Deep Learning for Computational Neuroimaging

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Leonie Henschel
aus
Göttingen

Bonn, Dezember 2023

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter:	Prof. Dr. Martin Reuter
2. Gutachter:	Prof. Dr. Reinhard Klein
Tag der Promotion:	23.08.2024
Erscheinungsjahr:	2024

Abstract

Non-invasive neuroimaging studies aim to increase our understanding of the brain in health and disease. The extraction and quantification of morphometric measures such as volume or cortical thickness from the incoming stream of data requires efficient, reliable, and accurate computational workflows. However, existing neuroimaging pipelines involve computationally intensive optimization steps and, thus, do not scale well to large cohort studies. Efficient deep learning networks have the potential to revolutionize image analysis but are so far limited in their applicability due to the primary focus on 1.0 mm voxel-based segmentation and insufficient validations. This thesis addresses these issues and contributes fast, accurate, and extensively validated open-source deep learning solutions for the automated processing of structural human brain magnetic resonance images.

With FastSurfer, a full deep learning based alternative to well-established, traditional neuroimaging pipelines is introduced. The methodological contributions include FastSurferCNN, a 2.5D convolutional neural network (CNN) for whole brain segmentation into 95 classes in under 1 minute, and RECON-SURF, a surface reconstruction stream including a novel spectral spherical embedding, fast mapping of the volume segmentations to the surfaces, and extraction of pointwise and regional thickness estimates. FastSurfer reduces processing times while outperforming traditional tools in terms of accuracy, reliability, and sensitivity. The thesis also addresses cortical surface segmentation with a novel view-aggregating polar parameterization network called p^3CNN. The view aggregation across different pole axis orientations alleviates distortions introduced by the non-isometric mapping and is shown to outperform spherical CNNs. With the voxel-size independent neural network (VINN) the thesis further introduces a tool for native image segmentation across a range of different voxel sizes. A novel network-integrated resolution-normalization layer uses a priori knowledge about the image resolution to internally transition between scales. The internal interpolation retains important image information and outperforms traditional scaling augmentation. VINN consistently achieves good results on both, low- and high-resolution images and is highly effective in combating methodological biases in multi-resolution settings. Extending the VINN concept by shifting not only scaling but also rigid transformations into the network, finally gives rise to the VINN with internal augmentations (VINNA). At the first layer scale transition, the multi-dimensional feature maps are flexibly rescaled and randomly transformed to diversify the training distribution. The shift from external to internal augmentations translates to improved accuracy measures in the context of newborn brain segmentation. Overall, the thesis contributions enable reliable, scalable big-data analysis with high robustness for a variety of datasets.

Acknowledgements

It is a pleasure to thank all those who made this thesis possible.

I would like to thank Prof. Martin Reuter for giving me the opportunity to join his group and write my Ph.D. thesis at the German Center for Neurodegenerative Diseases. I am also grateful for his valuable support and encouragement in the realization of this work.

I am very grateful to Prof. Reinhard Klein from the University of Bonn for accepting my thesis supervision and for his time spent reviewing my work.

A special thanks goes to all members of the Reuter group for their assistance and for creating such a pleasant working atmosphere.

Furthermore, I would like to thank Prof. Lilla Zöllei for giving me the opportunity to work at the Massachusetts General Hospital as a visiting research scholar in 2022.

Finally, I would like to express my gratitude to my family and my closest friend Niklas Domdei for their support in several ways throughout my studies and for encouraging me in the pursuit of my goals.

Contents

CHAPTER 1

Introduction

The brain is one of the most integral and complex parts of the human body. The conglomerate of 100 billion neurons and their interplay gives us consciousness, makes up our thoughts and emotions, and defines how we perceive the world around us. Just like us, it changes over the years and shows significant adaptability to both, intrinsic and extrinsic influences. The human brain will grow three times its size in the first year of life, continues to grow until about middle age, and starts shrinking as we get older [1]. Analogously, our learning ability and cognitive skills expand and slowly decline in the natural process of aging. Genetic as well as environmental factors can drastically influence this process and can cause developmental, neurodegenerative, or psychiatric disorders. Discrimination of potentially detrimental versus benign phenotypes or identification of early intervention points to reduce disease burden is increasingly important, specifically considering the rise in human life expectancy. Over the last decade, neuroimaging studies have used non-invasive imaging technologies, first and foremost Magnetic Resonance (MR) Imaging, to improve our understanding of brain organization in health and disease. The current, drastic increase in imaging cohort sizes inevitably requires fast computational methods to analyze and extract the treasure of information hidden within them. The extensive run times of traditional neuroimaging pipelines, unfortunately, limit their applicability here and aggravate the knowledge discovery process. Given their recent success in several image analysis tasks, Convolutional Neural Networks (CNNs) offers the potential to address this challenge by drastically improving runtime as well as performance. However, existing brain segmentation architectures are limited to volume-based processing in single-resolution set-ups and lack utilities such as the creation of cortical surface models and estimation of thickness or region of interest (ROI) measures.

Therefore, this Ph.D. thesis focuses on developing fast, scalable, reliable, and sensitive algorithms for the analysis of the human brain Magnetic Resonance Images (MRI). This work introduces the first deep learning based neuroimaging pipeline for segmentation and subsequent surface reconstruction called FastSurfer [2]. It is extensively validated, with an assessment of segmentation accuracy, generalizability, reliability, and sensitivity to disease effects. Unique architectural solutions for multi-resolution [3], head position variations [4], and spherical deep learning [5] extend the work with methodological innovations. Assessment of segmentation bias in multi- and single-resolution networks further highlights the need for resolution-independent segmentation tools [6]. The necessary background to evaluate these contributions is summarized in the following introduction sections. Due to the interdisciplinary nature of the thesis, both neuroimaging and deep learning concepts are

introduced.

1.1 Neuroscience

1.1.1 The Brain

The human brain is a soft organ consisting of a complex assembly of neurons and glial cells situated within the cranial cavity [7, 8]. The spatial organization is encoded in the genome and generally consistent across populations. The median longitudinal fissure divides the brain into two roughly equal parts called the left and right hemispheres [9, 10]. Each hemisphere consists of a 1–4.5 mm thick cortical gray matter (GM) sheet [11] called the cortex, the white matter (WM) as a coherent structure below it, and a collection of subcortical structures within or adjacent to the WM [8–10]. The corpus callosum, a bundle of WM fibers, connects the two hemispheres. The brain is immersed in cerebrospinal fluid (CSF) filling the area between the brain and skull.

The WM is composed of bundles of nerve fibers or axons essential for communication between different brain regions. These elongated tubes connect the neural cell body or soma and the dendrites within a neuron [12]. The axon is wrapped in myelin sheaths, a mixture of protein and lipids acting as a natural insulator and allowing high-speed conductivity of electrical signals [13]. The high-fat content makes the WM appear white and ultimately gives it its name. The GM, on the other hand, predominantly contains the soma, dendrites, and synapses which contain less fat and therefore appear grayish. The subcortical structures encompass areas such as the brainstem, thalamus, hippocampus, amygdala, putamen, or cerebellum, and the ventricular system [8]. The brain ventricular system consists of four connected, CSF-filled cavities called ventricles.

The cortex has a wrinkled, convoluted appearance defined by a series of sequential gyri (outward-facing bulges) and sulci or fissures (inward-facing grooves). The characteristic appearance stems from the rapid expansion of the cortex during embryonic development [1] and maximizes surface area in the restricted space. The depth, shape, and relative location of gyri and sulci serve as landmarks and allow separation of the cortex into distinct regions [9, 10, 14–16]. While the macroanatomy is generally consistent, substantial differences in size and shape of the sulci or gyri exist between individuals [17, 18] and even identical twins [19]. This flexibility can aggravate analysis efforts and needs to be considered when designing algorithms for brain analysis.

The brain also changes significantly in size and morphology throughout the lifespan (see Figure 1.1). Neuronal production starts as early as day 42 post-conception and continues until midgestation [1]. The myelination of neurons starts in the pons and cerebellar peduncles at birth and continues till 8-12 months where the WM and frontal, parietal, and occipital lobe are finally changing [23–25]. The morphology within the infant cohort alone (e.g. prenatal, neonates, six months old, or one year old) is therefore exceptionally diverse. The brain further increases in size by around four-fold up till the preschool period and changes in GM and WM structures continue up until adolescence [1]. This is also reflected in a change in cortical thickness, i.e. the distance from the WM-cortex to the cortex-CSF border. After reaching a maximum between 3-10 years of age, the cortical thickness slowly decreases with age due to loss of GM [26–28]. Various neurological disorders further show different patterns of progression and regionality of cortical thinning [29–50]. Thickness analysis can therefore provide valuable insights into the evolution and causative factors of a disease [11]. Neurodegenerative disorders such as Alzheimer's Disease (AD) for example, are often characterized by advanced cortical atrophy [29–31, 40, 44–50]. Further, subcortical structures like the hippocampus,

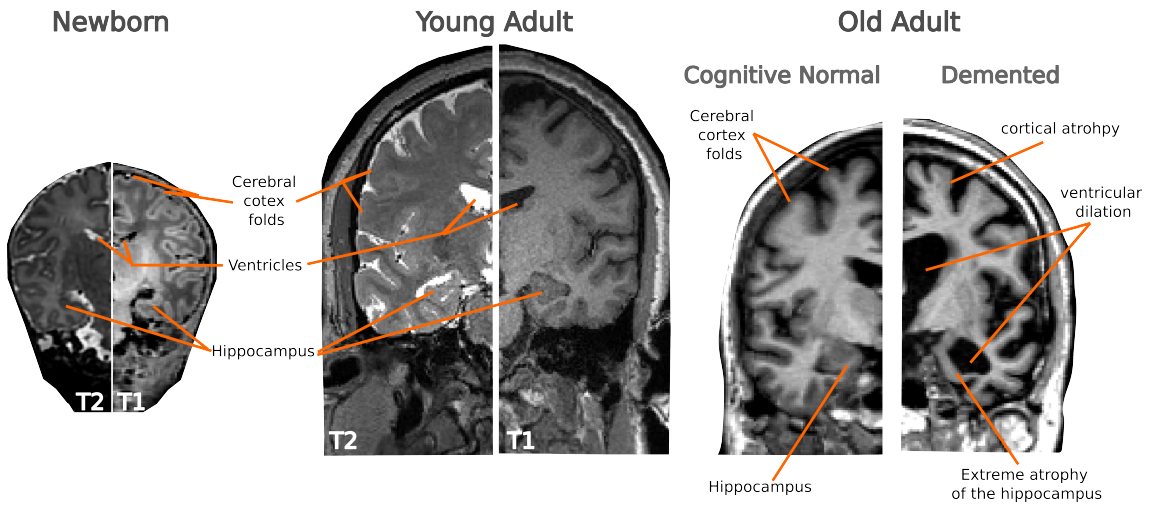

Figure 1.1: Comparison of brain structures across age and disease groups. Morphologies within the newborn brains (left) are smaller and contrast in images differs strongly from young adults (middle). Gray matter decreases while ventricular structures increase in older adults (right) with significant differences in the severity between cognitive normal and demented cases. Screenshots are from representative scans of the Developing Human Connectome Project (dHCP) [20] (newborn), Human Connectome Project (HCP) [21] (young adult), and OASIS1 [22] (old adults).

thalamus, or amygdala show significant volume reductions while the ventricles increase in size [46, 51–55]. Comparison of brain morphology across age-matched groups can therefore allow detection of neurodegeneration at an early stage [31, 56–58].

1.1.2 Neuroimaging Data Collections

To date, several population and cohort studies set out to assemble large repositories of neuroimaging data to identify sources of variability and distinguish normal from pathological developments. The data collection for younger populations, such as newborns or infants, is currently much sparser than for adults, but efforts to provide a comprehensive set of images across this stage of life are underway [20, 59, 60]. To date, the Developing Human Connectome Project (dHCP) provides the largest, easily accessible collection with 1173 fetal and/or neonatal participants in total [20].

As imaging has initially been optimized for adults, data availability is considerably larger and ever-increasing at a fast pace. While older data collections were also restricted to smaller cohorts, studies nowadays envision repetitive imaging of tens of thousands of participants. For example, the Alzheimer's Disease Neuroimaging Initiative (ADNI), a consortium dedicated to understanding changes in AD progression, started the first data collection phase including 821 participants with and without mild cognitive impairment (MCI) or AD in 2004 [61]. With the latest renewal, the collection has more than doubled in size with over 2000 enrolled participants [62] and is envisioned to screen up to 20,000 additional participants in the latest extension [63]. Similarly, the Open Access Series of Imaging Studies (OASIS) [22, 64–66] provides longitudinal and cross-sectional neuroimaging data for young, middle-aged, non-demented and demented older adults, and grew from a starting size of 416 in its first iteration [22] to over 2608 enrolled participants and 4325 imaging sessions in total [64–66]. Other atypical brain developments are reflected in neuroimaging collections as well, such

as psychiatric disorders [67–71] or autism [72–74]. Another large-scale data collection stems from prospective cohort studies focused on recording representative cross-sections of the general population to elucidate the effects and interactions of lifestyle, environmental, and genomic factors on disease development. The Rhineland Study (RS) [75], for example, recently started collecting biological data, including brain scans, for 30,000 adults from Bonn and the surrounding area. Over 30 years, each participant will be reexamined every three years, leading to a myriad of imaging data points. Similarly, the United Kingdom Biobank acquired 100,000 brain images over five years [76] and the Human Connectome Project (HCP) is in the process of gathering data from healthy humans, ranging from birth [20], to children [77], young [21], and older adults [78].

1.1.3 Magnetic Resonance Imaging

A common factor in all of the aforementioned studies is the recording of structural MR images to create a visual, computer-analyzable image of the brain at the macroscopic scale. MR Imaging is a method introduced as a clinical device at the beginning of the 1980s [79]. It allows the non-invasive visualization of soft body parts *in vivo* without ionizing radiation such as X-rays. Due to its ideal contrast for soft tissue, it is the modality of choice for structural brain imaging applications [80–82].

The principle of MR imaging is based on the protons within the nuclei of hydrogen atoms [80–82]. Their spin and positive electrical charge produce an electrical current and subsequently a magnetic field. Adding an external magnetic field results in the alignment in parallel or anti-parallel direction (Figure 1.2, A). The parallel orientation is the energetically preferred state and results in a net longitudinal magnetization (LM) along the direction of the external field (z-axis). The protons move around the alignment axis with the precession frequency ω_0, which is dependent on the magnetic field strength B_0 and can be calculated with the Larmor equation:

$$\omega_0 = \gamma B_0 \tag{1.1}$$

A stronger external field results in a greater precession frequency, where the exact amount is dependent on the tissue-dependent gyro-magnetic ratio γ [82]. When a radio frequency pulse at the precession frequency of the protons is added into the system, the protons synchronize and switch to the anti-parallel, high-energy position. Thus, the magnetization along the z-axis is reduced while simultaneously creating a transversal magnetization (TM) in the x-y plane. The movement of the TM field induces an electric current which can be measured from the outside.

Application of an external magnetic field gradient along the z-axis (GS) creates a slice-selective precession frequency and in turn allows targeted excitation with a matching radio frequency range [80] (Figure 1.2, B, left part). The resonance frequency, which describes the precession after synchronization with the radio frequency pulse, is also a function of the local magnetic field strength. Therefore, the same principle as before applies: adding a local gradient along the x-axis (GR) during the readout phase spreads out the recorded frequencies, and therefore encodes spatial information [80] (Figure 1.2, B, middle part). The information about the remaining y-axis is encoded via the phase of the frequency pulses (Figure 1.2, B, right part). The phrase describes the angle between the wave signal at an arbitrary origin t_0 (e.g. the beginning of a cycle) and time point t with $\phi(t) = 2\pi(t - t_0) \times frequency f$. Applying a gradient along the y-axis briefly alters the precession frequency. Once turned off, the segments return to the original frequency determined by GR, but with a phase-shift, i.e. the positions along the wave differ at time point t [81, 82]. The overlapping waves create a final read-out with

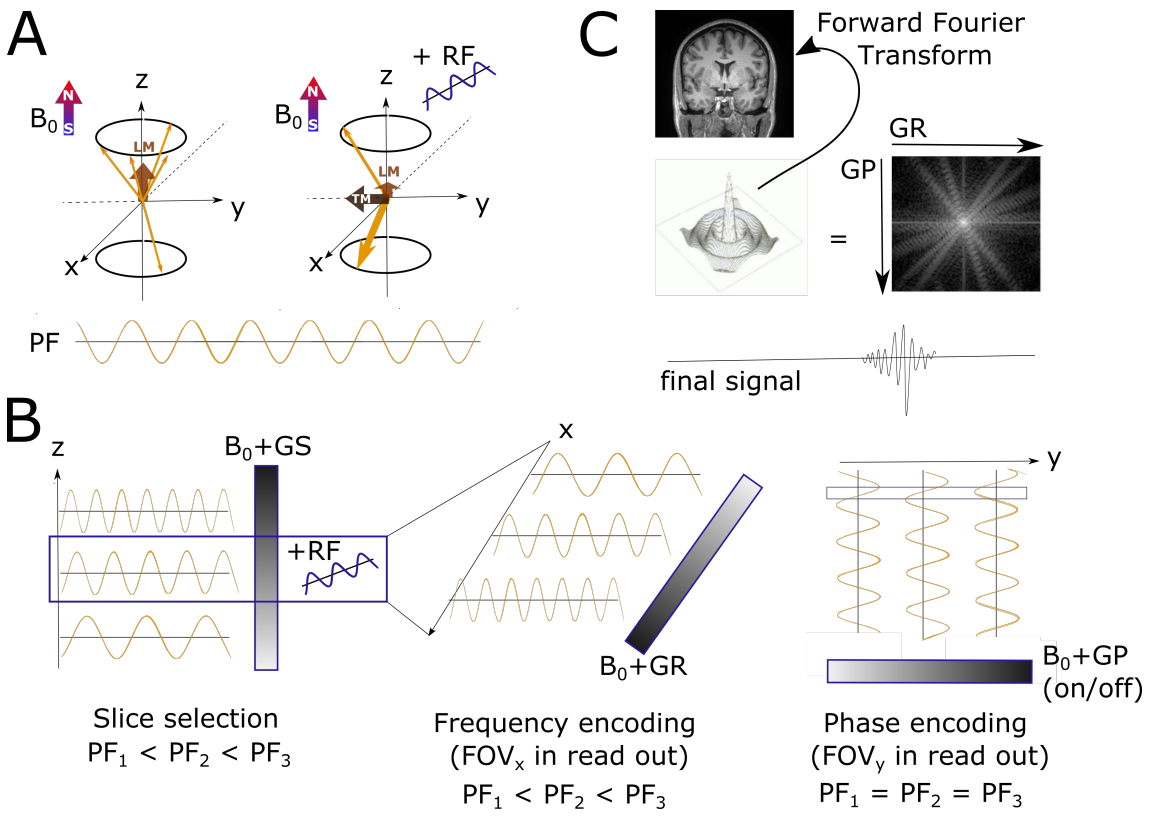

Figure 1.2: The principle of Magnetic Resonance Imaging. A: The protons (orange arrows) orient themselves in an external magnetic field B_0 and rotate with the precession frequency (PF) around the alignment axis. As more protons align themselves in the favorable position in the direction of the external field, a net longitudinal magnetization (LM) along the z-axis occurs. Excitation with a radio frequency (RF) pulse at PF, creates a net transversal magnetization (TM) along the z-axis and reduces the LM through synchronization and switch of protons to the anti-parallel position. B: Application of magnetic field gradients along the z-axis (GS) differentiates PFs along the z-axis and allows the selection of slices with a targeted RF. Subsequently, spatial resolution along the x-axis is encoded via another magnetic field gradient (GR) during the read-out phase. Application of a gradient along the y-axis (GP) shortly alters the PF and once turned off results in a phase-shift. C: The overlapping signals represent a rapidly decaying wave and are translated into image space using the Fast Fourier transform.

varying amplitude dependent on their phase shift. The final signal represents a rapidly decaying wave (Figure 1.2, C). Applying a fast Fourier transform translates the signal into an image readout [80, 82]. The overall amplitude is dissected into contributions by the different frequencies, translated to voxel intensities, and assigned to specific locations within the image based on the known magnetic field gradients. Here, the divisions within GR and GP represent the x- and y-ticks and should be equal to the chosen matrix resolution [80].

