x and y in Fig. 49 reach values of about 4° , therefore the mean absolute errors at the end of training take values of around 2° , which is close to the stimulus width of 1.6° . The mean absolute error of the pRF size parameter meanwhile reaches values close to 0.8° , which is half of the stimulus width. These correlations indicate an influence of the stimulus width on the training performance, where narrower stimuli might help in reducing training losses. In empirical settings, however, exceedingly narrow stimuli would necessarily lead to longer measurement times at a constant stimulus diameter or to smaller stimulated areas of the visual field at constant measurement times. This is because it takes more steps to cover the whole stimulated region with narrower stimuli than with broader ones. In practice, neither longer measurement times nor smaller stimulated regions are desirable.

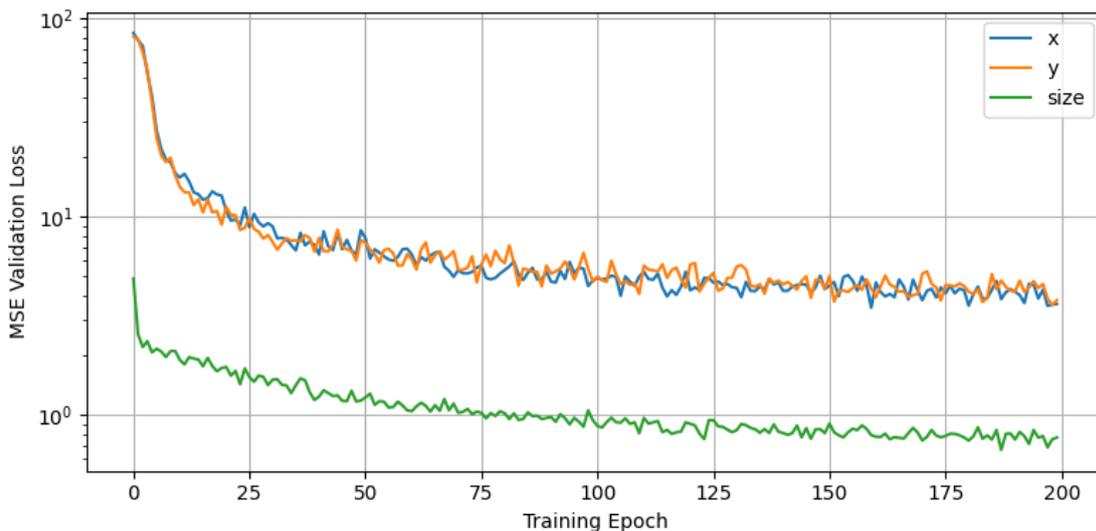


Figure 49: MSE validation loss separated into the losses for x , y and pRF size parameters for the **Stimsim24** stimulus. This shows the model struggles more with localization of center positions than with accurate size estimation.

4.2 Results on Stimsim24 Dataset

For the **Stimsim24** dataset, the difference between the top rated model and the second best model in terms of total rank-sums (see Table 5) is striking in the sense that no two variable features of the models are equal: They were trained with different noise models, different HRFs and different HRF delay ranges. However, when only taking the mean and median variance-explained rank-sums into account without the number of voxels above a threshold of 0.1 variance-explained, 3 of the top 4 models were trained with the **mrVista** HRF, indicating this to be an important factor for the goodness-of-fit on this dataset. This behaviour exemplifies the need for thorough analysis of **DeepRF** results and the fact that no single model can be a priori identified as best for a given dataset. Further, it is notable that the **SnapRF** model identified as best is a sub-model of the **RVy** model, which was trained with HRF delay in contrast to the best performing **RVn** model. This indicates that an HRF delay of zero seconds is a reasonable assumption for this dataset. This result might further indicate the need to train models with HRF delay with a less uniform sampling of delays, so that they can still recognize and successfully fit small delay values.

On the cortex plots (Fig. 42 and 43) it is clearly visible that **SnapRF** is capable of reproducing the retinotopy of the visual cortex similar to **mrVista**. There exist some small differences, but overall the results are highly comparable. This indicates the promise **SnapRF** has for fast and reliable pRF mapping results.

In the inter-method comparison of pRF estimates and derived quantities (Fig. 44), the overall agreement between the two methods is promising. Stark deviations are only visible in the size estimates, where **mrVista**'s estimates go to larger sizes, which can be easily explained by the size limit in **SnapRF** training, as well as in the variance-explained of estimates. The latter, of course, is due to **mrVista**'s optimization of the parameter and **SnapRF**'s ignorance to it.

From the mean deviations of estimates across 30 runs from the estimates on the averaged run (Fig. 45), several features can be identified. For the deviations in x ,

both methods scatter broadly with growing eccentricity, though **mrVista** does so a bit wider. In the x and y deviations in dependence of polar angle, both methods show a similar behaviour to that seen on the simulated data: Close to the y axes, the error in x is maximal, and close to the x axes, that in y is maximal. For the deviations in dependence of size, **SnapRF** shows a bias towards larger sizes for small eccentricities and high variance-explained values of the reference averaged run. **mrVista**, on the other hand, is more likely to underestimate the size compared to the reference run for larger eccentricities and larger sizes, as well as for lower variance-explained values. Further, the **mrVista** estimates go to larger sizes than the **SnapRF** estimates on the averaged run, because of the size limit in **SnapRF** training.

Finally, for the deviations in dependence of variance-explained, both methods fare comparably well, only in the size deviations does **mrVista** show a slight bias to larger sizes in the voxels with low variance-explained, and **SnapRF** a slight bias to larger sizes for higher variance-explained values.

Together, these results suggest that **SnapRF** gives comparable results to **mrVista** on low-noise empirical data in only a fraction of the time.

4.3 Results on NYU Retinotopy Dataset

For the NYU retinotopy dataset, the most important characteristic in the performance of **SnapRF** models seems to be the noise model and the HRF delay (Table 6). Here, as in the **Stimsim24** dataset, a **SnapRF** model with the new RS noise model achieved the best overall results. In contrast to the **Stimsim24** dataset however, the two top rated models across all categories (mean and median variance-explained, number above threshold) both feature a non-zero HRF delay in training, indicating that the data in this dataset is well explained by assuming a delay of the HRF.

In this dataset, the training losses are consistently higher than those in the **Stimsim24** dataset. Therefore, it is expected that mapping results with **SnapRF** are less reliable here than in the **Stimsim24** dataset.

To investigate the influence of stimulus configurations on the training loss, different stimuli were created from the **Stimsim24** and NYU stimuli by slicing the individual stimulus passes and putting them back together in different orders. These orders and their properties can be seen in Table 7.

Table 7: Properties of different stimulus configurations.

Configuration	Vertical Sweeps	Length [s]
original	Full	240
original reduced	Half	192
shuffled	Full	240
shuffled reduced	Half	192

The associated training losses for each of those configurations can be seen in Fig. 50. From these losses, it is clear the stimulus configuration only has a marginal influence on **SnapRF** training performance. Therefore, the reason for the worse training performance for the NYU stimulus must lie in the stimulus properties itself. In comparison with those of the **Stimsim24** stimulus, the NYU stimulus is almost twice as broad due to the larger area of visual field covered. Therefore, the resolution of the NYU stimulus in regards to visual field position as well as pRF size is lower

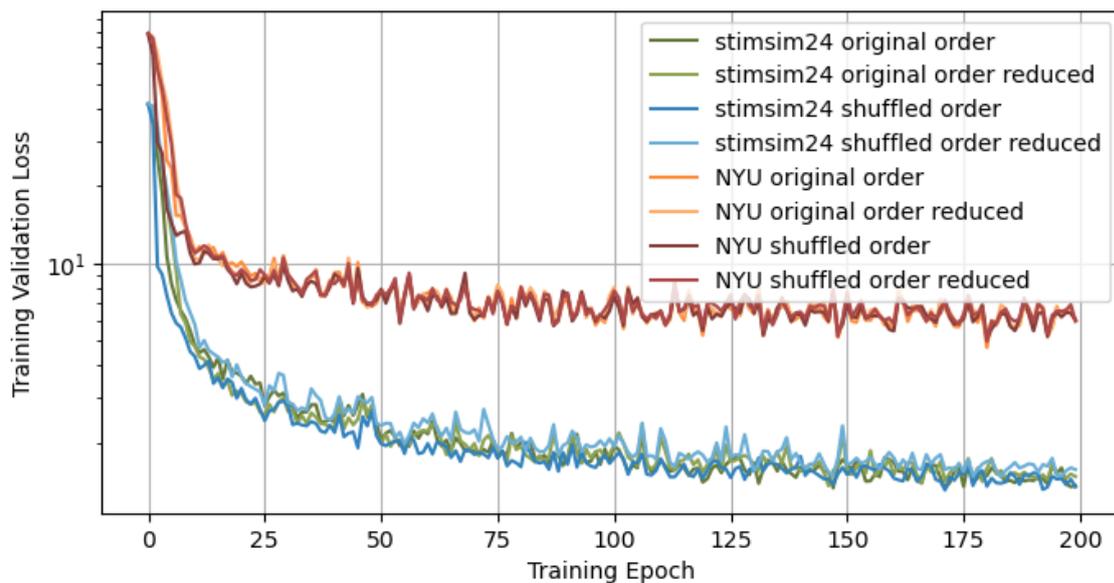


Figure 50: MSE validation loss for different configurations of the *Stimsim24* and NYU retinotopy dataset stimuli showing marginal influence of stimulus configuration compared to stimulus width on training performance.