Imaging Modalities

Different tissue types contain different amounts of macro- and micro-molecules which affect the emitted signal by changing the longitudinal relaxation time T1 and the transversal relaxation time T2

[80, 81]. T1 describes how long it takes the tissue to return to its low energy equilibrium state after the radio frequency pulse and recover 63% of the original z-axis magnetization. It is dependent on how often and how strongly the surrounding molecules or lattice absorb energy upon impact. T2 describes how fast the TM decreases to 37% of its original value. It depends on the magnetic properties of the surrounding nuclei which cause protons to change their precession frequency and push them out of phase. Liquids or water have a long T1 and T2 as the molecules are small and in rapid motion. Fat has a short T1 and T2 due to its large, rigid size and low innate energy. Weighting the constants against each other, results in images with different tissue contrasts. The two most commonly used sequences create T1-weighted (T1w) and T2-weighted (T2w) images (also referred to as T1 and T2 images).

T1w sequences reduce the time of repetition, i.e. the length between the radio frequency pulses, such that the signal between tissues with short and long T1 is maximized [80–82]. High water appears dark as the T1 is long and recovery of the original z-axis magnetization is slow. Fast repetition of the radio frequency pulse ensures the continuation of the low-intensity state. Fatty structures, on the other hand, appear bright due to the short T1 and faster recovery of the LM.

The T2w sequences show the opposite effect, with water exhibiting the highest signal intensity and fat the lowest. The contrast change is created by the time to echo which defines the time between the application of a radio frequency pulse and the peak of the echo [81]. A longer time to echo delays the rephasing of the protons and reset of the TM [82]. Water with its long T2 reduces the x-y magnetization slower than fat, resulting in less signal reduction over time, and a stronger signal for a long time to echo [80–82].

The sequence selection is determined by the structure or cohort of interest. From childhood onward population studies record T1w MRIs due to the solid contrast between the highly myelinated WM and low-fat GM [21, 22, 61, 64, 67, 72, 83]. Consequently, most analysis pipelines are also optimized for T1w images. During early developmental stages, the myelination is incomplete. Younger population cohorts therefore predominantly record T2w images which create a better contrast until around 6 months of age [20, 59, 60]. Due to their small head sizes, infants are further often recorded at higher imaging resolutions than adults (up to 0.5 mm).

Imaging Resolution

The resolution describes the size of the individual voxels in the 3D MRI [80]. A smaller voxel size equals a higher resolution and generally allows a more fine-grained visualization of structures. The voxel size is defined via the scanning parameters, specifically the field of view (FOV), the acquisition matrix M, and the slice thickness ST: $voxelsize = (FOV/M) \times ST$ [80]. The FOV defines the area to be scanned while the acquisition matrix defines the number of voxels within this area (see Figure 1.3). If the voxels have identical depth, height, and width, they are isotropic.

To date, most anatomical data resides at 1 mm spatial resolution which is considered a good compromise between acquisition time and image quality. As the resolution increases, the signal per voxel is reduced leading to a nonrecoverable decrease in the image quality described by the signal-to-noise ratio (SNR) [80, 84–86]. The smaller voxels contain less protons, therefore reducing the readout per voxel. For a high-quality image, the signal from the structures of interest must be higher than the noise. As the SNR scales with the voxel size, doubling the resolution of an isotropic volume translates to a 2^3 reduction in SNR [80, 86]. Longer scanning times, repeated scans, or higher field strength can alleviate this problem [80].

Over the years, developments in scanner design have substantially improved image quality of

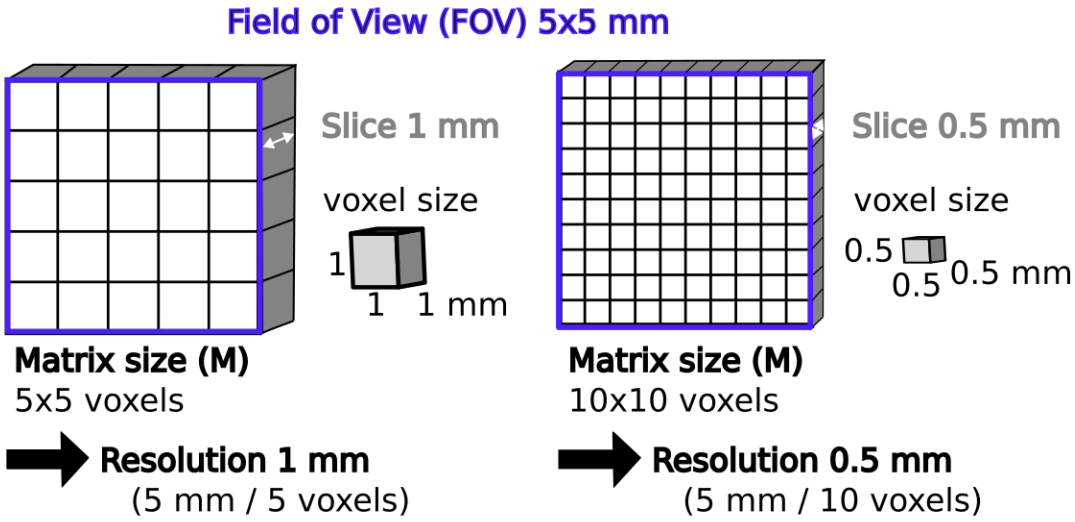

Figure 1.3: Image resolution is defined by the field of view FOV, acquisition matrix M, and slice thickness. The left low-resolution, 1.0 mm slice is defined by a FOV of 5×5 mm and M of 5×5. Using the same FOV, but a larger M (10×10) acquires more, smaller voxels and returns a high-resolution slice at 0.5 mm (right). Both slices have isotropic voxels with identical lengths in all three directions (x, y, z).

high-resolution MRI [87], partly bypassing the SNR problem by offering increased baseline signals via stronger magnets. The subsequent larger magnetic field strength nowadays allows recording at sub-millimeter resolutions for better detection, sampling, and measurements of fine-grained anatomical details [88, 89]. Better visualization of GM, WM, and CSF facilitates accurate delineation of the gray-white and gray-CSF tissue borders and improves cortical thickness estimates (see Figure 1.4)[88, 89]. The driving factor behind these improvements is a reduction in the so-called partial volume effect (PVE) [88–90]. Transition regions between multiple tissue classes can not be separated by a complete voxel, due to the resolution limit. Therefore, a mix of tissue types is present in different proportions within the voxel and contributes to its intensity. Assignment of these voxels to a single structure is difficult and often causes misclassifications. Higher resolutions alleviate this problem and hence improve the estimation of volume-based measurements. More and more imaging studies are therefore shifting from the de-facto 1.0 mm standard resolution to sub-millimeter acquisitions [75, 89].

1.1.4 Computational Neuroimaging

While MRI acquisition is a pivotal step, the actual morphometric information of interest, such as volume, density, thickness, or curvature, is still buried within the three-dimensional volume. W global intensity, morphometric features can not be extracted directly from the original MRI. Generating a segmentation, i.e. a division of the brain into anatomically meaningful regions, is essential to provide localized ROI measurements. Comparison of these measurements across different populations or ages allows the identification of subtle, localized alterations and subsequently underlying cause-effect relationships. Additionally, MRI segmentation is essential for structure visualization, delineation of pathological regions (e.g. lesions or tumors), and subsequently image-guided interventions and surgical planning.

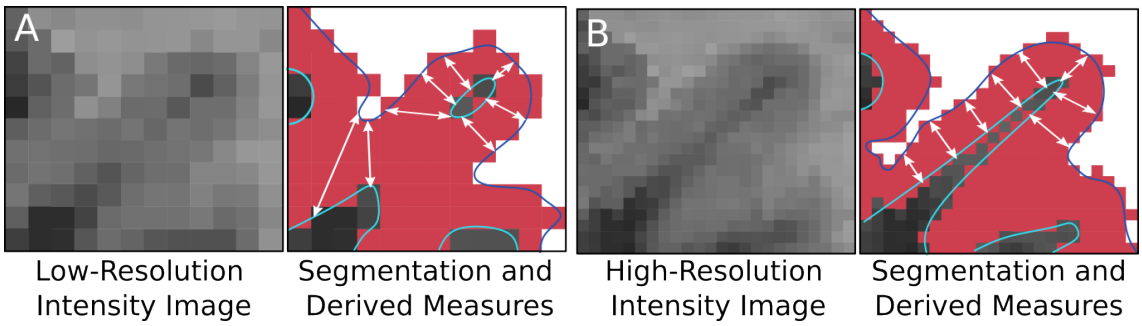

Figure 1.4: Image resolution affects the detail of discrete segmentation label maps and derived measures such as surface models and thickness. **A.** The low-resolution image is less detailed and causes partial volume effects (PVEs) by accumulating signals across tissue boundaries into larger voxels, whereas **B.** The high-resolution images and derived segmentations allow more precise region delineation and capture details, e.g. for improved shape or thickness analysis (white arrows). Figure adapted from [3].

Traditionally, human operators have created segmentations based on a priori defined landmarks. While still considered the gold standard today, manual labeling of brain MRIs is a complex, time-consuming task aggravated by the ambiguity of tissue boundaries, great anatomical variations between individuals, and the complex folding patterns of the cortex [91, 92]. Labeling a single 1 mm^3 brain scan from scratch takes a trained human operator 2-3 days [14] and region assignments can vary significantly between different raters [93]. The time requirement further increases substantially for higher resolution labels due to the cubic voxel increase. Hence, automated computational tools are pivotal to aiding or bypassing human operators, reducing processing times, and allowing standardization of the image analysis process.

1.1.5 FreeSurfer

One of the most extensively validated, commonly used tools in the neuroimaging community for automatic processing of whole brain MRIs is FreeSurfer [94]. While the challenges human operators face when segmenting brain MRI remain, FreeSurfer has developed methodologies to address them, first and foremost via the introduction of surface models. Surface models are meshes representing tissue borders in the brain, specifically the inner (GM-WM interface) and outer (GM-CSF interface) boundary. These surface meshes are generated from anatomical MRIs through a process known as cortical surface reconstruction. In FreeSurfer, the reconstruction can be divided into two parts, the volume and surface stream. In short, inhomogeneity removal creates a normalized intensity image subsequently segmented into 32 anatomical structures [95, 96]. Following triangular tessellation, deformation, and correction of topological errors yields an accurate and smooth representation of the surfaces [95]. Within the next two sections, both processes will be explained in more detail.

Volume Stream

Initially, the input MRI is conformed to an isotropic voxel size of 1 mm and a uniform size of $256 \times 256 \times 256$. The image intensities are further robustly rescaled to a range of 0 to 255. Afterward, intensity variations caused by the magnetic susceptibility artifacts and resonance frequency-field inhomogeneities of MR scanners are corrected to yield a uniform, normalized image [95]. This

process is commonly referred to as bias-field correction. After removing the skull from the 3D MRI volume, the image is segmented into WM and background in a two-step procedure. First, each voxel is classified to WM if the intensity is within a range of 90-140. Second, if more than 20% of the voxel nearest neighbors differ from its assigned label, it is considered for relabeling based on local geometric information [95]. The WM segmentation is further divided by two cutting planes. The sagittal plane extends from top to bottom along the corpus callosum and divides the left and right hemispheres. The coronal plane cuts off the pons and removes subcortical structures such as the brain stem. After removing any remaining holes, the resulting WM segmentation serves as a baseline for the triangular tesselation. In FreeSurfer, surface-based analysis is, however, restricted to the cortex, and volumetric techniques are used to create labels for the subcortical and ventricular structures, such as the hippocampus, amygdala, or lateral ventricles [96]. To this end, the image is aligned to a probabilistic atlas, which contains 32 structure priors for each voxel in the anatomically meaningful Talairach space [97, 98]. Considering these priors in addition to information about spatial arrangements of anatomical labels calculated with an anisotropic spatially nonstationary Markov Random Field on a training set yields a segmentation comparable to manual labels in terms of accuracy [96]. Within FreeSurfer, the resulting label map is called anatomical segmentation (*aseg*).

Surface Stream

The surface generation starts with the tessellation of the generated WM segmentation one hemisphere at a time [96]. Initially, the border voxels' square faces are modeled by two triangles. The resulting tessellation is a jagged, block-like representation of the connected WM voxels. While explicitly ensuring connectivity and avoiding self-intersections, the sphere is re-positioned by minimizing an energy function consisting of a two-factor spring property and an intensity-based term. The tangential term of the spring property imposes a smoothness constraint by penalizing each node for being too far from its neighbors. The normal term redistributes vertices based on local expansion requirements and encourages uniform spacing of the vertices. The intensity term finally promotes the positioning of the vertices such that their intensity value is close to the average white matter value of its neighboring border voxels. Movement along the gradient of the energy function results in a smooth, reconstructed WM surface called *white* in FreeSurfer. The outer surface representing the GM-CSF interface is constructed by deforming the white surface. The minimized energy function remains the same except for the intensity term. Instead of considering a local neighborhood, a predefined global target is used as the contrast between CSF and GM is sufficiently large. The resulting surface is called *pial* in FreeSurfer. As the surface generation is based on automatic segmentation, topological errors such as handles or small holes can not be avoided and need to be fixed. Detection of these errors is achieved by using the fundamental notion of topological equivalence as stated in Fischl et al. [94]: "If the original surface is not topologically equivalent to a sphere, then it is not possible to find a continuous homeomorphic mapping". Mapping the reconstructed surfaces to a sphere identifies non-invertible areas and allows efficient topology correction by discarding the tessellation and generating a new corrected mesh version in those areas only [99]. Considering geometric features, the defects are corrected via filling of holes (aka discontinuities in planar sheets) or cutting of existing handles (aka bridges between nonadjacent points in the cortex).

Besides the detection of topological defects, the spherical inflation can be used for alignment of the individuals' surface to a canonical reference atlas based on the cortical folding or curvature patterns (i.e. sulci and gyri) [100, 101]. In contrast to volume-based registration, the spherical approach stays

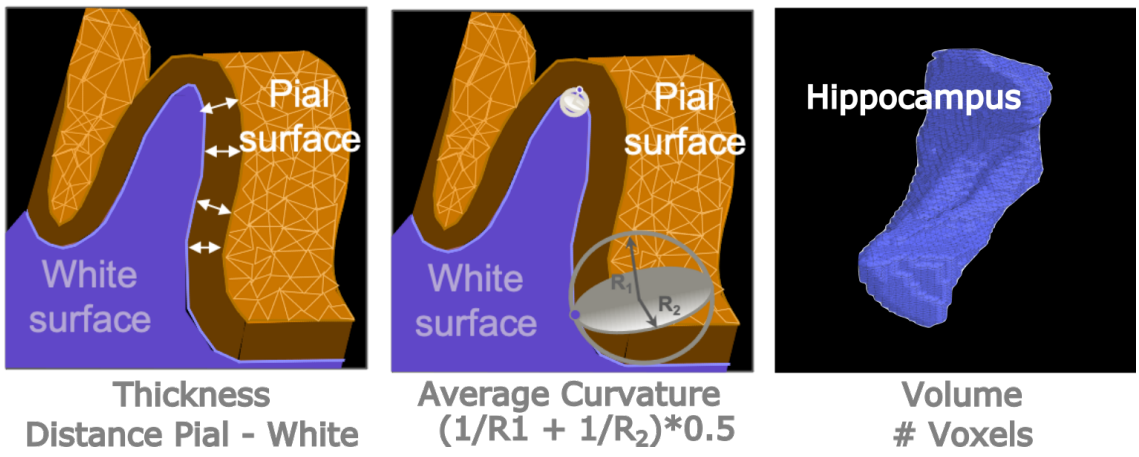

Figure 1.5: Surface-based and volume-based brain morphology measures. The cortical thickness (left) represents the distance between the pial and the white surface. Fitting a sphere to the white surface allows the calculation of curvature (middle) by averaging the inverse of the radius R1 and R2. The summation of the voxels assigned to a certain structure (e.g. the Hippocampus) estimates the volume (right).

true to the underlying geometry of the cortical surface while providing a well-defined coordinate system. In FreeSurfer, the reference atlas is called *FSAVERAGE*. Several manual segmentations have been defined in this space and can be propagated to the individual subject. The most commonly used atlas definitions are the Desikan-Tourville atlas [15] separating the cortex into 35 regions, the Desikan-Tourville-Klein atlas (DKT) [14] creating a segmentation into 31 regions, and the Destrieux atlas [16] providing a detailed subdivision of the cortex into 74 regions. The labeling of individual cortical regions is called *parcellation* in FreeSurfer. A final volume with cortical parcels, as well as subcortical regions, is generated by (i) correcting the initial cortical segmentation in the *aseg* with the surface information and (ii) propagating the labels of the individual cortical parcels from the surface into the volume. Using the Desikan-Killiany-Tourville atlas [14, 15], a final segmentation into 95 classes called *aparc.DKTatlas+aseg* can thus be generated.

At the end, morphometric information is readily available from both, the surfaces such as the thickness, curvature, or folding indices of the cortical structures, and the images such as volume or intensity for subcortical and ventricular structures. Figure 1.5 depicts the measurement of the thickness, average curvature, and volume for the cortex and hippocampus, respectively.

1.1.6 Open Challenges

While FreeSurfer offers many advantages, the extended processing time limits its applicability to large cohorts – specifically for up-and-coming high-resolution studies. As an example, a single FreeSurfer run on a standard 1.0 mm isotropic volume takes around 4-8 hours. Processing all current images of the envisioned 30,000 0.8 mm scans of the RS [75] would hence require more than 14 years without parallelization. This is a common problem for existing computational pipelines as their segmentation methods require extensive numerical optimization (for non-linear registration or segmentation), multiple image pre-processing steps, and careful hand-crafting of discriminative features [94, 102–105]. The considerable runtimes limit scalability and pose problems for the incoming

data stream. There is an urgent need in the neuroimaging community for efficient, accurate, and well-characterized processing pipelines. One potential solution to this challenge is the use of deep learning which has not only substantially reduced processing times to seconds but also reached human-level performance for image classification and segmentation tasks [106–108].

1.2 Deep Learning

Deep learning is a representation learning method in which compositions of simple, non-linear modules transform raw data step by step into more abstract forms to automatically discover an optimal representation for the task of interest, such as object detection, semantic segmentation, or classification [107]. The pivotal advantage of deep learning compared to conventional machine-learning techniques is the automatic nature of this extraction. The representations are learned from the data itself and do not require a priori hand-crafting of discriminative features.

The underlying technology behind deep learning are so-called Artificial Neural Networks (ANNs), a conglomerate of interconnected nodes or neurons organized in a directed multi-level, layered structure. The terms deep learning and neural networks are therefore often used interchangeably. The idea for ANN was born in 1943 when Warren McCulloch and Walter Pitts used propositional logic to model how a biological network of neurons work together to perform complex computations [109]. Within their model, each base unit called an artificial neuron has one or more binary inputs and one binary output. Here, a value of one indicates activation or an "on" state, while a value of zero represents inhibition or an "off" state. If the aggregated input values of these neurons are larger than an a priori-defined threshold parameter, the entire base unit is on and propagates its output.

By stacking multiple of these artificial units, any logical proposition can be computed in theory. Units not in the input or output layer are referred to as hidden units [107]. In 1958, the concept of the artificial neuron was generalized by Frank Rosenblatt with learnable weights for each input connection and real numbered in- and outputs [110]. The weighted sum of the inputs is passed to a step or activation function which outputs the positive class when the result exceeds a threshold (see Figure 1.6, top left). The novelty in the multi-layer perceptron lies within the ability to adapt the weights for each input connection based on Hebb's rule [111]. Within the perceptron, a connection weight between two units is increased when it enables a correct prediction. The perceptron therefore represents the first trainable neural network.

Importantly, deep learning networks are characterized by several (more than three) hidden layers that process data in a coordinated and adaptive fashion. Training of these deeper networks became possible by the introduction of the simple yet effective back-propagation algorithm [112], graphical processing units (GPUs) allowing massive parallelization of matrix and vector calculations, and non-linear activation functions such as the rectified linear unit (ReLU) [113, 114]. Together, those innovations allow the training of networks with hundreds of hidden layers and a speed-up of factor 10 to 50 [106, 107, 115–118]. As an example, segmentation of a 1.0 mm whole human brain into more than 30 classes is nowadays achievable in less than a minute on the GPU [2, 3, 119, 120].

For training an image segmentation network, pairs consisting of an intensity image x and corresponding ground truth y are needed. The ground truth represents the correct label assignment and is the output the network should ideally produce. Training a neural network comes down to finding a set of parameters $\theta = [W, b]$, that minimize the difference between the ground truth and the prediction made by the network. This parameter optimization task is achieved by gradient descent [107].

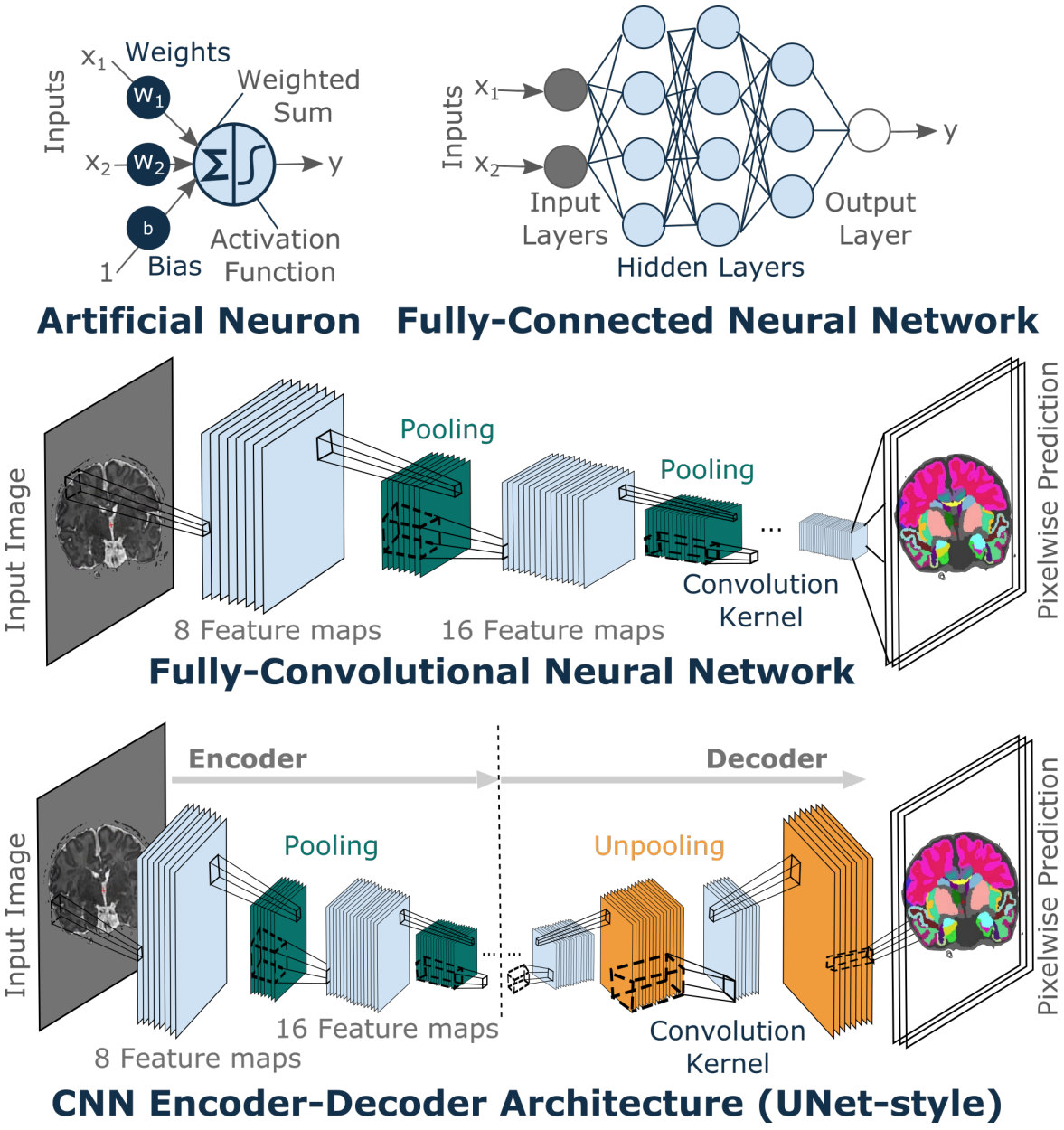

Figure 1.6: Artificial neurons (also called linear threshold units) calculate the output value by passing the weighted sum of the input signals through an activation function (top left). Stacking multiple layers of these units creates fully-connected neural networks (top right). Notice, that all units of a given layer are connected to all units of the following layer. Each unit also has an associated bias term (not shown). A fully convolutional neural network (FCNN, middle) repeats convolution and pooling layers and upsamples the final encoding back to the original input size to create a full pixel-wise prediction. Here, the units are arranged in a regular grid and each is only connected to a local patch or receptive field of the feature maps of the previous layer via convolutional filters (black rectangles extending across the entire depth). Encoder-Decoder architectures replace the single upsampling in FCNN with a subsequent combination of convolution and unpooling layers (bottom). Due to the resemblance to the letter U, the networks are also called U-Nets.

In a fully-connected network (Figure 1.6, top right), the units within the first layer apply a linear transform to the intensity image, represented as a linearized vector x with N elements or observations, by multiplication with a randomly initialized weight matrix W and addition of a bias term b (Equation (1.2)). The number of columns C and rows R in the weight matrix equals the number of observations in the input vector x and the number of neurons in the current layer, respectively. The bias term is equivalent to the intercept in a regression model and allows the network to learn a more complex function. Applying a non-linear activation function f on the weighted sum z_r creates the final output y_r of a given unit r in the first layer.

$$\begin{aligned} z_r &= b_r + \sum_{c=0}^{N} w_{c,r} \times x_c \\ y_r &= f(z_r) \end{aligned} \tag{1.2}$$

This process is repeated in all following layers, slowly transforming the input intensity image to a different representation. Due to the directionality of the signal flow, this type of network is called a feedforward neural network.

At the end of the network, the output prediction $\hat{y}$ is compared to the ground truth y using a loss function such as the sum of squared errors or cross-entropy [121]. The prediction error E is propagated layer by layer back through the network to measure the contribution from each connection using the chain rule [112]. During this backward propagation, the error gradient of the loss function with respect to the weight matrices W and the bias terms b determines how each element needs to be updated to minimize the error. The backward pass starts by computing the error gradient $\frac{\delta E}{\delta y_j} = \hat{y}_j - y_j$ for each of the $0...j$ output units [112]. Subsequent application of the chain rule allows computation in the penultimate layers y_{j-1} with

$$\frac{\delta E}{\delta y_{j-1}} = \frac{\delta E}{\delta y_j} * \frac{\delta y_j}{\delta y_{j-1}} \tag{1.3}$$

[112]. Repeating the chaining procedure propagates the error gradient up until the input layer. The weights and biases are updated by an amount proportional to the accumulated error gradient following the gradient descent formula [112]

$$w^l_{ji} = w^l_{ji} - \eta * \frac{\delta E}{\delta w^l_{ji}} \tag{1.4}$$

where w^l_{ji} represents a given weight connection between node j in layer $l-1$ and node i in layer l, and η is the chosen learning rate. Iterating the forward pass (aka creating a prediction based on the current weights and biases) and backward process (aka weight and bias update based on the respective error gradient) until the network converges to a solution ultimately allows the network to learn the desired function [112]. The backpropagation algorithm is a fundamental element of neural networks and ultimately permits the training of deeper networks in the first place.

Due to the inherent interconnectivity, fully-connected layers have, however, several downsides for image segmentation. The number of parameters is exceptionally large for full-size images and fully-connected networks are not robust to variations in object positions. Inspired by the optical processing systems in cats, these challenges have been addressed by replacing the classic linear transforms with convolutional kernels giving rise to a new class of neural networks called CNN [107].

1.2.1 Convolutional Neural Networks for Image Segmentation

The idea behind CNNs arises from the knowledge that distinct local motifs in an image are formed by highly correlated value groups that are not locally restricted [107]. An upward-pointing edge, for example, can occur anywhere within the image. By training the network, the weight parameters $\theta = [W, b]$ are updated to recognize a specific motif, independent from its position within the image. The hierarchical combination of these motifs in deeper layers allows the detection of more complex, high-level patterns, such as an eye or a nose. To achieve these attributes, the CNN repetitively uses convolution and pooling layers (scaling transitions) depicted in Figure 1.6.

Convolution Layer

The role of the convolutional layer is the detection of local conjunctions of features or motifs [107]. The input and output of convolutional layers are called feature maps. In the first layer, the input feature map is the original input image. Each unit in a convolutional layer is only connected to a local patch or receptive field of the feature maps of the previous layer via a set of so-called filters or kernels. During training, each of these kernels specializes in recognizing a certain pattern. The locally weighted sum of the input feature y^{l-1} with the kernel k of size (k_1, k_2) creates a pre-activation output signal for each position i, j in the output feature map z^k (Equation (1.5)). Normally, multiple kernels are applied, each producing a separate feature map output (e.g. 64 in Figure 1.7). The convolution kernel extends across all N feature maps of the previous layer $l-1$. Further, one bias term b_k for each kernel k exists. In contrast to the fully-connected networks, all units within a feature map share the same weights $w_{a,b}$ for a given kernel, significantly reducing the number of network parameters. Before applying the non-linear activation function f to create the final output feature map y^k for kernel k, it is common to add a batch normalization layer to reduce the internal covariate shift and improve network training [122].

$$\begin{aligned} z_{ij}^k &= bk + \sum_{a=0}^{k_1-1} \sum_{b=0}^{k_2-1} \sum_{c=0}^{N-1} w_{a,b}^k \times y_{(i+a),(j+b),c}^{l-1} \\ y_{ij}^k &= f(z_{ij}^k) \end{aligned} \tag{1.5}$$

The consecutive application of 64 convolution kernels of size 3×3, batch normalization, and activation function is depicted as H_a^j in Figure 1.7 (top). Such a sequence of convolution layers is also called a network block. Consecutive application of convolutions increases the receptive field size (colored squares in the bottom part of Figure 1.7) considered for the calculation of the feature response in the following layers at each voxel by $(k_1 - 1 \times k_2 - 1)$.

Dense Skip Connections

Dense connections mimic a deeper network while promoting feature reusability and facilitating gradient flow [123]. They have been implemented via concatenation of feature maps, i.e. the output of previous convolution layers are stacked on top of each other before passing the signal forward.

Considering a sequence of three convolutions $H_a^{1..n}$ within a given block a acting on the feature

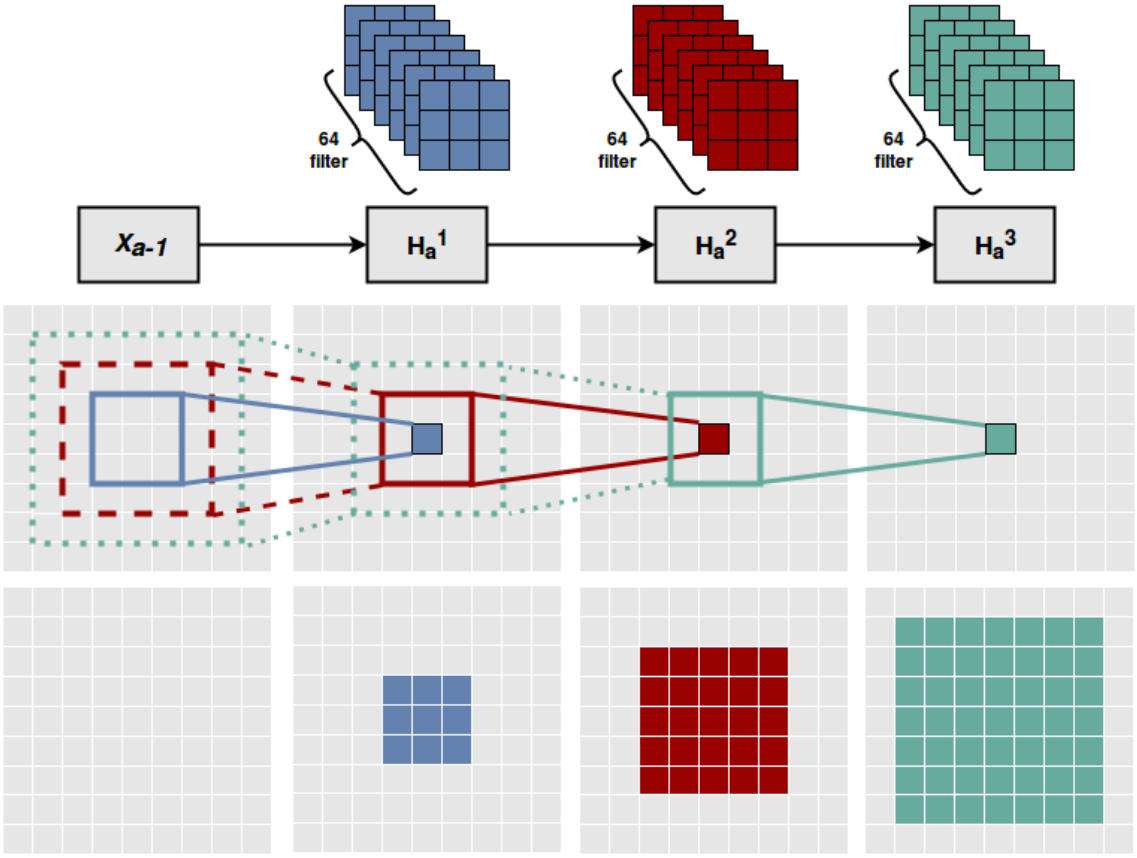

Figure 1.7: A sequence of three convolutions, each with 64 filters of size 3×3, on the input image X_{a-1} is performed inside each convolution block. The second row depicts the receptive field considered for calculating the feature response at the central colored unit on the input image (far left) and all subsequent feature maps. Due to zero padding along the borders, the size of all feature maps is constantly equal to the input map. The bottom row visualizes the growing receptive field size on the input image with each convolution. A series of these convolutions with a kernel size of 3×3 incrementally increases the receptive field, which in effect creates finer to coarser filters, e.g. H_a^3 operates on a 7×7 receptive field in contrast to a 3×3 field for H_a^1.

map output of the previous block X_{a-1} as depicted in Figure 1.7, the dense blocks can be described as:

$$\begin{aligned} \mathbf{y}_a^1 &= \quad \text{concat}(H_a^1(\mathbf{X}_{a-1}), \mathbf{X}_{a-1}) \\ \mathbf{y}_a^2 &= \quad \text{concat}(H_a^2(\mathbf{y}_a^1), \mathbf{y}_a^1, \mathbf{X}_{a-1}) \\ \mathbf{X}_a &= \qquad\qquad\qquad H_a^3(\mathbf{y}_a^2) \end{aligned} \tag{1.6}$$

Commonly, a 1×1 convolution is added at the end of a dense block to reduce the dimensionality of the final output map along the channel dimension. Dense connections are also often used to concatenate feature maps of distant network blocks. These so-called global skip connections allow the gradient to flow earlier to shallower parts by skipping intermediate layers.