than in the *Stimsim24* stimulus, which is visible in the *SnapRF* training loss when splitting it into the individual parameters (Fig. 51). When further splitting the pRF size loss function into bins of 1° , it is apparent that the higher loss for size predictions with the NYU retinotopy stimulus is mainly driven by small sizes, which cannot be resolved correctly due to the broader stimulus (Fig. 52). As with the *Stimsim24* stimulus, the losses seen here can be well explained quantitatively by the stimulus width: The mean squared errors in pRF size and x and y positions in Fig. 51 match the squared half width of the stimulus and squared width of the stimulus reasonably well. Any additional error is likely due to noise.

Of course, the lower spatial resolution of the stimulus will also affect the results of *mrVista*, but without any indication, so that this effect passes silently for standard analysis software. Thus, *DeepRF* training performance can be used to assess different stimulus designs and thereby help improve pRF mapping reliability.

In the inter-method reproducibilities for both subject median and group average results, similar features can be seen (Fig. 47). The winding shape of the eccentricity estimate comparison and the disparity in the polar angle estimates between the two methods are in stark contrast to the almost linear relationship between these estimates on the *Stimsim24* dataset. Interestingly, both methods have smaller maximum size estimates in this dataset, and an overall good agreement between the two. In variance-explained, however, the same phenomenon can be seen as in Fig. 44, where *mrVista* estimates almost always have higher variance-explained values than those of *SnapRF*.

The differences between the methods seen here are likely due to the relatively poor *SnapRF* training performance on the NYU retinotopy dataset. However, due to the broad stimulus used in this dataset, it is also possible that *mrVista* estimates are less reliable here than in the *Stimsim24* dataset.

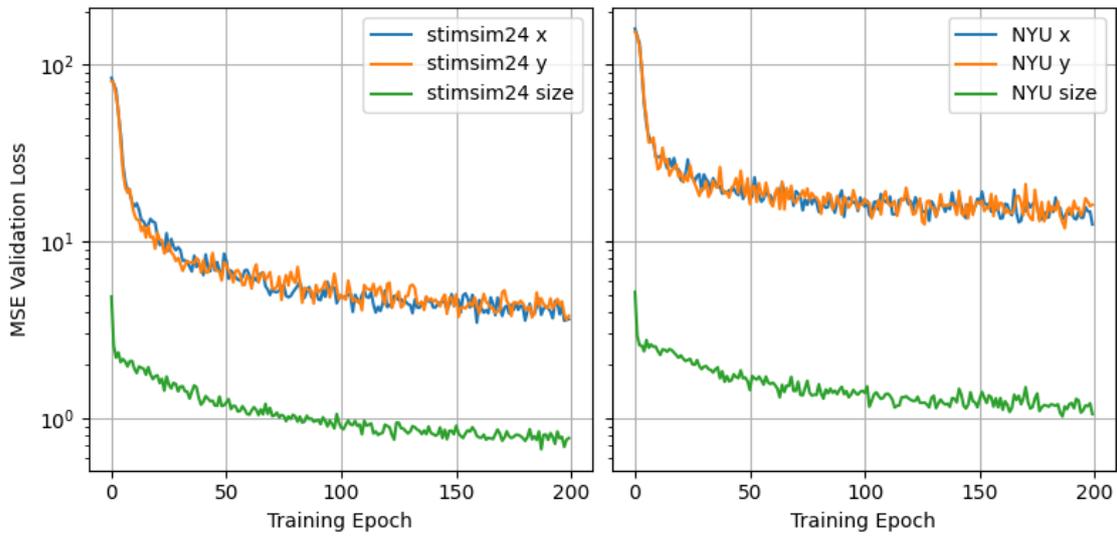


Figure 51: MSE validation loss separated into the losses for x , y and pRF size parameters for the Stimsim24 stimulus (left) and the NYU retinotopy dataset stimulus (right). Both center localization and size estimation is worse in training with the NYU stimulus compared to training with the Stimsim24 stimulus, though center localization more so than size estimation.

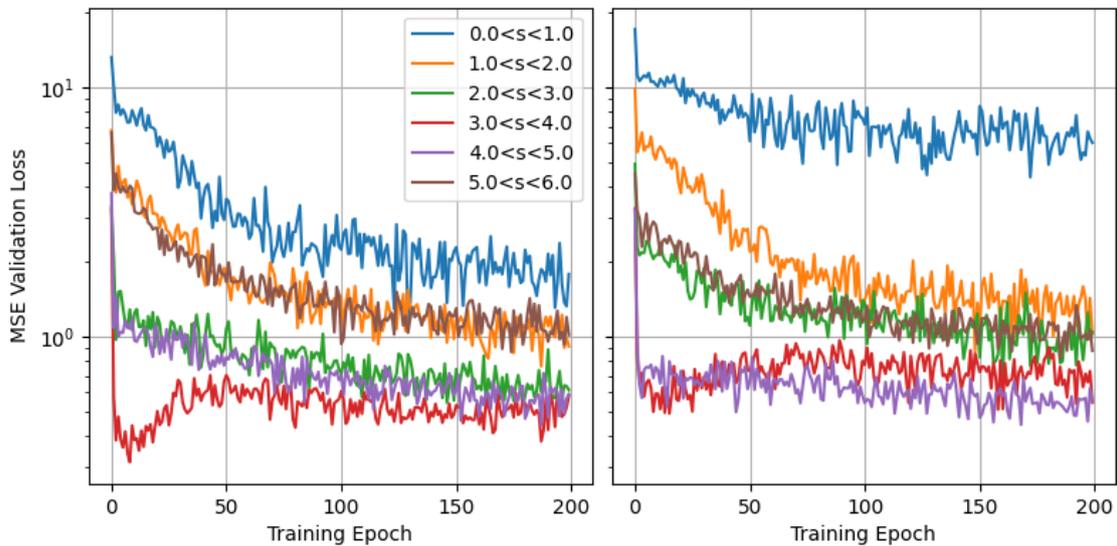


Figure 52: MSE pRF size validation loss separated into size bins of the ground truth for the Stimsim24 stimulus (left) and the NYU retinotopy dataset stimulus (right). In the legend, s describes the true size with which the timecourses were generated. Higher size estimation losses with the NYU stimulus are mainly driven by small sizes, which cannot be resolved due to the broader stimulus.

5 Conclusion and Outlook

In this thesis, the DeepRF approach for pRF mapping was reimplemented as SnapRF in order to overcome limitations of the original implementation in regards to the fairness and scope of comparisons with reference methods. In doing so, the approach was expanded and thoroughly tested on different datasets. To assess its suitability for pRF mapping applications under realistic conditions, its results were compared to that of the standard analysis tool mrVista.

On simulated data with known ground truth, SnapRF was shown to perform worse than mrVista under a wide range of noise conditions. Especially in the size estimates can a bias of SnapRF towards larger predictions be observed, which could feasibly be removed by training a SnapRF model with more datapoints generated with smaller pRF sizes. This, however, would introduce a bias towards smaller sizes, which is why it was not implemented for the purposes of this thesis, as the goal was to assess the model without artificial biases. Due to the multi-stage fitting approach of mrVista, however, the simulation study has shown SnapRF results at larger sizes to be slightly more reliable than those of the reference method.

An important part of the simulation study worthy of highlighting was the newly created RS noise model described in 2.2, which features noise components from resting state fMRI measurements and enabled a fair comparison between the two methods. Such an approach to modelling noise in fMRI simulation studies will doubtless prove useful for future studies. An important limitation of the RS noise model in its current form, however, is the small number of freely available datasets with a wide range of different properties (number of participants, sequence used, measurement site, etc.). Due to this small sample size, it is possible that features seen in a few subjects (such as relatively large movements or above-average heart rate, etc.) might have a strong influence on the characteristics of the generated noise. Nevertheless, the merits of the new RS noise model were proven in the analyses on empirical datasets. On these, SnapRF models trained with the RS noise model exhibited best performance in terms of a combined metric depending on the variance-explained of pRF estimates compared to models trained with other noise sources. These comparisons highlight that no "one-size-fits-all" approach can be employed where a single SnapRF model is suitable for all datasets. With this in mind, the results of comparisons on the simulated data might also indicate problems in SnapRF's ability to generalize on data with different noise characteristics than those it was exposed to during training. This further emphasizes the need for careful assessment of SnapRF model suitability for individual datasets.