Scale Transition Layer

Pooling operations represent fixed-factor integer down-scale transitions along the spatial dimensions (width and height) that reduce the feature map sizes [107]. The channel dimension (aka number of feature maps) is commonly kept constant. Akin to convolutional layers, each unit within a pooling layer is connected to a small, rectangular field of the previous feature maps. Calculating the maximum or average across the local patch effectively reduces all units within the receptive field to a single output [124–126]. Neighboring pooling units are commonly separated by a constant stride, reducing the feature map sizes by a fixed scale (commonly factor two). Doubling the number of convolution kernels with every pooling operation is a common design choice in modern convolutional architectures to trade off reduced spatial resolution for greater channel depth.

In effect, the downsampling represents a natural transition to smaller scales (i.e. resolutions) and allows a reduction in parameters, memory usage, and computational load as well as capture of a larger receptive field. Adding pooling between convolutional blocks further introduces resilience to small, local shifts and distortions into the CNN, giving rise to the important property of translation invariance [107, 121]. With invariance, the system evokes the same response, irrespective of the transformation of the input. In other words, a translational invariant system will recognize a feature such as an eye or nose independent of its location within the image.

In feedforward CNNs for classification tasks, the last layer is a fully-connected layer connecting all units from the respective feature maps and produces a single prediction value (Figure 1.6, top right). In early segmentation networks, segmentation maps were hence produced by subsequently classifying one voxel after the other [127, 128]. For large images, this procedure is very inefficient. Newer networks therefore use fixed-factor integer up-scale transitions via index unpooling or transposed convolutions to revert the coarsened, pooled feature maps to the original image size [129–132]. These so-called Fully Convolutional Neural Networks (FCNNs) can operate on an input of any size and return a corresponding segmentation of the original spatial dimensions. In the original FCNN, the coarse output maps created via the pooling operations are directly upsampled to the input image size using backwards-strided convolution with an output stride of f [129] (Figure 1.6, middle). Later works increased the pooled input step-wise by alternating convolution blocks with deconvolutions [130, 131] or index unpooling [132] (Figure 1.6, bottom). In the latter approach, the indices of the max or average values from the pooling operation are retained and used to position the values from a lower layer within the original size of the activations [132]. Theoretically, any sampling strategy that allows gradients to flow effectively through the network can be used instead of the mentioned scale-transition operations (i.e. pooling-/unpooling).

External Data Augmentation

External data augmentation describes the process of transforming the appearance of existing training instances to artificially increase the training set and avoid overfitting [108, 118, 133–136]. Augmentations are of special interest in medical image segmentation where the availability of labeled training data is sparse due to high annotation costs. Consequently, training datasets are often small and lack diversity (e.g. limited variations in scanning centers, MRI vendors, imaging protocols, diseases, age groups, or nationalities) leading to poor generalizability. Exposing the network to more variability through augmentation improves robustness and reduces potential bias from missing or unbalanced data characteristics [108, 118, 133–136]. Importantly, the transformation should create realistic instances

representing the distribution of interest. Common augmentations are affine transformations (rotation, translation, scaling), intensity augmentations (brightness, contrast, saturation, or hue changes), the addition of noise (Gaussian or white noise), background alterations, or random cutouts [136].

For image segmentation, the affine transformations must be applied to both, the intensity image and the ground truth segmentation to preserve spatial correspondence. Any affine transformation operation normally changes the coordinates of the regular grid points representing the image (pixel or voxel) [133]. The values at these new sampled points are unknown and must thus be estimated from the existing information by fitting a continuous function. A larger difference between the original and target position, orientation, and size, hence increases the ratio of explicit (i.e. measured) versus implicit (i.e. inferred) intensity information. Interpolation of discrete label maps further requires the usage of the nearest neighbor kernel and is thus susceptible to visual artifacts and loss of image quality [137–140]. So far, external augmentations are, however, still the method of choice in CNNs.

1.2.2 The U-Net - A New Framework for Biomedical Image Segmentation

A special type of FCNN that has become one of the most popular network choices for medical image segmentation is the U-Net [130, 141]. The name originates from the organization of the building blocks within its encoder-decoder architecture resembling the letter U. The encoder or contracting part alternates convolution blocks with pooling operations to extract features from the input signal. Intuitively, the encoder compresses high-level information (where are things in the image) into a condensed representation of its content (what is depicted). The decoder or expanding part is the key feature setting the U-Net apart from previous FCNNs [129]. Instead of a single upsampling operation to the original input image size at the end of the contracting part, the decoder is divided into stages of convolution layers and transposed convolutions mimicking the encoder design [130]. The resolution of the feature maps is therefore gradually increased at each stage allowing the decoder to synthesize the segmentation sequentially from crude, large-scale structures visible in the low-resolution feature maps, to small, fine-grained details present at the higher resolutions [141]. Global skip connections between the encoder and decoder aid in preserving the high-resolution contextual information at each level and allow the convolutional layers of the decoder to learn a precise representation. These connections allow gradients to flow efficiently and significantly speed up the learning process [130, 141]. The original implementation does not use any padding, and starting from a size of 64 changes the number of kernels after each sampling step by a factor of 2. The lower layers, therefore, have a larger number of channels with 1024 being the maximum in the bottleneck layer.

This network architecture has been proven to be highly useful for image segmentation: To this day, it is one of the most cited papers for biomedical segmentation, and new architectures based on its core ideas are continually published [142]. Many works in the field of deep learning for brain segmentation have taken inspiration from the U-Net as well [2, 119, 120, 143–150].

1.2.3 Deep Learning for Brain Segmentation

Brain segmentation is not a trivial task, specifically when a detailed delineation into multiple cortical structures is desired. First, short- and long-range contextual information is important to capture the shape and location of structures with respect to each other. Cortical parcellations are particularly difficult to segment as the intensity values are mostly identical across the cortical GM. Consequently, the network has to learn to distinguish them purely by their folding patterns and relative location.

Full-view 3D networks would therefore be an ideal choice in theory. Their practical application is, however, currently limited due to significant memory requirements specifically for predicting a larger number of classes and higher-resolution images. Segmenting downsampled images is also not an option due to insufficient structural resolution. Alternatively, both volume slice-based 2D, as well as patch-based 3D approaches in combination with view aggregation or sliding window schemes, have been proposed in recent years to leverage as much spatial information as possible. Patch-based 3D CNNs like DeepNat [151], 3D U-Net [152], AssemblyNet [150] or SLANT [144] have limited long-range context with a small field of view. Some of the architectures further require several additional, potentially error-prone processing steps including image registration to a standard space and bias field removal, as well as inference for many networks [144, 150]. Slice-based approaches trade local 3D context for a full field of view [119, 153–155]. To address the information loss in the third dimension, multiple 2D U-Nets can be trained separately on different representations, such as the axial, coronal, and sagittal anatomical views. Averaging the prediction probabilities across the networks regains 3D information and effectively creates a 2.5D approach. This concept was introduced in Quick Segmentation of NeuroAnaTomy (QuickNAT) [119]. Performance comparison between 2.5D and 3D approaches found limited differences, even after extensive optimization of 3D network architectures [156].

Second, the structures within the brain largely differ in size, hence creating a strong class imbalance. As an example, for a thirty-year-old cognitive normal male, the average total WM volume is around 50.000 mm^3, while all subcortical structures together comprise a volume of around 6.000 mm^3 [157]. Therefore, training instances do not represent structures equally biasing segmentation accuracy across classes. Additionally, correct delineation of tissue boundaries between structures is aggravated by PVE (see Section 1.1.3). Data sampling strategies [134, 158] or loss function weighting schemes [119, 154, 159] are commonly used to address these challenges.

Third, the estimation of reliability and sensitivity to real neuroanatomical effects is pivotal for the translation of the tools to clinical practice. However, networks for brain segmentation mainly estimate segmentation accuracy exclusively via overlap (e.g. Dice Similarity Coefficient (DSC)) or distance measures (e.g. Average Surface Distance (ASD), Hausdorff Distance (HD)).

Fourth, existing networks for brain segmentation are restricted to volume-based segmentation limiting their potential for neuroimage analysis. Traditional pipelines go far beyond image segmentation and provide utilities such as the creation of cortical surface models, estimation of thickness, construction of fiber tracts or functional connectivity graphs, and tools for group comparison, such as registration and statistical frameworks [94, 102, 160]. The surface-based analysis has proven pivotal for estimating indices of cortical morphology, such as cortical thickness (Figure 1.8), and improves reliability and detectability of effects [161–166]. Integrating developed deep learning frameworks into complete, self-contained imaging pipelines would greatly benefit the neuroimaging community and facilitate progression to next-generation deep learning tools.

Fifth, the aforementioned changes in appearance with age and disease need to be addressed to ensure robustness (see Figure 1.1). Considering the adult cohort, generating diverse training sets is possible at the de-facto standard resolution of 1.0 mm. High-resolution datasets, on the other hand, are sparse and lack diversity. As an example, the openly available RS [75] or HCP [21] are currently restricted to healthy cohorts and further differ in their scanning resolution. So far, published networks for brain segmentation predominantly follow a fixed-resolution scheme and do not support high- or multi-resolutions [119, 120, 134, 143–151]. While downsampling of the MRIs may be used to harmonize the data, structural information is lost and may introduce biases into the analysis.

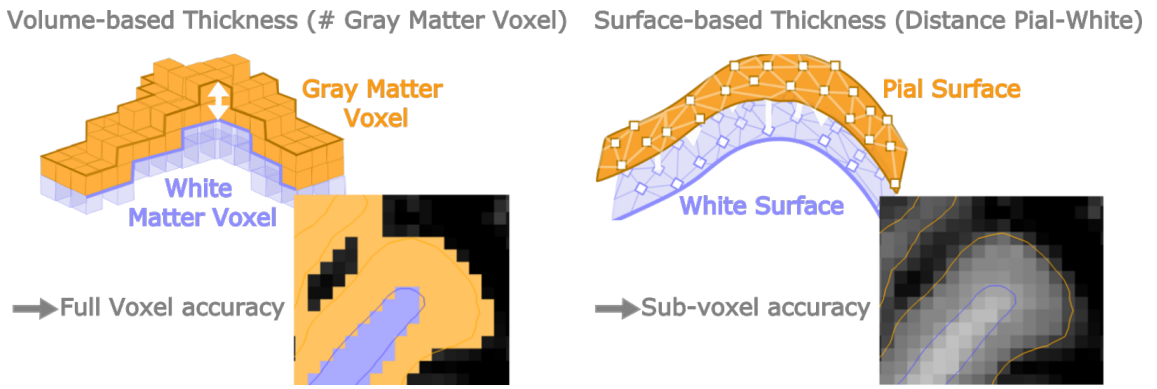

Figure 1.8: Surface-based morphology analysis improves thickness measures. Derivation of cortical thickness from segmentation maps is limited to the summation of voxels (left), while surface reconstruction allows estimation at a sub-voxel level (right). Here, thickness is estimated as the distance between the pial and the white surface.

Considering the rise in high-resolution studies and proven improvement of structural assessments in sub-millimeter MRIs, the development of resolution-independent tools allowing segmentation at any native resolution is a critical next development step.

Sixth, dedicated network solutions for age groups other than adults are scarce. Datasets for newborns, for example, are far harder to come by while displaying substantial variability in head morphometry and position within the scanner. These factors all pose significant challenges for method development. Tools developed for adults do not automatically work due to significant anatomical differences [167]. Network architectures addressing specifically the non-standardized head positions are currently not available and would provide a valuable contribution to the neuroimaging community.

1.3 Contributions of the Thesis

Extracting morphological information from brain MRIs is a pivotal step to reveal subtle pre-symptomatic disease effects, further our understanding of health and disease progression, and evaluate disease-modifying therapies. Non-invasive neuroimaging studies, therefore, collect large amounts of human brain imaging data across different populations and cohorts. The size and pace of these recordings have increased significantly over the years, creating the need for efficient, reliable, and accurate analysis methods. Traditional neuroimaging pipelines, however, rely on atlas-based registration processes, extensive feature extractions, and pre-processing steps, limiting their scalability. Given their recent success in several image analysis tasks, CNNs offers the potential to address this challenge by drastically improving runtime and performance. Existing architectures for brain segmentation, however, lack utilities such as the creation of cortical surface models, and estimation of thickness or ROI measures and are limited to volume-based processing in fixed, single-resolution settings.

This thesis therefore aims to advance the field by developing dedicated deep learning methods for the unbiased segmentation of multi-resolution whole brain MRI in adults and newborns. The contributions are detailed in the accompanying sections Chapter 2 to Chapter 6 and summarized below.

Contribution 1: FastSurfer - the first deep learning based end-to-end pipeline for the automated segmentation and surface reconstruction of adult brain MRI. With FastSurferCNN, a 2.5D CNN capable of whole brain segmentation into 95 classes in under 1 minute is presented. Integration into a novel surface reconstruction stream provides a full alternative to the well-established traditional neuroimaging pipeline FreeSurfer, including pointwise and ROI thickness analysis.

Contribution 2: Parameter space CNN for cortical surface segmentation. Comparison of a novel surface parameterization approach termed p^3CNN with a sphericalCNN directly operating on the geometric mesh reveals drawbacks of current geometric deep learning approaches and introduces the robust p^3CNN for cortical surface segmentation.

Contribution 3: The Voxel-size Independent Neural Network (VINN) for native resolution image segmentation. Innovative network architectures for synergistic usage of varying (high-)resolutions across a range of different voxel sizes using a novel network-integrated resolution-normalization layer and dedicated high-resolution loss function are developed and evaluated. The VINN network retains important image information at the native resolution while maintaining generalizability across potentially underrepresented groups. The developed concept has a broader impact on semantic segmentation and may prove useful beyond the primary application in neuroimaging.

Contribution 4: Identifying and combating bias in segmentation networks. Assessment of the effect of image resampling, scale augmentation, and resolution-independence reveals that single-resolution training settings translate to systematically biased predictions. Biases can effectively be reduced with multi-resolution approaches including the VINN.

Contribution 5: Orientation-independent whole brain segmentation for newborn MRIs. Development and evaluation of a novel internal positioning augmentation via a four-degree of freedom (4-DOF) transform module to address the issues of label sparsity and increased head position/orientation variability in the newborn cohort. Comparison to external augmentation approaches and traditional pipelines for newborn segmentation shows positive effects on segmentation performance with this paradigm shift.

Successful completion of the proposed work contributes a well-documented and validated open-source software framework for fast, deep learning based, multi-resolution neuroimage analysis, supporting efficient big data processing, pre-symptomatic disease studies, drug treatment assessment, and personalized medicine. Furthermore, the impact of the proposed novel deep learning methods may extend to other medical applications and research domains.

CHAPTER 2

FastSurfer - A Fast and Accurate Deep Learning Based Neuroimaging Pipeline

The following chapter is based on the publication [2]:

> **FastSurfer - A fast and accurate deep learning based neuroimaging pipeline**
> Authors: Leonie Henschel, Sailesh Conjeti, Santiago Estrada, Kersten Diers, Bruce Fischl, and Martin Reuter
> Published in: *NeuroImage*, volume 219:117012, 2020
> DOI: 10.1016/j.neuroimage.2020.117012

2.1 Individual Contribution

This publication was done in collaboration between Sailesh Conjeti, Martin Reuter, and me. I programmed and evaluated several different network architectures by extending the initial software written by Sailesh Conjeti. I tested the existing code components, wrote supporting algorithms (metric scripts, data processing, loss functions, as well as training and inference scripts), and verified the reproducibility and statistical significance of the segmentation results. I also assisted in the construction of the surface pipeline. I co-wrote parts of the RECON-SURF.sh script with Martin Reuter, formulated and programmed the surface metrics, group analysis, and Test-Retest analysis. I conducted the research and investigation process of both, the segmentation model and surface stream. I wrote the original draft of the introduction and discussion section. I co-wrote the material and methods part with Martin Reuter and Santiago Estrada and the results section with Martin Reuter.