On the Stimsim24 dataset, SnapRF was shown to be able to reproduce the retinotopic organization of the visual cortex in good agreement with mrVista. Further, overall good agreement between the methods was shown on a high SNR averaged measurement and in analyzing the reproducibility of estimates over a large number of measurements of single subjects. Additionally, due to SnapRF's pRF size limit in training, the pRF estimate results show better reproducibility in dependence of size than those of mrVista. It is important to note that the size limit in SnapRF is grounded in the physiological properties of the primary visual cortex, which this thesis focused on.

On the NYU retinotopy dataset, the agreement between the method estimates was less pronounced than expected based on the previous results. This discrepancy

can likely be explained by the properties of the stimulus used for this dataset. The dependence of **SnapRF** training success on visual stimulus design was uncovered by training **SnapRF** models with different stimulus configurations and by splitting the loss function into the individual pRF parameters and further into bins of the pRF size parameter. It was shown that broader stimuli lead to worse **SnapRF** performance due to the lower spatial resolution afforded by them compared to narrower stimuli. Standard analysis procedures are likely influenced in the same way, but without the warning signs present in the model training performance. Therefore, it is proposed that **DeepRF** training performance may serve as a guideline in pRF stimulus design. However, one has to note that **SnapRF** training performance is also influenced by training hyperparameters, such as the learning rate, the network architecture, and so on. Nevertheless, the assessment of training losses may help in designing ideal stimuli for best result reliability.

The results and interpretations presented herein open the door for multiple future research directions. New research could be conducted into **DeepRF**'s performance on other visual areas and the possibility to delineate different visual areas in the human brain based on **DeepRF** results. Further, the method holds great promise for the assessment of different stimulus designs, which might also benefit standard pRF mapping approaches. Finally, it might be worthwhile to investigate other network types as a basis, as several new ANN architectures have been developed since **DeepRF**'s inception. Transformer models, for example, have been shown to outperform ResNet architectures on computer vision tasks [109]. It is therefore conceivable that they might perform better on the task of pRF mapping as well.

While the results obtained in this thesis do not show **SnapRF** to be capable of better performance than the standard approach, the method still holds potential for future applications due to the substantial 191-fold reduction of testing time compared to **mrVista**. Further, the reusability of a once trained model to any dataset acquired with the same stimulus is an attractive feature of the method beyond the increase in testing speed.

In conclusion, the newly implemented RS noise model appears to be a better option than the **neuRosim** noise models in training for application on empirical data and is therefore suggested for future studies focused on machine learning approaches in fMRI analyses. However, the relatively poor performance of a model trained with one noise model on data generated with the other indicates a poor generalization of trained models to other noise sources than the ones trained with, so that a simple one-size-fits-all approach is not recommended when applying this method. The new noise model is further suggested for simulation studies as a way to remove biases towards machine learning models due to training procedures featuring the same noise types as simulated test data. Further development of the **DeepRF** approach might help in tackling the challenges the method faces and establish it as a reliable and power-saving alternative to current methods for pRF mapping.

List of Figures

1	Experimental setup for the Stern-Gerlach Experiment (adapted from [6]). Silver atoms are emitted from a furnace (1) and are collimated into a beam of atoms (2) before entering an inhomogeneous magnetic field (3). Quantum mechanics predicts a splitting of the beam into two separate components (4) while classically a broadening of the beam on the detector screen is expected (5).	2
2	Experimental evidence for spin quantization of silver atoms in an inhomogeneous magnetic field [5]. In the scale below, 20 units are equal to 1 mm.	3
3	Larmor precession of a nuclear spin around an external field with precession frequency ω_0 (adapted from [17]).	5
4	Schematic of Rabi's magnetic resonance method setup [19]. Magnets <i>A</i> and <i>B</i> create inhomogeneous magnetic fields with anti-parallel gradients, magnet <i>C</i> creates the oscillating field. The solid curves represent paths of molecules without change of magnetic moments in the apparatus. The dashed curves represent paths of molecules whose magnetic moments were changed in the excitatory field of the <i>C</i> magnet. Such molecules are not refocused onto the detector.	6
5	First observation of nuclear magnetic resonance [14]. At the resonance condition, a sharp drop in beam intensity at the detector is visible, as molecules are not refocused onto the detector screen.	6
6	Establishment of phase coherence between children of swings through pushes applied at the resonance frequency.	8
7	Tipping of the net magnetization vector in a coordinate system rotating with the Larmor frequency [17].	8
8	Relaxation or recovery of longitudinal magnetization M_z after a $\frac{\pi}{2}$ excitation pulse of the resonant RF field or after exposure to a static external magnetic field, respectively [17].	9
9	Occurrence of free induction decay after one $\frac{\pi}{2}$ pulse and a spin echo after a second rephasing pulse that reverses dephasing due to magnetic field inhomogeneities (adapted from [23]).	10
10	Emergence of a gradient echo after reversing a dephasing magnetic field gradient (adapted from [26]).	11
11	Measurement of T_2^* with a train of gradient echoes created through application of dephasing (De) and rephasing (Re) gradients after excitation with an RF pulse with a flip angle of α (adapted from [27]).	11
12	Schematic representation of how overlapping frequency and phase encoding magnetic field gradients in a selected slice define three-dimensional voxels of space that can be selectively measured (adapted from [32]). At step 0, the phase encoding gradient is not active, and the resulting measured signal $S_0(t)$ is a linear combination of the signals from voxels <i>A</i> and <i>B</i> . At step 1, the phase of voxel <i>B</i> was shifted by 180° relative to voxel <i>A</i> by the phase encoding gradient, and the measured signal $S_1(t)$ is the difference between the two individual signals. The combination of measurements allows the calculation of individual voxel signals.	13

13	Central parts of k-space (a) encode general features like contrast in images (b), while peripheral parts (c) encodes details but little contrast (d) (adapted from [33]). Arrows represent Fourier transforms between k-space and real space and vice-versa.	13
14	Rectangular motion through k-space in the echo planar imaging (EPI) technique [36].	14
15	T_1 and T_2 weighted images of the same brain slice [37].	14
16	Normalized canonical hemodynamic response function (HRF) that models the increase in bloodflow to an active brain region following a short stimulus.	16
17	The visual pathway in humans, shown from the retina to the visual cortex [54]. Note the contralateral endpoints for left and right hemifields, respectively.	17
18	Retinotopic maps of a dog created by Munk in the late 1870s [56]. The maps show the reorganization of the optic nerves at the optic chiasm (Ch) and the contralateral endpoints for input from the left an right hemifields, respectively. Additionally, the retinotopic organization of visual space on the visual cortex is shown.	18
19	One of the first retinotopic maps in humans, created by Holmes in 1944 from lesion studies of soldiers [60]. The retinotopic organization of the visual cortex is shown both in eccentricity and polar angle for the right visual hemifield.	18
20	Periodic stimulus (a) evokes observed wave of neural activity (b) [63].	19
21	Typical stimuli used in pRF mapping experiments [70].	20
22	Sweeping bar stimulus used in pRF mapping experiments (adapted from [70]).	20
23	Simulated timecourses for the same voxel with different noise levels. Note the peaks introduced by the higher noise that may be fitted as signal.	21
24	Artificial neuron that applies a nonlinear function to a weighted sum of inputs.	22
25	Schematic representation of a biological neuron in the brain.	23
26	Artificial neural network (ANN) constructed from layers of artificial neurons (ANs), with one input layer, a single hidden layer, and one output layer.	23
27	Schematic representation of a layer with a convolutional kernel filter in a convolutional neural network (CNN) (adapted from [73]).	24
28	Schematic representation of padding before convolution in a CNN (adapted from [73]). Padding can be used to increase the image size (a) or to keep it the same (b) after convolution with a filter kernel.	24
29	Schematic representation of pooling operations in a convolutional neural network (CNN).	25
30	Different nonlinear activation funtions.	26
31	Schematic representation of a block of layers in a residual neural network (ResNet) with identity skip connection [81]. The output of such a building block is given by the sum of the functions applied by the layers and the input to the layers itself.	26

32	Comparison of HRFs used by DeepRF and mrVista to model stimulation at $Time = 0$. Both HRFs have been used for this thesis.	29
33	Sweeping bar stimulus used in the Stimsim24 dataset (adapted from [70]).	33
34	Sweeping bar stimulus used in the NYU retinotopy dataset (adapted from [70]).	35
35	Background consisting of natural images over pink noise revealed by the stimulus aperture in the NYU retinotopy dataset.	35
36	SnapRF MSE validation loss for different learning rates (LR) showing a model with a learning rate of 0.001 converging faster and to lower loss values than with other learning rates.	40
37	Mean variance-explained values of 1000 voxels for different SnapRF CNR values and noise models showing comparable.	40
38	Almost linear relationship between number of voxels in a test set and the time taken for applying SnapRF in seconds.	41
39	MSE validation loss of model used for evaluation of SnapRF on simulated data converging towards a value close to 1. This is a higher loss than in the original DeepRF results but expectedly so due to higher noise and larger search radius in training SnapRF	42
40	pRF estimate errors for mrVista and SnapRF-NVn on 10,000 simulated voxels in dependence of eccentricity, pRF size, and variance-explained. Moving percentiles are also plotted.	43
41	MSE validation loss of model used for evaluation of SnapRF on the Stimsim24 dataset showing convergence to a value below 2. This is higher than in the simulation study, likely owing to the different type of noise used in training.	45
42	Comparison of eccentricity estimates of mrVista (top) and SnapRF-RVn (bottom) on the inflated brain surface of subject 001 for the averaged run.	45
43	Comparison of eccentricity estimates of mrVista (top) and SnapRF-RVn (bottom) on the inflated brain surface of subject 002 for the averaged run.	46
44	Voxel-wise comparison of pRF estimates of mrVista and SnapRF-RVn on averaged run for both Stimsim24 subjects. Only voxels with variance-explained > 0.1 for both methods are plotted. Blue voxels are from the left hemisphere, and orange voxels from the right hemisphere.	46
45	Mean pRF estimate deviations over 30 runs from estimates on averaged run in dependence of eccentricity, pRF size, and variance-explained of averaged run estimates. Data for both subjects of the Stimsim24 dataset were pooled. Moving percentiles are plotted additionally.	47
46	MSE validation loss of model used for evaluation of SnapRF on the NYU retinotopy dataset reaching lowest values of about 7, which is much higher than in previous datasets, indicating less predictive power of the trained model in this dataset.	49

47	Comparison of SnapRF-RDy and mrVista voxel-wise median subject pRF mapping results (top) and group-average pRF mapping results (bottom) on the NYU retinotopy dataset. Only voxels with variance-explained > 0.1 for both methods are plotted. Blue points are voxels from the left hemisphere, and orange ones from the right hemisphere.	49
48	Minimum MSE validation loss of SnapRF models trained with different combinations of search radius and CNR values for the Stimsim24 stimulus, showing how lower CNR values and larger search radii in training decrease model training performance.	52
49	MSE validation loss separated into the losses for x , y and pRF size parameters for the Stimsim24 stimulus. This shows the model struggles more with localization of center positions than with accurate size estimation.	53
50	MSE validation loss for different configurations of the Stimsim24 and NYU retinotopy dataset stimuli showing marginal influence of stimulus configuration compared to stimulus width on training performance.	55
51	MSE validation loss separated into the losses for x , y and pRF size parameters for the Stimsim24 stimulus (left) and the NYU retinotopy dataset stimulus (right). Both center localization and size estimation is worse in training with the NYU stimulus compared to training with the Stimsim24 stimulus, though center localization more so than size estimation.	56
52	MSE pRF size validation loss separated into size bins of the ground truth for the Stimsim24 stimulus (left) and the NYU retinotopy dataset stimulus (right). In the legend, s describes the true size with which the timecourses were generated. Higher size estimation losses with the NYU stimulus are mainly driven by small sizes, which cannot be resolved due to the broader stimulus.	56

List of Tables

1	SnapRF models trained per dataset.	31
2	Openneuro datasets used for RS noise models.	31
2	Openneuro datasets used for RS noise models (<i>Continuation</i>).	32
3	SnapRF pure testing times on simulated data.	41
4	SnapRF and mrVista testing times on simulated dataset.	42
5	Variance-explained Rank-Sums for SnapRF Models on the Stimsim24 Dataset.	44
6	Variance-explained Rank-Sums for SnapRF Models on the NYU Retinotopy Dataset.	48
7	Properties of different stimulus configurations.	54

Acronyms

- ANN** Artificial Neural Network. 22, 23, 25
- BIDS** Brain Imaging Data Structure. 28
- BOLD** Blood Oxygenation Level Dependent. 15, 33–37
- CBF** Cerebral Blood Flow. 15
- CBV** Cerebral Blood Volume. 15
- CNN** Convolutional Neural Network. 23, 24
- CNR** Contrast-to-Noise Ratio. 20, 29, 38, 39, 59
- DCNN** Deep Convolutional Neural Network. 24
- EPI** Echo Planar Imaging. 13, 15, 32, 33, 35, 36
- FID** Free Induction Decay. 10
- fMRI** Functional Magnetic Resonance Imaging. iii, 1, 15–17, 19, 20, 27–30, 32–36, 55, 56
- GLM** General Linear Model. 19
- GPU** Graphical Processing Unit. 27, 28
- HRF** Hemodynamic Response Function. 15, 16, 28, 29, 31, 32, 40, 51, 52, 58, 59
- LGN** Lateral Geniculate Nucleus. 16
- ML** Machine Learning. 21, 22
- MRI** Magnetic Resonance Imaging. 12, 15, 27, 30, 32, 33
- MSE** Mean Squared Error. 30, 38, 43, 47, 49, 50, 54, 60
- NMR** Nuclear Magnetic Resonance. 2, 4–7, 15
- NORDIC** NOise Reduction with DIstribution Corrected PCA. 33
- PC** Principal Component. 30
- PCA** Principal Component Analysis. iii, iv, 30
- PET** Positron Emission Tomography. 15
- pRF** Population Receptive Field. iii, iv, 19, 20, 27–30, 33, 35, 38–43, 45–47, 49–51, 53–56, 58–60

- ReLU** Rectified Linear Unit. 25
- ResNet** Residual Neural Network. 25
- RS** Resting State. 30–32, 34, 36–38, 49, 52, 55
- RSS** Residual Sum of Squares. 19, 20, 27
- SGE** Stern-Gerlach Experiment. 2–4
- SNR** Signal-to-Noise Ratio. 35, 43, 55
- TE** Echo Time. 31–33, 35
- TR** Repetition Time. 15, 30–33, 35

References

- [1] J. Thielen, U. Güçlü, Y. Güçlütürk, L. Ambrogioni, S. E. Bosch, and M. A. J. van Gerven. “DeepRF: Ultrafast population receptive field mapping with deep learning”. In: (Aug. 2019). DOI: 10.1101/732990. URL: <http://dx.doi.org/10.1101/732990>.
- [2] F. Bloch. “Nuclear Induction”. In: *Physical Review* 70.7–8 (Oct. 1946), pp. 460–474. ISSN: 0031-899X. DOI: 10.1103/physrev.70.460. URL: <http://dx.doi.org/10.1103/PhysRev.70.460>.
- [3] A. Abragam and H. Y. Carr. “The Principles of Nuclear Magnetism”. In: *Physics Today* 14.11 (Nov. 1961), pp. 56–58. ISSN: 1945-0699. DOI: 10.1063/1.3057238. URL: <http://dx.doi.org/10.1063/1.3057238>.
- [4] Z. P. Liang and P. C. Lauterbur. *Principles of magnetic resonance imaging*. IEEE Press Series on Biomedical Engineering. Piscataway, NJ: IEEE Publications, Oct. 1999.
- [5] W. Gerlach and O. Stern. “Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld”. In: *Zeitschrift für Physik* 9.1 (Dec. 1922), pp. 349–352. ISSN: 1434-601X. DOI: 10.1007/bf01326983. URL: <http://dx.doi.org/10.1007/BF01326983>.
- [6] Tatoute. *Illustration of Stern-Gerlach experiment*. accessed on February 2, 2025. 2014. URL: https://commons.wikimedia.org/wiki/File:Stern-Gerlach_experiment.svg.
- [7] O. Stern. “Ein Weg zur experimentellen Prüfung der Richtungsquantelung im Magnetfeld”. In: *Zeitschrift für Physik* 7.1 (Dec. 1921), pp. 249–253. ISSN: 1434-601X. DOI: 10.1007/bf01332793. URL: <http://dx.doi.org/10.1007/BF01332793>.
- [8] G. Uhlenbeck and S. Goudsmit. “Ersetzung der Hypothese vom unmechanischen Zwang durch eine Forderung bezüglich des inneren Verhaltens jedes einzelnen Elektrons”. In: *Die Naturwissenschaften* 13.47 (Nov. 1925), pp. 953–954. ISSN: 1432-1904. DOI: 10.1007/bf01558878. URL: <http://dx.doi.org/10.1007/BF01558878>.
- [9] T. E. Phipps and O. Stern. “Über die Einstellung der Richtungsquantelung”. In: *Zeitschrift für Physik* 73.3–4 (Mar. 1932), pp. 185–191. ISSN: 1434-601X. DOI: 10.1007/bf01351212. URL: <http://dx.doi.org/10.1007/BF01351212>.
- [10] R. Frisch and E. Segre. “Über die Einstellung der Richtungsquantelung. II”. In: *Zeitschrift für Physik* 80.9–10 (Sept. 1933), pp. 610–616. ISSN: 1434-601X. DOI: 10.1007/bf01335699. URL: <http://dx.doi.org/10.1007/BF01335699>.
- [11] P. Güttinger. “Das Verhalten von Atomen im magnetischen Drehfeld”. In: *Zeitschrift für Physik* 73.3–4 (Mar. 1932), pp. 169–184. ISSN: 1434-601X. DOI: 10.1007/bf01351211. URL: <http://dx.doi.org/10.1007/BF01351211>.
- [12] E. Majorana. “Atomi orientati in campo magnetico variabile”. In: *Il Nuovo Cimento* 9.2 (Feb. 1932), pp. 43–50. ISSN: 1827-6121. DOI: 10.1007/bf02960953. URL: <http://dx.doi.org/10.1007/BF02960953>.