2.2 Summary

2.2.1 Motivation

Automated segmentation of structural MRI, the labeling of anatomical ROIs in images, is a pivotal task in neuroimaging analysis. It allows quantification of segmented brain structures with respect to different measures (e.g. volume or thickness) and consequently serves as a starting point to quantitatively

model healthy development and understand disease effects or age-related changes. Structural MRI segmentations are further important to relate functional and diffusion signals to anatomical regions, and hence establish a structure-function relationship. Due to the increase in both, population and clinical studies, there is a need for efficient, scalable, and reliable software in the neuroimaging community. Traditional neuroimage pipelines, such as FreeSurfer [94], SPM [102], or FSL [103], are currently the only way to handle the incoming flow of data. However, these methods involve computationally intensive optimization steps and thus do not scale well to large cohorts. Orders of magnitude faster deep learning networks have the potential to revolutionize MRI analysis, but existing methods [144, 145, 150, 151, 153–155, 168] are so far primarily focusing on voxel-based segmentation and only sparsely validated. Evaluation of reliability and sensitivity to real neuroanatomical effects as well as surface-based analysis, which has proven pivotal e.g. to correctly estimate cortical thickness, has so far not been addressed in comparative publications.

This paper fills this gap and is the first publication integrating a novel deep learning framework (FastSurferCNN) into a complete, self-contained imaging pipeline called FastSurfer. FastSurfer provides a fully automated FreeSurfer alternative for volumetric morphology (in under 1 minute) and surface-based thickness analysis (within approximately 60 minutes, depending on image quality and process parallelization).

2.2.2 Network Architecture of FastSurferCNN

The volumetric analysis utilizes FastSurferCNN (Figure 2.1), a U-Net-like CNN with four dense encoder-decoder blocks equipped with long- and short-range skip connections [123, 130], to create a whole-brain segmentation into 95 anatomical regions based on the Desikan-Killiany-Tourville atlas [14, 15].

To recapture spatial information in the third dimension, FastSurferCNN uses a 2.5D approach: a multi-view aggregation combines the probability maps of three separately trained 2D Encoder-Decoder networks (one per anatomical view, see Figure 2.1, right side). FastSurferCNN further contains two methodological advances tailoring performance towards accurate segmentation of cortical and subcortical brain structures. The first methodological improvement is the competitive dense block (CDB) where concatenations commonly found in dense blocks are replaced with the efficient maxout operation [169–171]. As depicted in Figure 2.1 (left side), the encoder and decoder consist of four CDBs separated by a bottleneck layer. Max pooling and index unpooling implement scale transitions between the blocks. Each CDB contains three layers each consisting of a probabilistic ReLU (pReLU), a 5x5 convolution, and a batch-normalization. Within the blocks, the layer outputs are combined via maxout instead of concatenations. This change reduces the number of trainable parameters by one-half (to approximately $1.8 * 10^6$) while improving segmentation accuracy on both, cortical and subcortical structures by 0.3 % DSC and 5 % average HD.

The second methodological improvement, the spatial information aggregation (SPI), adds 3D context to each single view CNN. Stacking the three preceding, the current, and the three succeeding 2D slices into a 7-channel input image, synergistically combines the advantages of 3D patches (local neighborhood) and 2D slices (global view). The SPI is particularly useful for the cortical structures, where segmentation accuracy improves by 1.9 % (DSC) and 12.7 % (average HD).

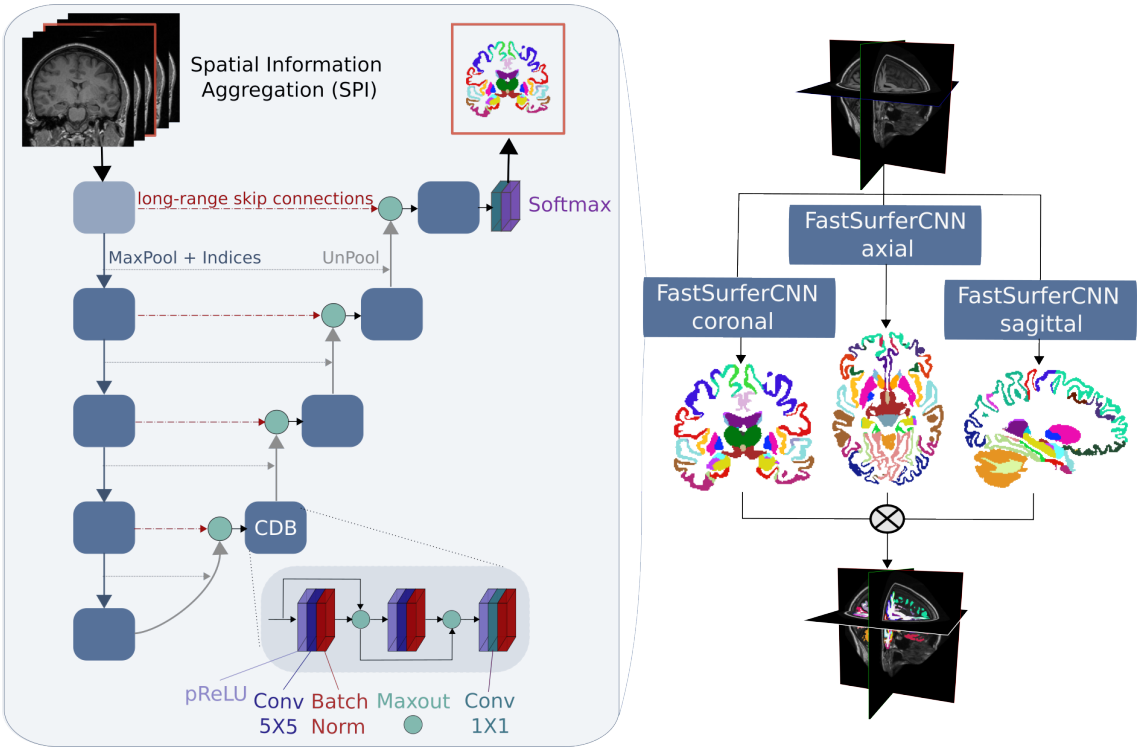

Figure 2.1: Architectural design of FastSurferCNN - Four encoder and decoder blocks separated by a bottleneck layer and connected via long-range skip connection formulate the base. The competitive Dense Blocks (CDB) include three layers of consecutive probabilistic rectified linear unit (pReLU), 5x5 convolution, and batch-norm. The layer outputs are combined via maxout. Spatial information aggregation (SPI) provides the 3D context within each 2D FCNN by presenting three neighboring slices in addition to the segmentation target. The 2.5D architecture originates from averaging the probability maps of the three separately trained 2D FCNNs (one per anatomical view).

2.2.3 Automatic Generation of Surfaces and Derived Measures - Surface-Based Processing

Based on the accurate 3D whole brain segmentation from FastSurferCNN the surface pipeline called RECON-SURF reconstructs cortical surfaces, thickness maps, and summary statistics in cortical regions. A significant speed-up compared to FreeSurfer can be achieved by omitting several steps that have become obsolete due to the presence of the brain segmentation, such as skull stripping and non-linear atlas registration. Furthermore, in contrast to FreeSurfer's iterative inflation (Figure 2.2, upper part), FastSurfer introduces an efficient spherical mapping via a novel spectral approach (Figure 2.2, lower part). More specifically, the spectral projection solves the Laplace-Beltrami Eigenvalue problem $\Delta f = -\lambda f$ [172, 173] on the original cortical surface mesh to obtain the first three non-constant Eigenfunctions with smallest Eigenvalues (entitled $f1$, $f2$, and $f3$ in Figure 2.2). After correcting sign flips and swaps, these functions parameterize the surface smoothly in lateral-medial (x), anterior-posterior (y), and superior-inferior directions (z). Subsequent scaling of the 3D Eigenfunction vector to unit length for each vertex projects the 3D spectral embedding to the sphere. Due to the smoothness of the Eigenfunctions and the well-behaved global ellipsoid-like geometry of the brain,

this fast approach reduces self-folds.

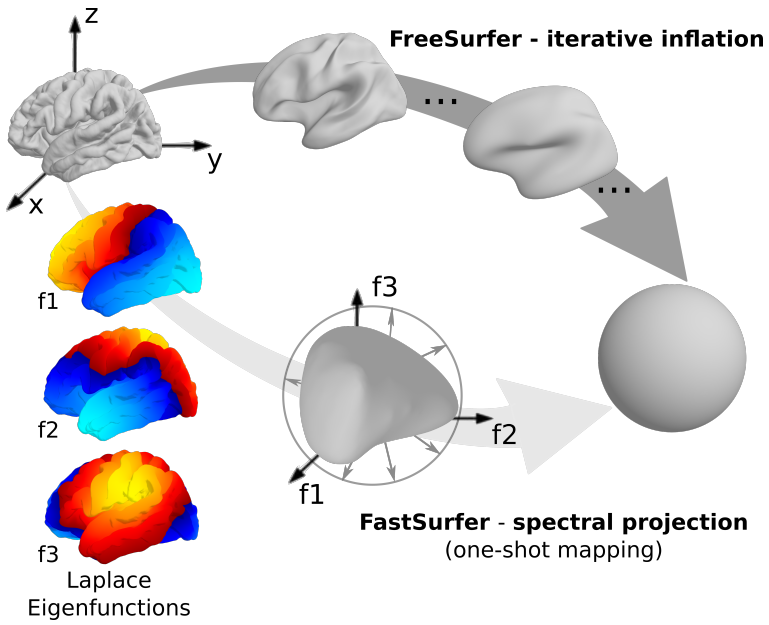

Figure 2.2: Spectral projection for one shot mapping from surface to sphere in FastSurfer - Instead of the time-consuming iterative inflation from FreeSurfer (top), the spectral projection (bottom) maps the first three non-zero Eigenfunctions ($f1$ to $f3$) of the Laplace operator with smallest Eigenvalues to the sphere by scaling the 3D Eigenfunction vector to unit length for each vertex. The 3D spectral embedding parameterizes the surface smoothly in anterior-posterior ($f1 = y$), lateral-medial ($f2 = x$), and superior-inferior ($f3 = z$) directions.

Further, instead of the time-intensive non-linear spherical atlas registration and segmentation in FreeSurfer, the segmentations are directly mapped from the image onto the surface. Subsequently, surface ROI statistics, such as thickness and curvature averages per region can directly be calculated. The introduced changes improve mesh quality and reduce the number of topological defects from an average of 27.2 to 23.1. Overall, processing time per hemisphere reduces from 41.6 minutes in the original FreeSurfer stream to 25.4 minutes in FastSurfer.

2.2.4 Sustainability and Dependability - A Valid Alternative to FreeSurfer?

For sustainability FastSurfer is extensively validated: DSC and average HD assert high segmentation accuracy and generalizability. On five independent test sets FastSurferCNN significantly outperforms state-of-the-art methods QuickNAT [119] and 3D U-Net [152], both with respect to FreeSurfer and a manual standard as a reference. Segmentation accuracy improves on unseen downsampled and defaced images from the HCP [21], with a final DSC of 0.87 on the cortex and 0.85 on the subcortical structures. Based on 180 participants from ADNI [61], FastSurferCNN generalizes well across different vendors (final DSC of 0.90/0.86 on GE, 0.89/0.86 on Philips, and 0.90/0.87 on Siemens for the cortical and subcortical structures, respectively) and disease states (consistent DSC of 0.90/0.87 for cognitive normal participants and 0.89/0.86 for diagnosed Alzheimer's Disease patients).

Compared to FreeSurfer, FastSurfer exhibits high Test-Retest reliability, with an average Intraclass Correlation Coefficient (ICC) of 0.9 on cortical and 0.99 on subcortical regions for 20 Test-Retest

subjects from OASIS1 [22]. In a group comparison between 273 non-demented and 97 demented subjects from OASIS1 [22], FastSurfer increases sensitivity relative to FreeSurfer and robustly detects reduced cortical thickness in regions associated with dementia as well as subcortical volume differences.

In addition to the performance improvement, FastSurfer simultaneously reduces processing time by a margin. Segmentation results are available within 1 minute and surface-based thickness analysis within 1 hour (plus optionally 30 minutes for group registration) compared to 4 hours with FreeSurfer (under optimal conditions, i.e. parallel mode with 4 threads).

2.2.5 Conclusion

Overall, FastSurfer combines the speed of supervised deep learning approaches with the convenience of the broad spectrum of surfaced-based features and analysis methodologies provided by traditional neuroimaging pipelines. Reliability and sensitivity are on par with FreeSurfer while the processing time is greatly reduced. As such, FastSurfer addresses a significant need within the community for efficient, stable, reliable, and sensitive pipelines for automated neuroimage analysis.

CHAPTER 3

Parameter Space CNN for Cortical Surface Segmentation

The following chapter is based on the publication [5] (reproduced with permission from Springer Nature):

> **Parameter Space CNN for Cortical Surface Segmentation**
> Authors: Leonie Henschel and Martin Reuter
> Published in: Tolxdorff, T., Deserno, T., Handels, H., Maier, A., Maier-Hein, K., Palm, C. (eds) *Bildverarbeitung für die Medizin 2020. Informatik aktuell*. Springer Vieweg, Wiesbaden, 2020
> DOI: 10.1007/978-3-658-29267-6_49

3.1 Individual Contribution

I developed, implemented, tested, and validated all models, software, and supporting algorithms. I had the original idea to create a view-aggregated 2D parameterization approach to reduce distortions introduced by non-isometric projections and compare this method to a spherical neural network for the segmentation of the human cortex. I also originated the idea to improve the spherical network by dense connections and a spherical pooling approach. The research goal and conceptualization were discussed and fine-tuned with Martin Reuter. He provided valuable feedback, financial support, computing resources, and guidance throughout the development phase. Martin reviewed the paper and refined sections of the original draft.

3.2 Summary

3.2.1 Motivation

Cortical surface segmentation plays a crucial role in various neuroimaging applications. Surface-based signals such as curvature or thickness encode valuable anatomical information and allow better mapping of functional activations, improve subject registration, or estimate myelination content. To simplify analyses, cortical surfaces are commonly represented as spheres [175]. The paper "Parameter Space

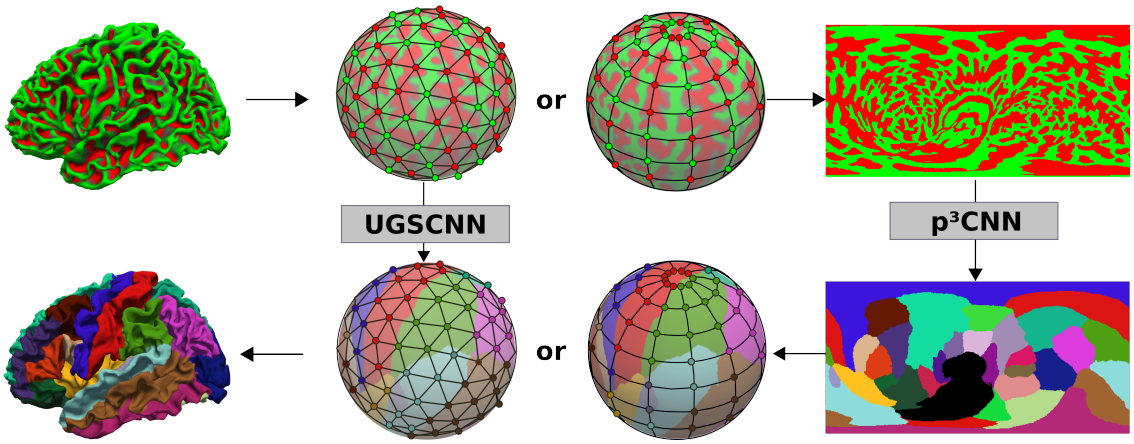

Figure 3.1: Two segmentation networks are compared: a spherical CNN (UGSCNN [174]) on the icosahedron (left) and the proposed view-aggregation on 2D spherical parameter spaces (p^3CNN, right). Both operate on curvature maps (top row) and thickness (not shown) for cortical segmentation of the cortex (bottom row). Figure and description originally from [5]. Reproduced with permission from Springer Nature.

CNN for Cortical Surface Segmentation" proposes a new approach for segmentation of the human cortex using voxel grid-based deep learning in combination with a 2D spherical parameterization approach and view aggregation. The novel method termed p^3CNN is compared to the only available state-of-the-art unstructured grid CNN for semantic segmentation at the time of publication (UGSCNN [174]) and highlights limitations of spherical deep learning approaches.

3.2.2 UGSCNN: Spherical CNN on Unstructured Grids

The only available geometric deep learning network for segmentation at the time of publication is a spherical CNN called UGSCNN [174], where a linear combination of parameterized differential operators weighted by a learnable parameter represents the convolutional kernel. The spherical domain is approximated by an icosahedral spherical mesh (Figure 3.1, UGSCNN) to allow well-defined coarsening of the grid in the downsampling step [174]. It serves as the reference method for segmentation accuracy comparisons.

3.2.3 Parameter Space CNN

Convolution filters used in traditional CNNs do not work on the unstructured triangular mesh representations of the sphere. To circumvent this problem, the parameter space CNN transforms the signal from the sphere to an image-like 2D representation (Figure 3.1, p^3CNN) using a spherical coordinate parameterization. In turn, established CNN architectures become applicable. Here, a multi-modal 2D Dense U-Net [123] with four dense encoder and decoder blocks forms the network base. Each dense block consists of a sequence of three 2D convolution layers with 64 kernels of size 3x3. Between the blocks, an index-preserving max-pooling operation halves the feature map size.