- [13] I. I. Rabi. “On the Process of Space Quantization”. In: *Physical Review* 49.4 (Feb. 1936), pp. 324–328. ISSN: 0031-899X. DOI: 10.1103/physrev.49.324. URL: <http://dx.doi.org/10.1103/PhysRev.49.324>.
- [14] I. I. Rabi, J. R. Zacharias, S. Millman, and P. Kusch. “A New Method of Measuring Nuclear Magnetic Moment”. In: *Physical Review* 53.4 (Feb. 1938), pp. 318–318. ISSN: 0031-899X. DOI: 10.1103/physrev.53.318. URL: <http://dx.doi.org/10.1103/PhysRev.53.318>.
- [15] C. J. Gorter and L. J. F. Broer. “Negative result of an attempt to observe nuclear magnetic resonance in solids”. In: *Physica* 9.6 (June 1942), pp. 591–596. ISSN: 0031-8914. DOI: 10.1016/s0031-8914(42)80073-7. URL: [http://dx.doi.org/10.1016/S0031-8914\(42\)80073-7](http://dx.doi.org/10.1016/S0031-8914(42)80073-7).
- [16] C. J. Gorter and Henry A. Boorse. “Bad luck in attempts to make scientific discoveries”. In: *Physics Today* 20.1 (Jan. 1967), pp. 76–81. ISSN: 1945-0699. DOI: 10.1063/1.3034140. URL: <http://dx.doi.org/10.1063/1.3034140>.
- [17] M. A. Brown and R. C. Semelka. *MRI: Basic Principles and Applications*. Wiley, Aug. 2003. ISBN: 9780471467939. DOI: 10.1002/0471467936. URL: <http://dx.doi.org/10.1002/0471467936>.
- [18] I. I. Rabi. “Space Quantization in a Gyating Magnetic Field”. In: *Physical Review* 51.8 (Apr. 1937), pp. 652–654. ISSN: 0031-899X. DOI: 10.1103/physrev.51.652. URL: <http://dx.doi.org/10.1103/PhysRev.51.652>.
- [19] I. I. Rabi, S. Millman, P. Kusch, and J. R. Zacharias. “The Molecular Beam Resonance Method for Measuring Nuclear Magnetic Moments.” In: *Physical Review* 55.6 (Mar. 1939), pp. 526–535. ISSN: 0031-899X. DOI: 10.1103/physrev.55.526. URL: <http://dx.doi.org/10.1103/PhysRev.55.526>.
- [20] E. M. Purcell, H. C. Torrey, and R. V. Pound. “Resonance Absorption by Nuclear Magnetic Moments in a Solid”. In: *Physical Review* 69.1–2 (Jan. 1946), pp. 37–38. ISSN: 0031-899X. DOI: 10.1103/physrev.69.37. URL: <http://dx.doi.org/10.1103/PhysRev.69.37>.
- [21] F. Bloch, W. W. Hansen, and M. Packard. “The Nuclear Induction Experiment”. In: *Physical Review* 70.7–8 (Oct. 1946), pp. 474–485. ISSN: 0031-899X. DOI: 10.1103/physrev.70.474. URL: <http://dx.doi.org/10.1103/PhysRev.70.474>.
- [22] E. L. Hahn. “Spin Echoes”. In: *Physical Review* 80.4 (Nov. 1950), pp. 580–594. ISSN: 0031-899X. DOI: 10.1103/physrev.80.580. URL: <http://dx.doi.org/10.1103/PhysRev.80.580>.
- [23] A. D. Elster. accessed on February 2, 2025. 2024. URL: <https://mriquestions.com/spin-echo1.html>.
- [24] S. Meiboom and D. Gill. “Modified Spin-Echo Method for Measuring Nuclear Relaxation Times”. In: *Review of Scientific Instruments* 29.8 (Aug. 1958), pp. 688–691. ISSN: 1089-7623. DOI: 10.1063/1.1716296. URL: <http://dx.doi.org/10.1063/1.1716296>.
- [25] E. L. Hahn. “Detection of sea-water motion by nuclear precession”. In: *Journal of Geophysical Research* 65.2 (Feb. 1960), pp. 776–777. ISSN: 0148-0227. DOI: 10.1029/jz065i002p00776. URL: <http://dx.doi.org/10.1029/JZ065i002p00776>.

- [26] A. D. Elster. accessed on February 2, 2025. 2024. URL: <https://mriquestions.com/gradient-echo.html>.
- [27] A. D. Elster. accessed on February 2, 2025. 2024. URL: <https://mriquestions.com/multi-echo-gre.html>.
- [28] J. Z. Bojorquez, S. Bricq, C. Acquitter, F. Brunotte, P. M. Walker, and A. Lalande. “What are normal relaxation times of tissues at 3 T?” In: *Magnetic Resonance Imaging* 35 (Jan. 2017), pp. 69–80. ISSN: 0730-725X. DOI: 10.1016/j.mri.2016.08.021. URL: <http://dx.doi.org/10.1016/j.mri.2016.08.021>.
- [29] P. C. LAUTERBUR. “Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance”. In: *Nature* 242.5394 (Mar. 1973), pp. 190–191. ISSN: 1476-4687. DOI: 10.1038/242190a0. URL: <http://dx.doi.org/10.1038/242190a0>.
- [30] P. Mansfield and P. K. Grannell. “NMR “diffraction” in solids?” In: *Journal of Physics C: Solid State Physics* 6.22 (Nov. 1973), pp. L422–L426. ISSN: 0022-3719. DOI: 10.1088/0022-3719/6/22/007. URL: <http://dx.doi.org/10.1088/0022-3719/6/22/007>.
- [31] A. D. Elster. accessed on February 2, 2025. 2024. URL: <https://mriquestions.com/how-to-locate-signals.html>.
- [32] A. D. Elster. accessed on February 2, 2025. 2024. URL: <https://mriquestions.com/what-is-phase-encoding.html>.
- [33] R. Mezrich. “A perspective on K-space.” In: *Radiology* 195.2 (May 1995), pp. 297–315. ISSN: 1527-1315. DOI: 10.1148/radiology.195.2.7724743. URL: <http://dx.doi.org/10.1148/radiology.195.2.7724743>.
- [34] P. Mansfield. “Multi-planar image formation using NMR spin echoes”. In: *Journal of Physics C: Solid State Physics* 10.3 (Feb. 1977), pp. L55–L58. ISSN: 0022-3719. DOI: 10.1088/0022-3719/10/3/004. URL: <http://dx.doi.org/10.1088/0022-3719/10/3/004>.
- [35] M. Barth, F. Breuer, P. J. Koopmans, D. G. Norris, and B. A. Poser. “Simultaneous multislice (SMS) imaging techniques”. In: *Magnetic Resonance in Medicine* 75.1 (Aug. 2015), pp. 63–81. ISSN: 1522-2594. DOI: 10.1002/mrm.25897. URL: <http://dx.doi.org/10.1002/mrm.25897>.
- [36] A. D. Elster. accessed on February 2, 2025. 2024. URL: <https://mriquestions.com/echo-planar-imaging.html>.
- [37] A. Jahn. accessed on January 6, 2025. 2019. URL: https://andysbrainbook.readthedocs.io/en/stable/MRtrix/MRtrix_Course/MRtrix_00_Diffusion_Overview.html.
- [38] P. T. Fox and M. E. Raichle. “Stimulus rate dependence of regional cerebral blood flow in human striate cortex, demonstrated by positron emission tomography”. In: *Journal of Neurophysiology* 51.5 (May 1984), pp. 1109–1120. ISSN: 1522-1598. DOI: 10.1152/jn.1984.51.5.1109. URL: <http://dx.doi.org/10.1152/jn.1984.51.5.1109>.