To create parameter space "images" to train on, the longitude/colatitude coordinate system parameterizes the sphere. Each vertex position on the sphere (x, y, z) is described by (i) the azimuthal angle $\varphi \in [0, 2\pi]$, (ii) the polar angle $\theta \in [0, \pi]$ and (iii) the radius r=100 via the spherical

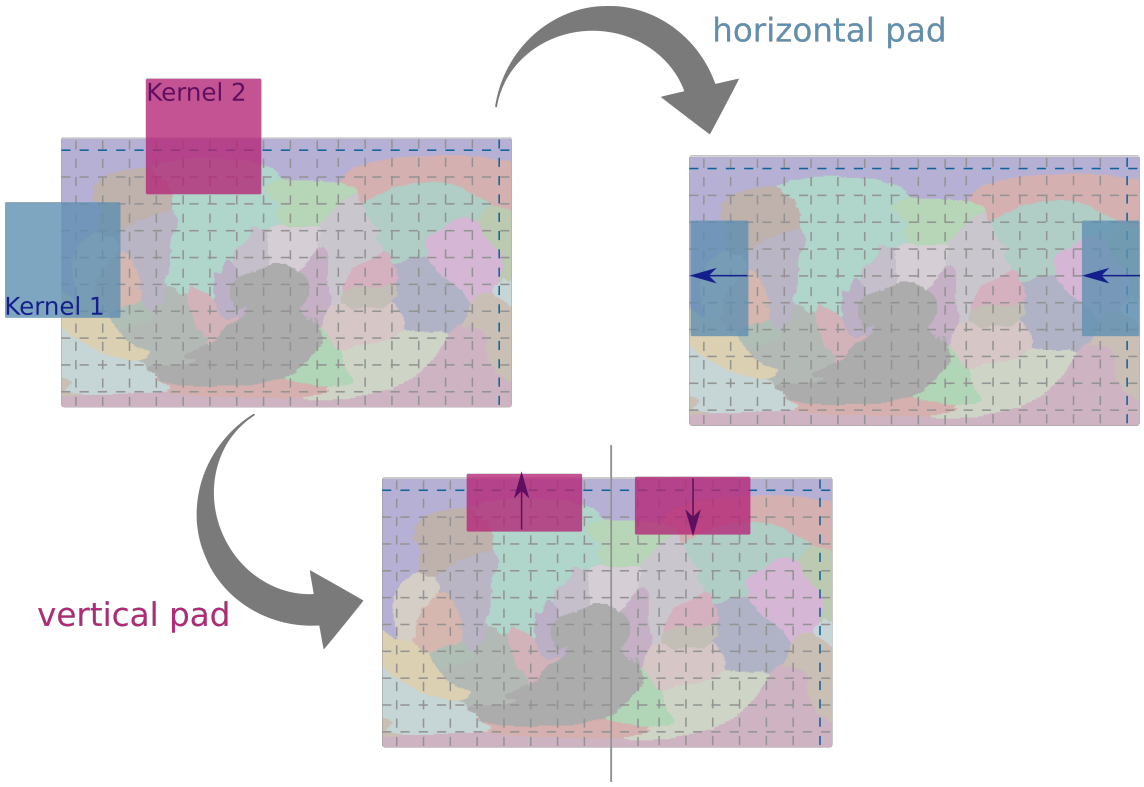

Figure 3.2: Imitation of spherical signal processing in 2D planar projection by padding. Horizontal padding with values from the opposing side allows a smooth transition from left to right mimicking transitions around the sphere. Vertical padding with mirroring of the split top or bottom half imitates transition across the poles.

parameterization:

$$x = r\ \sin\theta\cos\varphi, \quad y = r\ \sin\theta\sin\varphi, \quad z = r\ \cos\theta \tag{3.1}$$

After the transformation step, the signal of interest (thickness, curvature, or label map) is sampled at the given coordinates (φ, θ) on the left and right hemisphere and projected onto the 2D parameter grid (Figure 3.3). To further enforce spherical topology, appropriate padding is used. Before each convolution, the left and right borders are extended with values from the opposing side to provide a smooth, horizontal transition (Figure 3.2, horizontal pad). The top and bottom borders are padded by splitting them in half and mirroring about sideways around the center thereby modeling a transition across the poles (Figure 3.2, vertical pad).

A perfect (isometric) mapping between plane and sphere does not exist. The Mercator projection used here results in a non-uniform distribution of sample points with singularity issues at the poles. To address these issues, the corresponding angles θ are shifted by half the grid width when sampling the (φ, θ) parameter space to the (i, j) grid (see Figure 3.3). Further, rotation of the grid such that the poles are located along the x-, y- and z-axis diversifies sampling, and allows training of one network per rotation followed by a final aggregation of the resulting probability maps: (i) First, the label probabilities of each network are mapped to the original WM spherical mesh by computing a distance-weighted average of the three closest vertices on the sphere to each target vertex. (ii) Second,

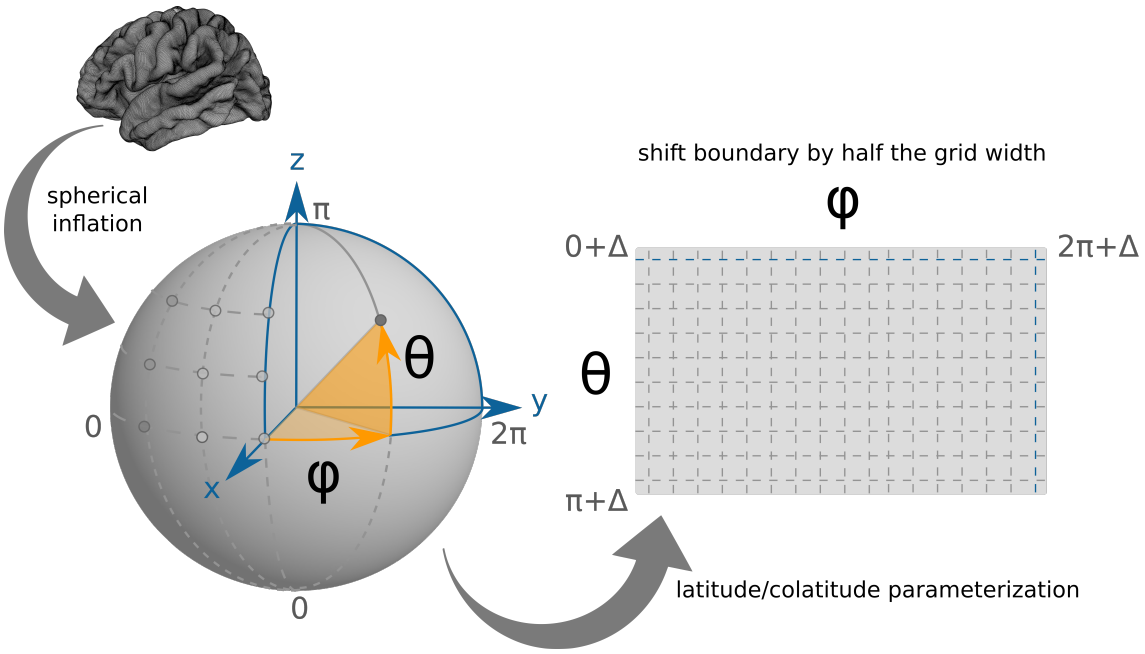

Figure 3.3: From surface to plane - transformation of the FreeSurfer signals to the plane for pCNN. Coordinates of the spherical plane described by angles (φ, θ) are projected to the 2D parameter space grid (i, j). To avoid singularity issues at the poles, the polar angle is shifted by half the grid width.

the three probability maps are averaged on a vertex-by-vertex basis to produce the final label map. Due to the view aggregation across three parameter spaces, this approach is called p^{3}CNN.

3.2.4 Segmentation Accuracy - Unstructured versus Structured Grids

The segmentation accuracy of the UGSCNN, the parameter space CNN (pCNN), and the view aggregation (p^{3}CNN) are compared by computing the surface-based DSC in the subject space on the original brain surface both with respect to FreeSurfer and a manual standard (Mindboggle) (Figure 3.4). Surprisingly, the spherical CNN (green) reaches the lowest DSC across all datasets with an average DSC of 0.76 and is already outperformed by the simple spherical parameterization approach (pCNN in light blue) despite the non-linear distortions induced by the latitude/colatitude parameterization. Finally, the view aggregation approach (dark blue, p^{3}CNN) further increases the segmentation accuracy and reaches the highest DSC for all datasets (all above 0.9 with respect to FreeSurfer, with an average of 0.93). The p^{3}CNN additionally improves the consistency of the segmentation accuracy. The variation across subjects is much lower when using view aggregation compared to the single view network. For both, the average and minimal DSC p^{3}CNN stays within the same range (0.03 to 0.08) whereas the pCNN is less consistent (0.06 to 0.21). Possibly, errors introduced by unequal sampling at the pole and equator regions are compensated by the inclusion of information from the other two views in which the structures might be more evenly sampled (different local attention).

The results show that spherical CNNs are not yet capable of outperforming parameterization approaches for cortical segmentation tasks despite the promise of a non-distorted operating space. Comparison of novel spherical or geometric CNN approaches should therefore not only include

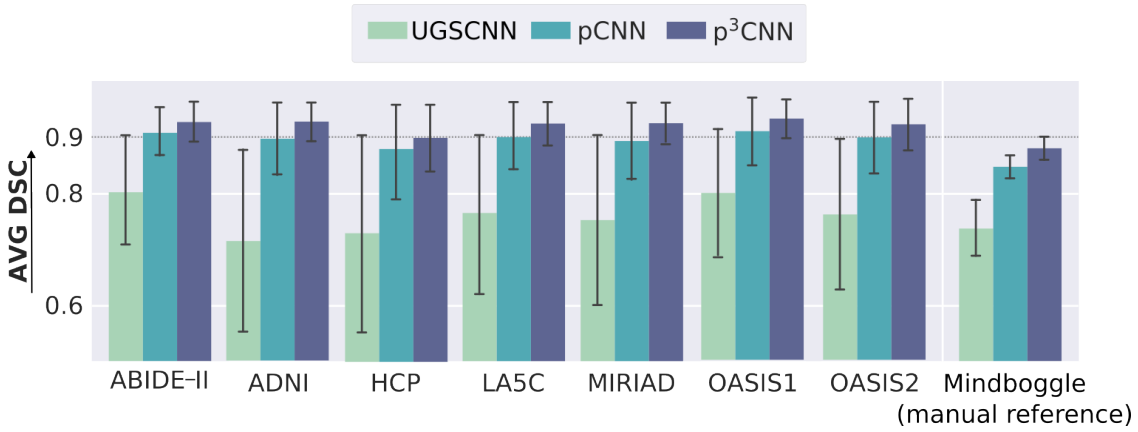

Figure 3.4: Segmentation performance improves with parameter space CNN with view aggregation (p^3CNN). The spherical CNN (UGSCNN, 1st bar) reaches the lowest DSC across 7 datasets with respect to FreeSurfer as a reference as well as a manual standard (Mindboggle). The single view parameterization approach (pCNN, 2nd bar) improves the average Dice Similarity Coefficient to around 0.9. The view aggregation across three grid rotations (p^3CNN, 3rd bar) reaches the best result across all datasets.

existing geometric methods but more importantly view-aggregating 2D segmentation networks in the spherical parameter space as a baseline.

3.3 Extension

The work presented in the following section covers extensions to the research idea, which have not been published. For the parameterization approach, the view aggregation influence is analyzed in more detail and an alternative flattening approach, the diamond projection [176], is contrasted to the chosen polar parameterization. For the unstructured grid approach, modifications of the underlying UGSCNN architecture are ablatively evaluated. This includes the replacement of the residual network blocks with dense connections and a new spherical pooling approach.

The model training and evaluation relies on the same brain MRI datasets (LA5C [67], ADNI [61], MIRIAD [83], OASIS [22, 64], ABIDE-II [72]). The training and validation set contain 160 and 25 participants, respectively. The ablative evaluations are based on a testing set of 48 participants. All sets are balanced with regard to gender, age, diagnosis, and MR field strength. FreeSurfer [94] annotations of the cortical regions according to the Desikan–Killiany–Tourville (DKT) protocol atlas [14, 15] represent the ground truth. Additionally, manual labels for 101 participants from Mindboggle [14] serve as an independent test set.

For consistency, all set-ups have the same architecture, i.e. four encoding-decoding layers, a frequency-weighted cross-entropy loss function, and an equal number (64) and dimension (3×3) of convolutional kernels. During training, a batch size of 16 and an initial learning rate of 0.01 is used. The learning rate is reduced every 20 epochs with $\gamma = 0.9$. Under these settings, the parameterization approaches use one NVIDIA V100 GPU while the UGSCNN versions require eight.

3.3.1 The Influence of Multi-View Aggregation

The first extension to the paper evaluates how an increase in the number of aggregated views changes the segmentation performance. To this end, the pole positions are rotated such that the resulting pole axis is equally inclined to the original coordinate axes (x, y, and z).

Given the unit vectors of $\vec{v}_x$, $\vec{v}_y$, and $\vec{v}_z$

$$\hat{v}_i = \frac{\vec{v}_i}{|\vec{v}_i|} \tag{3.2}$$

and considering that the direction cosines have to be identical between the resulting vector $\vec{a}$ and the unit vectors, the dot product of both have to be identical:

$$\hat{v}_i \cdot \vec{a} = \left|\hat{v}_i\right| |\vec{a}| \cos\theta_i \Rightarrow \cos\theta_i = \frac{\hat{v}_i \cdot \vec{a}}{\left|\hat{v}_i\right| |\vec{a}|} \Rightarrow \cos\theta = \frac{\hat{v}_i \cdot \vec{a}}{|\vec{a}|} \Rightarrow \hat{v}_x \cdot \vec{a} = \hat{v}_y \cdot \vec{a} = \hat{v}_z \cdot \vec{a} \tag{3.3}$$

Considering the dual basis property and given that the magnitude of the vector can be neglected, the final vector will be proportional to the sum of the cross products of the unit vectors.

$$\vec{a} \propto \hat{v}_x \times \hat{v}_y + \hat{v}_y \times \hat{v}_z + \hat{v}_z \times \hat{v}_x \tag{3.4}$$

Combinations of base axis x with y/-y, and z/-z then result in four new north pole to center axis definitions. By rotating the original x-axis towards the respective new positions, four new grid orientations are created. By repetition of the process with each newly generated base system, the number of axes positions increases by $4 * 3^n$ in each iteration.

As visible in Figure 3.5, the average DSC improves with each additional view. The segmentation accuracy specifically benefits from the first additions (view 3 and view 4) which increase the average DSC by 2.5-3 % and reduce variations across subjects. With four views, all segmentations reach an

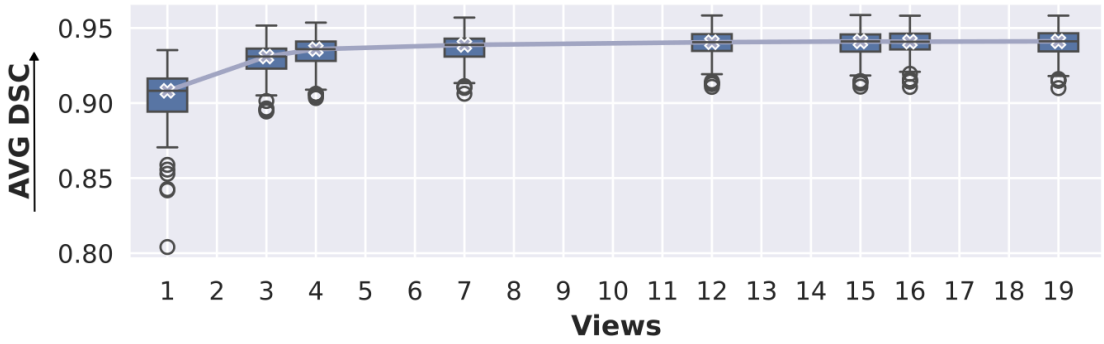

Figure 3.5: Segmentation performance improves with additional views for parameter space CNN with view aggregation (pCNN to p^{19}CNN). Starting from the single view pCNN and adding more view options (p^xCNN with x=number of views) increases segmentation performance measured via Average Dice Similarity Coefficient (AVG DSC) across 32 cortical structures and reduces inter-subject variations. The median values (marked with a blue x on the boxplots) improve up to 12 views and afterward run into saturation. Overall, the p^{12}CNN reaches an improvement by 3.6 % compared to pCNN.

average DSC of above 0.9. The view aggregation further runs into a saturation with no significant improvement above 12 views resulting in a median DSC of 0.9388. Compared to the single view pCNN network, this represents an improvement by 3.6 % DSC.

3.3.2 Spherical Flattening via Diamond Projection

Additionally, an alternative spherical flattening approach called diamond projection is evaluated [176]. This parameterization promises fewer distortions in regions with high latitudes when compared to the conventional Mercator or polar projection.

Instead of mapping full edges to the poles, a diamond shape is used where the two sub-triangles represent one-half of the sphere each. In this case, the spherical representation is mapped to α, β instead of (φ, θ) with α' and β' forming a diamond region with

$$|\alpha'| + |\beta'| \leq 1, -1 \leq |\beta'| < 1 \tag{3.5}$$

.

Rotating the diamond region clockwise by 45 degrees constructs a rectangle with $|\alpha| \leq 1$, $|\beta| \leq 1$ and represents a parameterization space with $\alpha = \alpha' cos(\pi/4) - \beta' sin(\pi/4)$ and $\beta = \alpha' sin(\pi/4) + \beta' cos(\pi/4)$. This then defines θ and φ as:

$$\theta := \frac{\pi * \alpha'}{1 - |\beta'|}, \quad \varphi := \frac{\pi * \beta'}{2} \tag{3.6}$$

The mapping to Cartesian coordinates is identical to the latitude/co-latitude transformation described in Section 3.2.3. The padding is adjusted such that the border values are taken from the adjacent side (i.e. considering the rectangle, the left and bottom sides form a pair as well as the top and right sides). Figure 3.6 presents the 2D image representation from both mappings at the top (diamond left, polar right) and a comparison of the two methods with respect to the average DSC across five different datasets at the bottom. While promising less distortion at the poles, the diamond projection (first box) does not outperform the polar projection (second box). The average DSC is consistently smaller, with a difference between 0.79 % DSC (on MIRIAD) and 0.05 % DSC (on ABIDE-II). Both approaches are, however, highly accurate with a DSC of above 0.92 on all datasets. Specifically, the polar projection reaches a DSC of 0.935 for LA5C, 0.937 for ADNI and MIRIAD, 0.939 for OASIS1 and ABIDE-II, and 0.938 for OASIS2. As the diamond projection does not improve performance, the polar projection remains the method of choice for the pCNN approach.