- [39] P. T. Fox and M. E. Raichle. “Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects.” In: *Proceedings of the National Academy of Sciences* 83.4 (Feb. 1986), pp. 1140–1144. ISSN: 1091-6490. DOI: 10.1073/pnas.83.4.1140. URL: <http://dx.doi.org/10.1073/pnas.83.4.1140>.
- [40] P. T. Fox, M. A. Mintun, M. E. Raichle, F. M. Miezin, J. M. Allman, and D. C. Van Essen. “Mapping human visual cortex with positron emission tomography”. In: *Nature* 323.6091 (Oct. 1986), pp. 806–809. ISSN: 1476-4687. DOI: 10.1038/323806a0. URL: <http://dx.doi.org/10.1038/323806a0>.
- [41] Arno Villringer, Bruce R. Rosen, John W. Belliveau, Jerome L. Ackerman, Randall B. Lauffer, Richard B. Buxton, Yong-Sheng Chao, Van J. Wedeenand, and Thomas J. Brady. “Dynamic imaging with lanthanide chelates in normal brain: Contrast due to magnetic susceptibility effects”. In: *Magnetic Resonance in Medicine* 6.2 (Feb. 1988), pp. 164–174. ISSN: 1522-2594. DOI: 10.1002/mrm.1910060205. URL: <http://dx.doi.org/10.1002/mrm.1910060205>.
- [42] John W. Belliveau, Bruce R. Rosen, Howard L. Kantor, Richard R. Rzedzian, David N. Kennedy, Robert C. McKinstry, James M. Vevea, Mark S. Cohen, Ian L. Pykett, and Thomas J. Brady. “Functional cerebral imaging by susceptibility-contrast NMR”. In: *Magnetic Resonance in Medicine* 14.3 (June 1990), pp. 538–546. ISSN: 1522-2594. DOI: 10.1002/mrm.1910140311. URL: <http://dx.doi.org/10.1002/mrm.1910140311>.
- [43] Bruce R. Rosen, John W. Belliveau, James M. Vevea, and Thomas J. Brady. “Perfusion imaging with NMR contrast agents”. In: *Magnetic Resonance in Medicine* 14.2 (May 1990), pp. 249–265. ISSN: 1522-2594. DOI: 10.1002/mrm.1910140211. URL: <http://dx.doi.org/10.1002/mrm.1910140211>.
- [44] J. W. Belliveau, D. N. Kennedy, R. C. McKinstry, B. R. Buchbinder, R. M. Weisskoff, M. S. Cohen, J. M. Vevea, T. J. Brady, and B. R. Rosen. “Functional Mapping of the Human Visual Cortex by Magnetic Resonance Imaging”. In: *Science* 254.5032 (Nov. 1991), pp. 716–719. ISSN: 1095-9203. DOI: 10.1126/science.1948051. URL: <http://dx.doi.org/10.1126/science.1948051>.
- [45] K. K. Kwong, A. L. Hopkins, J. W. Belliveau, D. A. Chesler, L. M. Porkka, R. C. McKinstry, D. A. Finelli, G. J. Hunter, J. B. Moore, R. G. Barr, and B. R. Rosen. “Proton NMR imaging of cerebral blood flow using H₂17O”. In: *Magnetic Resonance in Medicine* 22.1 (Nov. 1991), pp. 154–158. ISSN: 1522-2594. DOI: 10.1002/mrm.1910220116. URL: <http://dx.doi.org/10.1002/mrm.1910220116>.
- [46] Bruce R. Rosen, John W. Belliveau, Bradley R. Buchbinder, Robert C. McKinstry, Leena M. Porkka, David N. Kennedy, Michelle S. Neuder, C. Richard Fisel, Hannu J. Aronen, Kenneth K. Kwong, Robert M. Weisskoff, Mark S. Cohen, and Thomas J. Brady. “Contrast agents and cerebral hemodynamics”. In: *Magnetic Resonance in Medicine* 19.2 (June 1991), pp. 285–292. ISSN: 1522-2594. DOI: 10.1002/mrm.1910190216. URL: <http://dx.doi.org/10.1002/mrm.1910190216>.

- [47] L. Pauling and C. D. Coryell. “The Magnetic Properties and Structure of Hemoglobin, Oxyhemoglobin and Carbonmonoxyhemoglobin”. In: *Proceedings of the National Academy of Sciences* 22.4 (Apr. 1936), pp. 210–216. ISSN: 1091-6490. DOI: 10.1073/pnas.22.4.210. URL: <http://dx.doi.org/10.1073/pnas.22.4.210>.
- [48] Keith R. Thulborn, John C. Waterton, Paul M. Matthews, and George K. Radda. “Oxygenation dependence of the transverse relaxation time of water protons in whole blood at high field”. In: *Biochimica et Biophysica Acta (BBA) - General Subjects* 714.2 (Feb. 1982), pp. 265–270. ISSN: 0304-4165. DOI: 10.1016/0304-4165(82)90333-6. URL: [http://dx.doi.org/10.1016/0304-4165\(82\)90333-6](http://dx.doi.org/10.1016/0304-4165(82)90333-6).
- [49] S. Ogawa and T. M. Lee. “Magnetic resonance imaging of blood vessels at high fields: In vivo and in vitro measurements and image simulation”. In: *Magnetic Resonance in Medicine* 16.1 (Oct. 1990), pp. 9–18. ISSN: 1522-2594. DOI: 10.1002/mrm.1910160103. URL: <http://dx.doi.org/10.1002/mrm.1910160103>.
- [50] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann. “Neurophysiological investigation of the basis of the fMRI signal”. In: *Nature* 412.6843 (July 2001), pp. 150–157. ISSN: 1476-4687. DOI: 10.1038/35084005. URL: <http://dx.doi.org/10.1038/35084005>.
- [51] G. M. Boynton, S. A. Engel, G. H. Glover, and D. J. Heeger. “Linear Systems Analysis of Functional Magnetic Resonance Imaging in Human V1”. In: *The Journal of Neuroscience* 16.13 (July 1996), pp. 4207–4221. ISSN: 1529-2401. DOI: 10.1523/jneurosci.16-13-04207.1996. URL: <http://dx.doi.org/10.1523/JNEUROSCI.16-13-04207.1996>.
- [52] A. M. Dale and R. L. Buckner. “Selective averaging of rapidly presented individual trials using fMRI”. In: *Human Brain Mapping* 5.5 (1997), pp. 329–340. ISSN: 1065-9471. DOI: 10.1002/(sici)1097-0193(1997)5:5<329::aid-hbm1>3.0.co;2-5. URL: [http://dx.doi.org/10.1002/\(SICI\)1097-0193\(1997\)5:5%3C329::AID-HBM1%3E3.0.CO;2-5](http://dx.doi.org/10.1002/(SICI)1097-0193(1997)5:5%3C329::AID-HBM1%3E3.0.CO;2-5).
- [53] G. M. Boynton, S. A. Engel, and D. J. Heeger. “Linear systems analysis of the fMRI signal”. In: *NeuroImage* 62.2 (Aug. 2012), pp. 975–984. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2012.01.082. URL: <http://dx.doi.org/10.1016/j.neuroimage.2012.01.082>.
- [54] Mads00. *Neural pathway diagram*. accessed on January 6, 2025. 2016. URL: https://commons.wikimedia.org/wiki/File:Neural_pathway_diagram.svg#/media/File:Neural_pathway_diagram.svg.
- [55] H. Munk. *Über die Functionen der Grosshirnrinde: gesammelte Mittheilungen mit Anmerkungen*. Hirschwald, 1890. URL: <https://books.google.at/books?id=dPc-AAAAYAAJ>.
- [56] H. Munk. *Über die Functionen der Grosshirnrinde: gesammelte Mittheilungen mit Anmerkungen*. Hirschwald, 1890, p. 24. URL: <https://books.google.at/books?id=dPc-AAAAYAAJ>.

- [57] H. Munk. *Über die Functionen der Grosshirnrinde: gesammelte Mittheilungen mit Anmerkungen*. Hirschwald, 1890, pp. 89–90. URL: <https://books.google.at/books?id=dPc-AAAAYAAJ>.
- [58] T. Inouye. *Die Sehstörungen bei Schußverletzungen der kortikalen Sehsphäre*. urn:nbn:de:hbz:38m:1-55887. Leipzig: Wilhelm Engelmann, 1909.
- [59] G. Holmes. “Disturbances of visual orientation”. In: *British Journal of Ophthalmology* 2.9 (Sept. 1918), pp. 449–468. ISSN: 0007-1161. DOI: 10.1136/bjo.2.9.449. URL: <http://dx.doi.org/10.1136/bjo.2.9.449>.
- [60] G. Holmes. “The organization of the visual cortex in man”. In: *Proceedings of the Royal Society of London. Series B - Biological Sciences* 132.869 (Apr. 1945), pp. 348–361. ISSN: 2053-9193. DOI: 10.1098/rspb.1945.0002. URL: <http://dx.doi.org/10.1098/rspb.1945.0002>.
- [61] E. A. DeYoe, P. Bandettini, J. Neitz, D. Miller, and P. Winans. “Functional magnetic resonance imaging (fMRI) of the human brain”. In: *Journal of Neuroscience Methods* 54.2 (Oct. 1994), pp. 171–187. ISSN: 0165-0270. DOI: 10.1016/0165-0270(94)90191-0. URL: [http://dx.doi.org/10.1016/0165-0270\(94\)90191-0](http://dx.doi.org/10.1016/0165-0270(94)90191-0).
- [62] S. A. Engel, G. H. Glover, and B. A. Wandell. “Retinotopic organization in human visual cortex and the spatial precision of functional MRI”. In: *Cerebral Cortex* 7.2 (Mar. 1997), pp. 181–192. ISSN: 1460-2199. DOI: 10.1093/cercor/7.2.181. URL: <http://dx.doi.org/10.1093/cercor/7.2.181>.
- [63] S. A. Engel, D. E. Rumelhart, B. A. Wandell, A. T. Lee, G. H. Glover, E.-J. Chichilnisky, and M. N. Shadlen. “fMRI of human visual cortex”. In: *Nature* 369.6481 (June 1994), pp. 525–525. ISSN: 1476-4687. DOI: 10.1038/369525a0. URL: <http://dx.doi.org/10.1038/369525a0>.
- [64] M. I. Sereno, A. M. Dale, J. B. Reppas, K. K. Kwong, J. W. Belliveau, T. J. Brady, B. R. Rosen, and R. B. H. Tootell. “Borders of Multiple Visual Areas in Humans Revealed by Functional Magnetic Resonance Imaging”. In: *Science* 268.5212 (May 1995), pp. 889–893. ISSN: 1095-9203. DOI: 10.1126/science.7754376. URL: <http://dx.doi.org/10.1126/science.7754376>.
- [65] S. O. Dumoulin and B. A. Wandell. “Population receptive field estimates in human visual cortex”. In: *NeuroImage* 39.2 (Jan. 2008), pp. 647–660. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2007.09.034. URL: <http://dx.doi.org/10.1016/j.neuroimage.2007.09.034>.
- [66] J. D. Victor, K. Purpura, E. Katz, and B. Mao. “Population encoding of spatial frequency, orientation, and color in macaque V1”. In: *Journal of Neurophysiology* 72.5 (Nov. 1994), pp. 2151–2166. ISSN: 1522-1598. DOI: 10.1152/jn.1994.72.5.2151. URL: <http://dx.doi.org/10.1152/jn.1994.72.5.2151>.
- [67] W. Zuiderbaan, B. M. Harvey, and S. O. Dumoulin. “Modeling center-surround configurations in population receptive fields using fMRI”. In: *Journal of Vision* 12.3 (Mar. 2012), pp. 10–10. ISSN: 1534-7362. DOI: 10.1167/12.3.10. URL: <http://dx.doi.org/10.1167/12.3.10>.

- [68] K. N. Kay, J. Winawer, A. Mezer, and B. A. Wandell. “Compressive spatial summation in human visual cortex”. In: *Journal of Neurophysiology* 110.2 (July 2013), pp. 481–494. ISSN: 1522-1598. DOI: 10.1152/jn.00105.2013. URL: <http://dx.doi.org/10.1152/jn.00105.2013>.
- [69] P. Zeidman, E. H. Silson, D. S. Schwarzkopf, C. I. Baker, and W. Penny. “Bayesian population receptive field modelling”. In: *NeuroImage* 180 (Oct. 2018), pp. 173–187. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2017.09.008. URL: <http://dx.doi.org/10.1016/j.neuroimage.2017.09.008>.
- [70] D. Linhardt, M. Pawloff, A. Hummer, M. Woletz, M. Tik, M. Ritter, U. Schmidt-Erfurth, and C. Windischberger. “Combining stimulus types for improved coverage in population receptive field mapping”. In: *NeuroImage* 238 (Sept. 2021), p. 118240. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2021.118240. URL: <http://dx.doi.org/10.1016/j.neuroimage.2021.118240>.
- [71] K. Hornik, M. Stinchcombe, and H. White. “Multilayer feedforward networks are universal approximators”. In: *Neural Networks* 2.5 (Jan. 1989), pp. 359–366. ISSN: 0893-6080. DOI: 10.1016/0893-6080(89)90020-8. URL: [http://dx.doi.org/10.1016/0893-6080\(89\)90020-8](http://dx.doi.org/10.1016/0893-6080(89)90020-8).
- [72] G. Cybenko. “Approximation by superpositions of a sigmoidal function”. In: *Mathematics of Control, Signals, and Systems* 2.4 (Dec. 1989), pp. 303–314. ISSN: 1435-568X. DOI: 10.1007/bf02551274. URL: <http://dx.doi.org/10.1007/BF02551274>.
- [73] M. Stewart. accessed on January 23, 2025. 2019. URL: <https://towardsdatascience.com/simple-introduction-to-convolutional-neural-networks-cdf8d3077bac>.
- [74] X. Glorot and Y. Bengio. “Understanding the difficulty of training deep feed-forward neural networks”. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Ed. by Yee Whye Teh and Mike Titterton. Vol. 9. Proceedings of Machine Learning Research. Chia Laguna Resort, Sardinia, Italy: PMLR, 13–15 May 2010, pp. 249–256. URL: <https://proceedings.mlr.press/v9/glorot10a.html>.
- [75] S. K. Kumar. *On weight initialization in deep neural networks*. 2017. DOI: 10.48550/ARXIV.1704.08863. URL: <https://arxiv.org/abs/1704.08863>.
- [76] Ahmet Yilmaz and Riccardo Poli. “Successfully and efficiently training deep multi-layer perceptrons with logistic activation function simply requires initializing the weights with an appropriate negative mean”. In: *Neural Networks* 153 (Sept. 2022), pp. 87–103. ISSN: 0893-6080. DOI: 10.1016/j.neunet.2022.05.030. URL: <http://dx.doi.org/10.1016/j.neunet.2022.05.030>.
- [77] X. Glorot, A. Bordes, and Y. Bengio. “Deep Sparse Rectifier Neural Networks”. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Ed. by Geoffrey Gordon, David Dunson, and Miroslav Dudík. Vol. 15. Proceedings of Machine Learning Research. Fort Lauderdale, FL, USA: PMLR, Nov. 2011, pp. 315–323. URL: <https://proceedings.mlr.press/v15/glorot11a.html>.