3.3.3 A Case for Spherical CNNs - Network Improvements

To not only push the performance of the parameterization approaches but also investigate if the spherical CNN can be further improved, the UGSCNN is extended twofold: (i) through the introduction of dense connections (+ DenseBlock) and (ii) by a novel average spherical pooling mechanism (+ Pool).

Dense Encoder Blocks

In the original UGSCNN [174], the convolutional operations are embedded in residual blocks. Here, the extension to a densely connected spherical CNN is described. In brief, the network comprises a

sequence of four dense encoder and decoder blocks separated by a bottleneck layer and connected by skip connections [130]. The dense blocks consist of three convolutional layers, each preceded by a batch-normalization layer and followed by a ReLU. In between the convolution operations, the feature maps of the preceding convolutions are concatenated to create short-range skip connections within the network [123].

Structured Pooling Spherical CNN

Encoder-decoder-like architectures require coarsening of the underlying grid structure to effectively capture a broader receptive field and reduce feature map sizes. In standard FCNNs, average or max-pooling is the method of choice. The following block introduces a structured pooling spherical CNN (+ Pool) by proposing a novel average pooling strategy based on the triangle mesh of the refined icosahedron in addition to the aforementioned dense skip connections. First, the vertex-based signals are mapped to the triangles by averaging the values at the three associated vertices (Fig. 3.7, 1). In the next step, all four triangles belonging to the same super-triangle at the lower resolution are averaged to obtain the value of the super-triangle (Fig. 3.7, 2). Finally, this signal is mapped back to the vertices at the lower resolution (Fig. 3.7, 3), weighting the influence of each triangle on the vertices by the area it covers on the sphere to account for slight size imbalances.

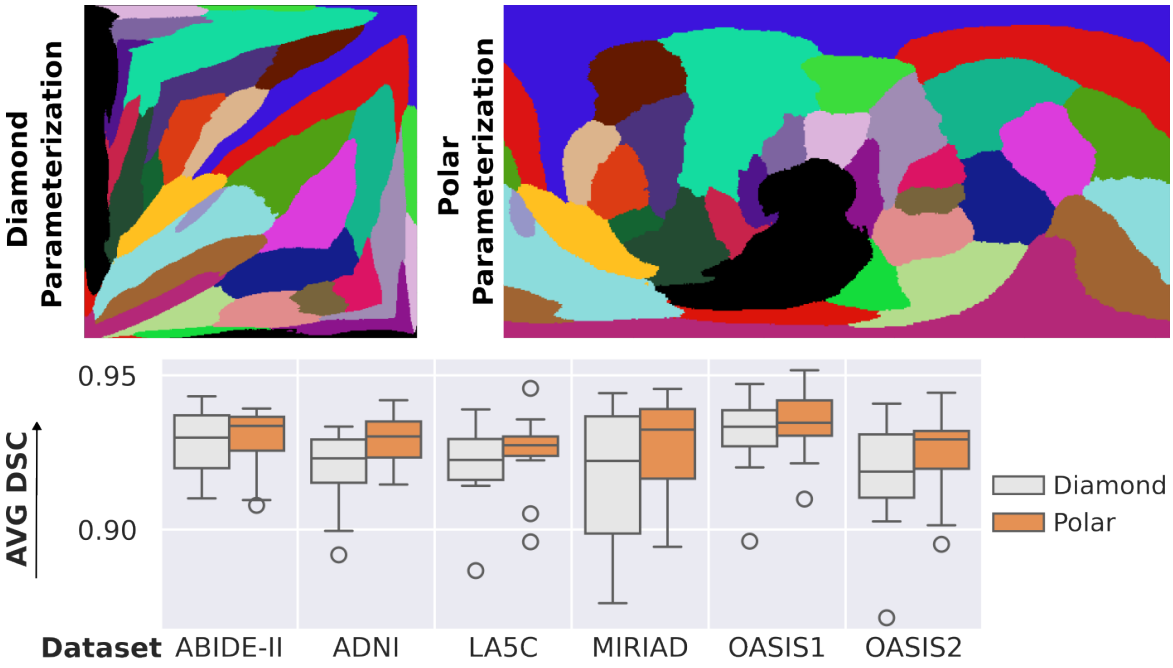

Figure 3.6: Spherical flattening via the diamond projection (top left) versus polar projection (top right). Across five different datasets, segmentation performance measured by DSC is inferior for the diamond projection (first box). Polar projection (second box) reaches the highest values with a DSC of 0.934 on average.

Experimental Results

Figure 3.8 represents the average DSC across all 32 cortical regions evaluated on five different datasets (ABIDE-II, ADNI, LA5C, OASIS1, and OASIS2) with respect to FreeSurfer as a reference, and on the manual labels of Mindboggle. The proposed modifications to the UGSCNN improve segmentation accuracy by a significant margin. Replacing residual units with dense blocks increases the average DSC by 5 % with the largest improvement on ADNI (9.6 %) and MIRIAD (9.1 %). Segmentation performance on the manual labels improves by 4.25 %. The introduction of the average spherical pooling (+ Pool) has the most profound effect with an improvement of the DSC to > 0.9 for all datasets except Mindboggle, where a DSC of 0.86 is reached. Overall, performance improves by more than 10 % across all datasets, with the largest improvement for ADNI (25.5 %) and the smallest for OASIS1 (10.1 %). Combining both modifications (+ DenseBlock + Pool) signifies the best method with a final DSC of around 0.92 for ABIDE-II, ADNI, LA5C, and MIRIAD, 0.93 for OASIS1, 0.91 for OASIS2, and 0.87 for Mindboggle. Compared to the parameterization approaches, the optimized

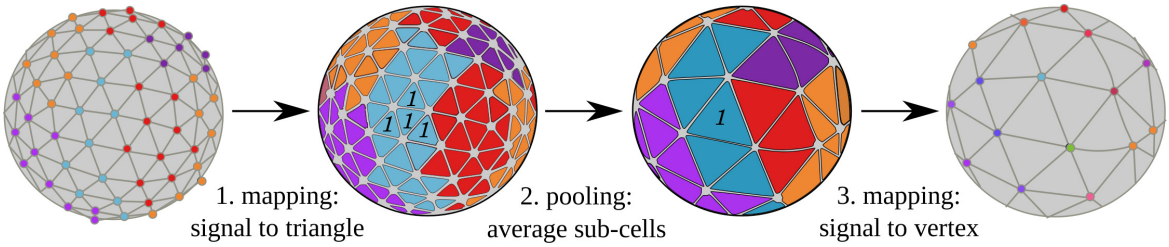

Figure 3.7: Average spherical pooling. Signals are mapped from vertices to triangles (1), sub-cells are pooled (2), and the final signal transitioned back to the vertices (3).

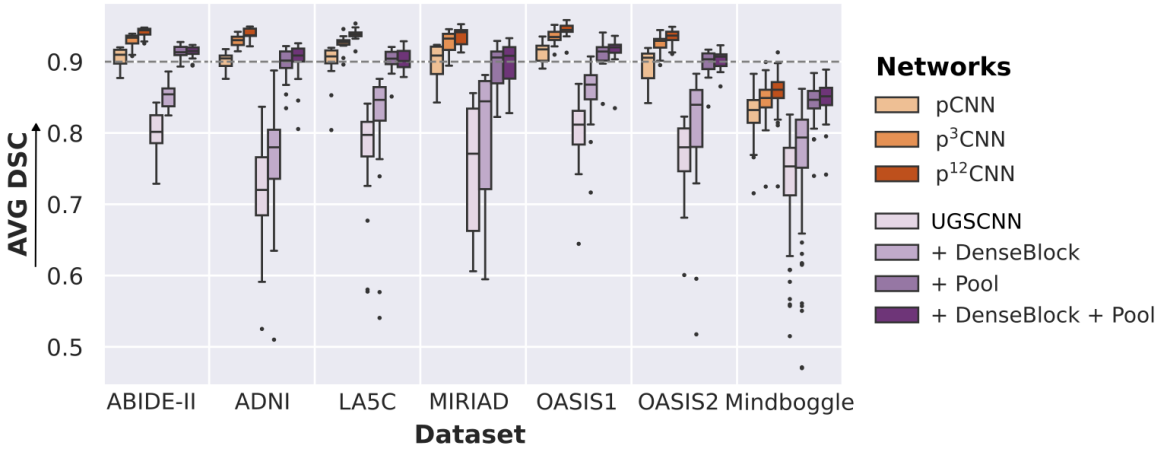

Figure 3.8: Dense blocks and spherical pooling improve segmentation performance of UGSCNN to above 0.9 on five datasets with respect to FreeSurfer and 0.85 with respect to manual labels (Mindboggle). While the original UGSCNN (fourth box) is not even close to the parameterization approaches (pCNN, p^3CNN and p^{12}CNN; first to third box), the addition of dense blocks (fifth box) and pooling (sixth box) improve performance. The combination of both ablations (seventh box) increases the DSC to a similar level as the parameterization approach.

UGSCNN slightly outperforms the pCNN on all datasets except for LA5C. A DSC increase by 1.32 % on Mindboggle represents the biggest improvement, followed by ABIDE-II with 0.93 %. The view aggregation approaches (p^3CNN and p^{12}CNN) show a higher segmentation performance. The p^{12}CNN reaches a 1.5 to 3.5 % larger DSC compared to the best UGSCNN (+DenseBlock + Pool). The results indicate that improvement for spherical deep learning models is possible. Yet, the segmentation performance on the cortical surface is still slightly inferior to the view-aggregating networks.

CHAPTER 4

Resolution-Independence in Deep Learning

The following chapter is based on the publication [3]:

> **FastSurferVINN: Building resolution-independence into deep learning segmentation methods-A solution for HighRes brain MRI**
> Authors: Leonie Henschel*, David Kügler*, and Martin Reuter, *=co-first authors
> Published in: *NeuroImage*, volume 251:118933, 2022.
> DOI: 10.1016/j.neuroimage.2022.118933

4.1 Individual Contribution

David Kügler and I jointly conceptualized the voxel-size independent neural network and developed the underlying methodology. I contributed the idea for the simultaneous inclusion of multiple resolutions during training to leverage the advantages of low- and high-resolution data. The initial idea to shift the scaling augmentation into the network was proposed by David Kügler. He also implemented the first version of the resolution-normalization layer. I extended the concept by exploring different interpolation kernels, positions of the layer within the network, alternating skip connections, a variable Gaussian noise augmentation offset, and the addition of an adaptive attention mechanism. I also contributed the idea, implementation, and evaluation of the high-resolution loss and its components. I implemented all network architectures, extended and updated the resolution-normalization layer, and wrote the code for training and inference. I also tested all existing code components and verified the replicability of the results. All formal analyses, research experiments, and evaluations were done by me. I also wrote the original draft. The research goal and conceptualization were further discussed and fine-tuned with Martin Reuter. He provided valuable feedback, financial support, and computing resources.

4.2 Summary

4.2.1 Motivation

Leading neuroimaging studies have pushed MRI acquisition resolutions below 1.0 mm for improved structure definition and morphometry. Yet, only a few automated image analysis pipelines [88, 177–180] have been validated for high-resolution settings. Already lengthy processing times are further drastically extended for sub-millimeter scans due to the cubic voxel increase, hindering the applicability of traditional tools to large cohort studies. Efficient deep learning approaches for brain segmentation [2, 119, 120, 144–151], on the other hand, do not support native image resolution segmentations so far, and require resampling of in- and/or outputs to one fixed resolution (usually 1.0 mm). As a consequence, important structural information is lost. While dedicated high-resolution networks may seem like a viable alternative, the lack of a standard submillimeter resolution would require training a plethora of different models and raises generalizability concerns – specifically considering the still limited availability of diverse high-resolution data with sufficient coverage of scanner, age, diseases, or genetic variance.

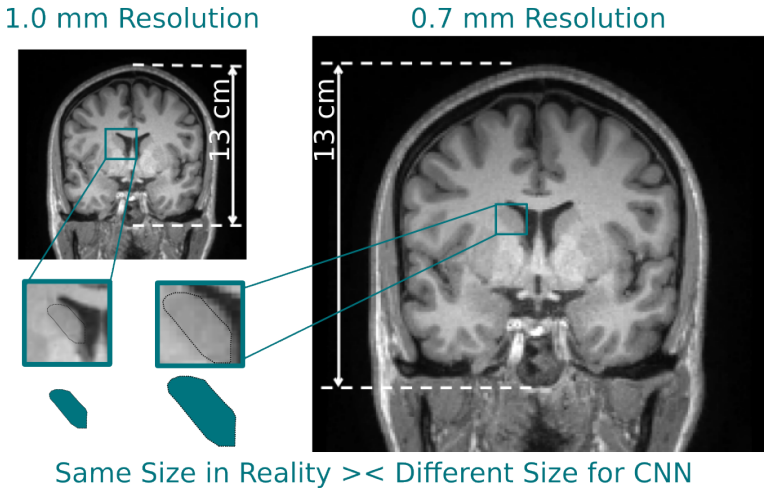

Figure 4.1: Resolution differences translate to size differences in CNNs. CNNs perceive images as a grid of uniform voxels without knowledge about the distance of grid points. Hence, high-resolution images (e.g. 0.7 mm, right) and anatomical features within appear larger compared to low-resolution images (1.0 mm, left). In reality, the head and structure sizes are identical.

Direct inclusion of multiple resolutions in the training corpus and the addition of external scale augmentation (exSA) can circumvent generalization issues and reduce potential biases from unbalanced or missing resolution data. However, as CNNs perceive images as a grid of uniform voxels without knowledge about the distance between the grid points, increasing the resolution simply translates to a larger grid (i.e. more voxels rather than smaller voxels). In consequence, anatomical structures appear enlarged as depicted in Figure 4.1. Due to this resolution-ignorance, all convolutional layers have to allocate network capacity to learn to generalize across vast scale differences which in reality do not occur. Further, exSA requires lossy nearest-neighbor interpolation of discrete label maps hence risking information loss and interpolation artifacts detrimental to high-resolution segmentation [137–140].

4.2.2 Voxel-size Independent Neural Networks

To circumvent this problem, the paper proposes the Voxel-size Independent Neural Network (VINN) which implements the concept of resolution-independence: MRI segmentation at any native resolution without external resampling during both training and inference. Leveraging the unaltered multi-resolution training corpus allows VINN to synergistically benefit from the diversity of widely available 1.0 mm datasets and superior structural details of high-resolution MRIs as depicted on the left in Figure 4.2.

VINN achieves compatibility with a range of resolutions for input and output by explicitly including the native image resolution as a separate input to the network. A flexible interpolation-based re-scaling operation within VINN, called the network-integrated resolution-normalization layer, uses the available image resolution information to calculate sampling coordinates and subsequently transforms the feature maps via an interpolation operation (Figure 4.2, green layer). This internal normalization to a fixed resolution (FixedRes, orange in Figure 4.2) allows VINN to disentangle perceived voxel versus actual structure size differences. Compared to CNNs + exSA, more resources are available to focus on structural details at a common metric scale. By replacing the first layer fixed-factor pooling and unpooling operations normally used in U-Net-like architecture to transition between scales advantageously places the resolution-normalization layer at a position where information loss naturally occurs (down- or up-scaling via pooling). In addition, interpolating feature maps of vectors instead of single scalar image slices extends the available contextual information before the interpolation step and may subsequently support a smoother transition between resolutions (multi- (yellow) and fixed-resolution (orange) in Figure 4.2).

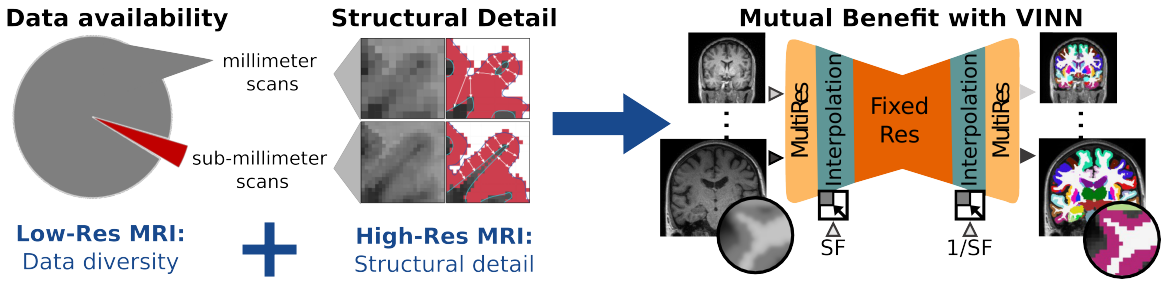

Figure 4.2: Synergistic usage of low- and high-resolution datasets with VINN. Low-resolution images exist for a diverse number of scanner types, nationalities, diseases, or age groups while high-resolution images provide rich, structural detail. The resolution-independent VINN allows segmentation of any resolution both during training and inference and therefore mutually benefits from the opposing advantages across resolutions. A network-integrated resolution-normalization layer (green) relies on the supplied scaling factor SF, the quotient of image resolution, and the inner network resolution (FixedRes, orange), to allow the transition to and from the multi-resolution block (yellow). Subsequently, predictions always remain at the native image resolution.