- [78] S. Ioffe and C. Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. 2015. DOI: 10.48550/ARXIV.1502.03167. URL: <https://arxiv.org/abs/1502.03167>.
- [79] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*. 2015. DOI: 10.48550/ARXIV.1502.01852. URL: <https://arxiv.org/abs/1502.01852>.
- [80] Kaiming He and Jian Sun. *Convolutional Neural Networks at Constrained Time Cost*. 2014. DOI: 10.48550/ARXIV.1412.1710. URL: <https://arxiv.org/abs/1412.1710>.
- [81] K. He, X. Zhang, S. Ren, and J. Sun. *Deep Residual Learning for Image Recognition*. 2015. DOI: 10.48550/ARXIV.1512.03385. URL: <https://arxiv.org/abs/1512.03385>.
- [82] A. Veit, M. Wilber, and S. Belongie. *Residual Networks Behave Like Ensembles of Relatively Shallow Networks*. 2016. DOI: 10.48550/ARXIV.1605.06431. URL: <https://arxiv.org/abs/1605.06431>.
- [83] G. Lerma-Usabiaga, N. Benson, J. Winawer, and B. A. Wandell. “A validation framework for neuroimaging software: The case of population receptive fields”. In: *PLOS Computational Biology* 16.6 (June 2020). Ed. by Saad Jbabdi, e1007924. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1007924. URL: <http://dx.doi.org/10.1371/journal.pcbi.1007924>.
- [84] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization*. 2014. DOI: 10.48550/ARXIV.1412.6980. URL: <https://arxiv.org/abs/1412.6980>.
- [85] M. M. Himmelberg, J. W. Kurzawski, N. C. Benson, D. G. Pelli, M. Carrasco, and J. Winawer. “Cross-dataset reproducibility of human retinotopic maps”. In: *NeuroImage* 244 (Dec. 2021), p. 118609. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2021.118609. URL: <http://dx.doi.org/10.1016/j.neuroimage.2021.118609>.
- [86] M. Welvaert, J. Durnez, B. Moerkerke, G. Verdoolaege, and Y. Rosseel. “neurosim: AnRPackage for Generating fMRI Data”. In: *Journal of Statistical Software* 44.10 (2011). ISSN: 1548-7660. DOI: 10.18637/jss.v044.i10. URL: <http://dx.doi.org/10.18637/jss.v044.i10>.
- [87] S. Moeller, P. K. Pisharady, S. Ramanna, C. Lenglet, X. Wu, L. Dowdle, E. Yacoub, K. Uğurbil, and M. Akcakaya. “NOise reduction with Distribution Corrected (NORDIC) PCA in dMRI with complex-valued parameter-free locally low-rank processing”. In: *NeuroImage* 226 (Feb. 2021), p. 117539. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2020.117539. URL: <http://dx.doi.org/10.1016/j.neuroimage.2020.117539>.
- [88] O. Esteban, C. Markiewicz, R. W. Blair, C. Moodie, A. I. Isik, A. Erramuzpe Aliaga, J. Kent, M. Goncalves, E. DuPre, M. Snyder, H. Oya, S. Ghosh, J. Wright, J. Durnez, R. Poldrack, and K. J. Gorgolewski. “fMRIPrep: a robust preprocessing pipeline for functional MRI”. In: *Nature Methods* (2018). DOI: 10.1038/s41592-018-0235-4.

- [89] O. Esteban, R. Blair, C. J. Markiewicz, S. L. Berleant, C. Moodie, F. Ma, A. I. Isik, A. Erramuzpe, J. D. Kent, M. Goncalves, E. DuPre, K. R. Sitek, D. E. P. Gomez, D. J. Lurie, Z. Ye, R. A. Poldrack, and K. J. Gorgolewski. “fMRIPrep”. In: *Software* (2018). DOI: 10.5281/zenodo.852659.
- [90] K. Gorgolewski, C. D. Burns, C. Madison, D. Clark, Y. O. Halchenko, M. L. Waskom, and S. Ghosh. “Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python”. In: *Frontiers in Neuroinformatics* 5 (2011), p. 13. DOI: 10.3389/fninf.2011.00013.
- [91] K. J. Gorgolewski et al. “Nipype”. In: *Software* (2018). DOI: 10.5281/zenodo.596855.
- [92] J. L. R. Andersson, S. Skare, and J. Ashburner. “How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging”. In: *NeuroImage* 20.2 (2003), pp. 870–888. ISSN: 1053-8119. DOI: 10.1016/S1053-8119(03)00336-7. URL: <https://www.sciencedirect.com/science/article/pii/S1053811903003367>.
- [93] M. Jenkinson, P. Bannister, M. Brady, and S. Smith. “Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images”. In: *NeuroImage* 17.2 (2002), pp. 825–841. ISSN: 1053-8119. DOI: 10.1006/nimg.2002.1132. URL: <http://www.sciencedirect.com/science/article/pii/S1053811902911328>.
- [94] R. W. Cox and J. S. Hyde. “Software tools for analysis and visualization of fMRI data”. In: *NMR in Biomedicine* 10.4-5 (1997), pp. 171–178. DOI: 10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L.
- [95] D. N. Greve and B. Fischl. “Accurate and robust brain image alignment using boundary-based registration”. In: *NeuroImage* 48.1 (2009), pp. 63–72. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2009.06.060.
- [96] J. D. Power, A. Mitra, T. O. Laumann, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen. “Methods to detect, characterize, and remove motion artifact in resting state fMRI”. In: *NeuroImage* 84.Supplement C (2014), pp. 320–341. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2013.08.048. URL: <http://www.sciencedirect.com/science/article/pii/S1053811913009117>.
- [97] Y. Behzadi, K. Restom, J. Liao, and T. T. Liu. “A component based noise correction method (CompCor) for BOLD and perfusion based fMRI”. In: *NeuroImage* 37.1 (2007), pp. 90–101. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2007.04.042. URL: <http://www.sciencedirect.com/science/article/pii/S1053811907003837>.
- [98] T. D. Satterthwaite, M. A. Elliott, R. T. Gerraty, K. Ruparel, J. Loughhead, M. E. Calkins, S. B. Eickhoff, H. Hakonarson, R. C. Gur, R. E. Gur, and D. H. Wolf. “An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data”. In: *NeuroImage* 64.1 (2013), pp. 240–256. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2012.08.052. URL: <http://linkinghub.elsevier.com/retrieve/pii/S1053811912008609>.

- [99] R. Patriat, R. C. Reynolds, and Rasmus M. Birn. “An improved model of motion-related signal changes in fMRI”. In: *NeuroImage* 144, Part A (Jan. 2017), pp. 74–82. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2016.08.051. URL: <http://www.sciencedirect.com/science/article/pii/S1053811916304360> (visited on 01/18/2017).
- [100] C. Lanczos. “Evaluation of Noisy Data”. In: *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis* 1.1 (1964), pp. 76–85. ISSN: 0887-459X. DOI: 10.1137/0701007. URL: <http://epubs.siam.org/doi/10.1137/0701007>.
- [101] N. C. Benson, K. W. Jamison, M. J. Arcaro, A. T. Vu, M. F. Glasser, T. S. Coalson, D. C. Van Essen, E. Yacoub, K. Ugurbil, J. Winawer, and K. Kay. “The Human Connectome Project 7 Tesla retinotopy dataset: Description and population receptive field analysis”. In: *Journal of Vision* 18.13 (Dec. 2018), p. 23. ISSN: 1534-7362. DOI: 10.1167/18.13.23. URL: <http://dx.doi.org/10.1167/18.13.23>.
- [102] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee. “N4ITK: Improved N3 Bias Correction”. In: *IEEE Transactions on Medical Imaging* 29.6 (2010), pp. 1310–1320. ISSN: 0278-0062. DOI: 10.1109/TMI.2010.2046908.
- [103] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee. “Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain”. In: *Medical Image Analysis* 12.1 (2008), pp. 26–41. ISSN: 1361-8415. DOI: 10.1016/j.media.2007.06.004. URL: <http://www.sciencedirect.com/science/article/pii/S1361841507000606>.
- [104] Y. Zhang, M. Brady, and S. Smith. “Segmentation of brain MR images through a hidden Markov random field model and the expectation - maximization algorithm”. In: *IEEE Transactions on Medical Imaging* 20.1 (2001), pp. 45–57. ISSN: 0278-0062. DOI: 10.1109/42.906424.
- [105] A. M. Dale, B. Fischl, and M. I. Sereno. “Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction”. In: *NeuroImage* 9.2 (1999), pp. 179–194. ISSN: 1053-8119. DOI: 10.1006/nimg.1998.0395. URL: <http://www.sciencedirect.com/science/article/pii/S1053811998903950>.
- [106] A. Klein, S. S. Ghosh, F. S. Bao, J. Giard, Y. Häme, E. Stavsky, N. Lee, B. Rossa, M. Reuter, E. C. Neto, and A. Keshavan. “Mindboggling morphometry of human brains”. In: *PLOS Computational Biology* 13.2 (2017), e1005350. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1005350. URL: <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005350>.
- [107] V. S. Fonov, A. C. Evans, R. C. McKinstry, C. R. Almlil, and D. L. Collins. “Unbiased nonlinear average age-appropriate brain templates from birth to adulthood”. In: *NeuroImage* 47, Supplement 1 (2009), S102. DOI: 10.1016/S1053-8119(09)70884-5.
- [108] K. Hoskin and Vos. “The ‘awful idea of accountability’: inscribing people into the measurement of objects”. In: 2021. URL: <https://api.semanticscholar.org/CorpusID:259255720>.

- [109] X. Chen, C.-J. Hsieh, and B. Gong. *When Vision Transformers Outperform ResNets without Pre-training or Strong Data Augmentations*. 2021. DOI: 10.48550/ARXIV.2106.01548. URL: <https://arxiv.org/abs/2106.01548>.