The sampling calculation relies only on the scale factor SF: the quotient of the resolution information of the inner normalized scale $\text{Res}_{\text{fixed}}$, a tune-able hyperparameter set to 0.8 mm throughout the experiments, and the input image $\text{Res}_{\text{native}}$. The addition of parameter α sampled from a Gaussian distribution slightly augments the scale factor SF ($SF = \text{Res}_{\text{fixed}}/\text{Res}_{\text{native}} + \alpha$), introduces small resolution variations to the sampling, and increases the robustness of the latent space interpolation. Specifically, the presence of alpha allows for augmentations at the actual anatomical rather than the voxel size. The interpolation itself is performed by applying a sampling kernel to the input map

$U \in \mathbb{R}^{H_{\text{native}} \times W_{\text{native}} \times C}$ to retrieve the value at a particular pixel in the output map $V \in \mathbb{R}^{H_{\text{fixed}} \times W_{\text{fixed}} \times C}$. At the end of the decoder, the operation is reversed using the inverse scale factor $1/SF$, returning the feature maps to the native image resolution. Identical sampling for each channel C further preserves spatial consistency. Different interpolation strategies can be defined based on the selection of the sampling kernel. Theoretically, any kernel with definable (sub-)gradients is applicable. Within the paper, the bi-linear sampling kernel is elicited as the best-performing method.

In the application to whole brain segmentation, the trained version of VINN is referred to as FastSurferVINN. In addition to VINN, respective baseline models are introduced for single-resolution and multi-resolution networks with and without exSA, referred to as FastSurferCNN* and FastSurferCNN* + exSa, respectively. Neither has previously been implemented or compared for submillimeter whole brain segmentation and represents a valuable contribution on its own. For maximal fairness, the training of all networks occurs under equal conditions (network parameters, datasets, and hardware).

FastSurferVINN consistently outperforms state-of-the-art scale augmentations across nine independent test sets by a significant margin ($p < 0.01$, Wilcoxon signed-rank test [181]). A final DSC of 88.64 for the subcortical and 88.23 for the cortical structures is reached. Improvements are consistent for submillimeter (89.62 DSC, 0.25 mm ASD) and 1.0 mm scans (86.67 DSC, 0.296 mm ASD). The comparison to single-resolution networks shows that FastSurferVINN does not only benefit high-resolution but also competitively reaches upper-bound performance in cases where a large and diverse training corpus does exist. Further, performance can be pushed by increasing the training corpus even when creating an imbalance between low- and high-resolutions. Expanding the 1.0 mm component by factor 20 does not introduce a resolution bias, but rather improves segmentation performance across all resolutions (0.7 to 1.0 mm).

FastSurferVINN outperforms data augmentation approaches most clearly in inter- and extrapolation scenarios, achieving the best results for a variety of unseen submillimeter resolutions. The detailed cortical structures targeted with high-resolution acquisition show a significant performance increase with a final DSC of 87.50 on 0.7 mm, 88.28 on 0.8 mm and 84.97 on 0.9 mm scans, respectively. The ASD improves by 19.7 %, 10 %, and 11 % compared to traditional scale augmentation.

One potential factor promoting the improved segmentation metrics with FastSurferVINN is the circumvention of nearest-neighbor label interpolation for discrete segmentation maps. Nearest-neighbor kernels are known to create jagged segmentation maps and potentially reduce important structural details [137–140]. Within the paper, the nearest-neighbor interpolation kernels systematically affected segmentation performance negatively. Usage of nearest-neighbor kernels within the resolution normalization step leads to a reduction in performance by approximately 2 % DSC and 10 % ASD compared to bi-linear, bi-cubic, or area kernels. Similarly, on the subcortical structures, the addition of exSA to FastSurferVINN deteriorates performance to a similar level as FastSurferCNN* + exSA. Overall, nearest-neighbor seems to have a systematic negative effect on segmentation performance and should, hence, be avoided.

4.2.3 High-resolution Composite Loss Function

In addition to the VINN, the paper introduces a high-resolution specific loss function focusing on areas strongly affected by PVE and compares it to an adaptive attention mechanism [182]. A strong PVE complicates the delineation of tissue borders and specifically affects narrow gyri and sulci [88, 89]. The level of detail in high-resolution images reduces PVE and improves segmentations and derived

thickness measures (see Figure 4.2, Structural Detail). Compared to the whole brain volume, however, variations in tissue and border assignment represent marginal changes in the loss, questioning accurate capture during network training. To emphasize the effect of wrong predictions within narrow WM strands, deep sulci, and tissue boundaries, a dedicated weight mask is added into the composite loss function of logistic loss and Dice loss [154]. A binary closing operation on the GM labels defines the WM strand and deep sulci mask ($\omega_{\text{WM/Sulci}}$). Erosion of the brain mask accentuates pixels at the cortex boundary and creates the outer gray matter mask (ω_{GM}). The addition of ω_{GM} and $\omega_{\text{WM/Sulci}}$ to the loss overall aims to shift the underlying decision boundary to closely mimic the target segmentation in PVE-affected locations.

Training FastSurferVINN with the new high-resolution loss function significantly improves segmentation performance on the cortical structures (corrected $p < 10^{-5}$) with a final DSC of 89.3 and an ASD of 0.209 mm. Consistent with expectations, segmentation accuracy on the subcortical structures does not change significantly. The adaptable attention mechanism, on the other hand, does not improve segmentation performance on neither the cortex nor the subcortical structures if the total number of parameters of the network with and without attention is equal. This indicates that network capacity (i.e. the number of learnable parameters) rather than adaptive attention is the critical factor here.

4.2.4 Conclusion

FastSurferVINN is a fast, voxel-size independent neural network that scales well to large datasets and enables seamless integration of a variety of resolutions, both during training and inference. In turn, both low- and high-resolution scans synergistically benefit from each other's favorable properties, i.e. dataset diversity and superior structural information. Compared to scaling augmentations, the VINN architecture avoids lossy label interpolation, features an internal resolution-normalization to disentangle perceived anatomical versus voxel size variations, and supports smooth transitions between resolutions while assuring spatial consistency. Overall, FastSurferVINN promises improved generalization to future high-resolution datasets without retraining, reduces possible dataset biases, and curtails necessary labor and time-intensive manual labeling efforts. Furthermore, the VINN concept represents an efficient resolution-independent segmentation method for wider application.

CHAPTER 5

Identifying and combating bias in segmentation networks by leveraging multiple resolutions

The following chapter is based on the publication [6] (reproduced with permission from Springer Nature):

> **Identifying and combating bias in segmentation networks by leveraging multiple resolutions**
> Authors: Leonie Henschel, David Kügler, Derek S. Andrews, Christine W. Nordahl, and Martin Reuter
> Published in: *Medical Image Computing and Computer Assisted Intervention (MICCAI), Lecture Notes in Computer Science*, vol 13435. Springer, 2022.
> DOI: 10.1007/978-3-031-16443-9_34

5.1 Individual Contribution

I formulated the original research question and devised the interpolation methods, models, as well as validation metrics for the CNNs and VINN. In joint discussions with David Kügler and Martin Reuter, I developed the goal and aim of the paper. I assembled appropriate datasets to allow visualization of the resolution biases and implemented the models, training, inference, and validation scripts. I conducted all the research experiments and wrote the original paper draft.

5.2 Summary

5.2.1 Motivation

In deep learning, limited data availability clashes with the assumption of representative training sets and raises concerns about the generalizability of trained networks. Investigating potential biases arising from this situation is of uttermost importance in medical settings where transparency and known applicability limitations are pivotal. Despite this importance, exploration of bias in deep learning for neuroanatomical segmentation is virtually nonexistent. This work, for the first time,

evaluates how heterogeneous resolutions between groups impact deep learning networks for brain segmentation.

5.2.2 Effect of Image Resampling, Scale Augmentation, and Resolution-Independence in Multi-Resolution Settings

In this scenario, two groups with known differences (cortical volume in adults versus children, hippocampal volume in controls versus patients with Alzheimer's Disease or mild cognitive impairment) are chosen and their MRIs are considered to exist at only one resolution during training (high-resolution (Group H) or low-resolution (Group L)). Inference on the actual high-resolution scans of Group L allows assessment of the resolution-derived segmentation bias.

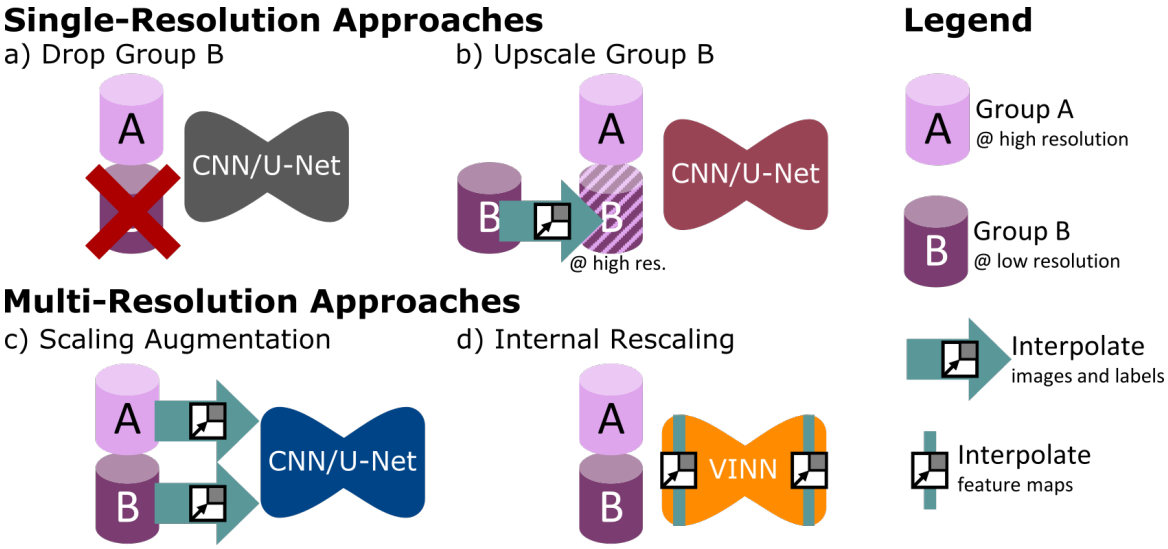

Figure 5.1: Four approaches to harmonize the resolution-biased datasets for the segmentation network: a) drop the lower-resolution data, b) upsample the low-resolution dataset to the higher resolution, c) randomly rescale both datasets as augmentation, and d) incorporate a resampling to the higher resolution into the network. Figure and caption adapted from [6]. Reproduced with permission from Springer Nature.

Single-resolution networks dropping the lower-resolution Group L (Figure 5.1, a), or trained on resampled images (Figure 5.1, b) are contrasted to multi-resolution networks trained with either external scaling augmentation (Figure 5.1, c) or internal resampling (aka VINN [3]) (Figure 5.1, d). All approaches are treated equally with identical architecture (number of parameters, layers, blocks), dataset, training, and hardware parameters.

Combining a robust signed volume measure catching systematic errors such as over- or undersegmentation and DSC to assess overall accuracy allows bias estimation. An unbiased network should capture group differences by keeping the modes between the distributions separated and the median values close to the manual labels. The DSC should ideally be high for both, Group H and Group L, and the signed volume error should be close to a median value of zero, indicating neither over- nor undersegmentation. Calculation of the Area-under-the-Curve (AUC) for predicting the group assignment (children or adults) based on thresholding the gray matter volume investigates how effects in segmentation performance influence downstream tasks.

5.2.3 Single-Resolution Networks

In the evaluation, single-resolution approaches, unfortunately, collapse to a uniform distribution and do not reach good accuracy on high-resolution scans for group L, present only at low-resolution during training. Further, single-resolution approaches consistently underestimate the volumes for group L. The detected volume bias also propagates to classification tasks, where the distinction between adults and children based on cortical volume is impaired. These effects are most pronounced for approaches where training occurs only in the high-resolution group (out-of-distribution effect). While resampling generally improves performance, it initiates a bias when shifting to lower resolutions.

5.2.4 Multi-Resolution Networks

Multi-resolution approaches, on the other hand, accurately reproduce the modes of volume distribution across different resolution settings, reaching a DSC above 0.9 and a median volume error around zero across all evaluated resolutions and groups. Note, that the networks only encountered low-resolution images of group L during training. Hence, learned high-resolution information from group H seems to translate effectively to group L during inference by including multi-resolution information directly in the training process. Subsequent classification tasks based on the prediction with the multi-resolution networks allow an effective distinction between adults and children with an AUC above 0.9 across all resolutions.

5.2.5 Conclusion

Single-resolution networks work well on their training distribution but fail to generalize group differences across resolutions. On the other hand, scale augmentation in CNNs as well as alternative architectures with build-in resolution independence like VINN can help reduce the observed segmentation biases. These multi-resolution networks effectively improve information transfer between high- and low-resolution images. Overall, processing high-resolution images with single-resolution deep learning based analysis pipelines should be avoided. The findings are highly relevant considering future high-resolution image acquisitions in neuroimaging but also relate to other fields dealing with resolution differences.

CHAPTER 6

VINNA for Neonates - Orientation Independence through Latent Augmentations

The following chapter is based on the publication [4]:

> **VINNA for Neonates - Orientation Independence through Latent Augmentations**
> Authors: Leonie Henschel, David Kügler, Lilla Zollei, and Martin Reuter
> Published in: *Imaging Neuroscience*, vol 2:1-26, 2024
> DOI: 10.1162/imag_a_00180

6.1 Individual Contribution

I formulated the initial idea to extend the scale invariance of the VINN architecture to rotation and translations to address elevated levels of head motions within the neonate cohort. I discussed the evolution and development of the methodology with David Kügler, Lilla Zöllei, and Martin Reuter. All software and testing pipelines were written by me. I formally analyzed, validated, and evaluated the experiments, and conducted the research and investigation process. I wrote the original paper draft, including all visualizations.

6.2 Summary

6.2.1 Motivation

Whole brain segmentation of young age groups is a challenging task and substantially different from adults (Figure 6.1). Tools developed for adults fail due to the substantial differences in head and structure sizes as well as contrast differences [167]. The neonate cohort (20-40 weeks post gestation), which is the focus of this paper, is characterized by large variations of head positions due to the scanning conditions and overall smaller anatomy. Padding is often used to stabilize the child's head and to occupy the space between the head coil and the participant which can not easily be standardized. Further, no standard resolution exists across available cohorts. Additionally, the number of available datasets is extremely limited. As an example, only two cohort studies are publicly available for the

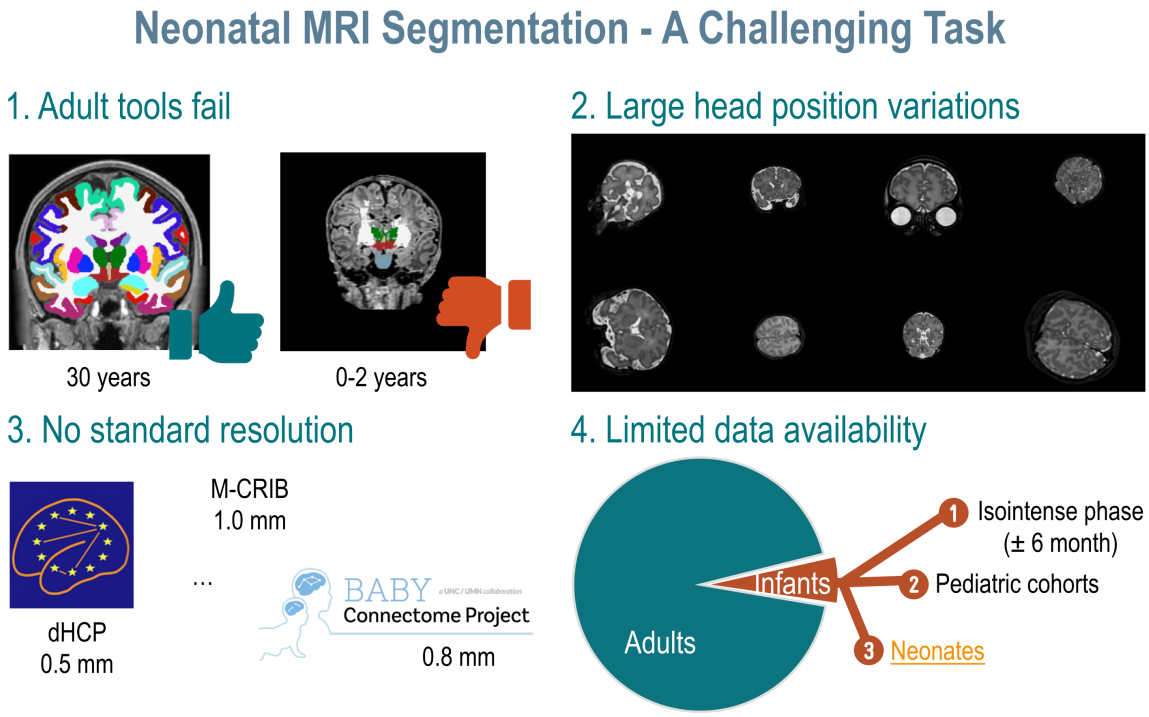

Figure 6.1: Challenges in neonatal MRI segmentation. First, tools developed for adults fail on neonates due to the substantial anatomical differences and contrast changes. Second, the neonate cohort is characterized by a large variation of head positions. Third, no standard resolution exists across available cohorts. Fourth, data availability is low and substantially less diverse than for adults.

newborn phase: the dHCP [20] and the Baby Connectome Project (BCP) [59]. Corresponding labeled data is even more difficult to find and diverse with respect to the label definitions, partly due to the still ongoing changes within the developing brain. Of the two datasets mentioned above, only the dHCP provides reference segmentations. Overall, specific tools able to cope with these challenges are needed.

A few pipelines for neonate segmentation already exist (e.g. iBEAT [183], infantFS [184], or the dhcp-pipeline [20]). These traditionally rely on time-consuming non-linear spatial registration procedures to address variations in spatial positioning of the heads and resample images to a common resolution. These interpolations can result in loss of information, specifically for down-sampling operations. Furthermore, registration is particularly challenging for newborns due to lower tissue contrast and a common cause of errors.

Deep learning tools can provide an alternative. However, considering the limited data availability for newborns, the variability in resolution and head positioning existing in the cohort is likely underrepresented in the publicly available datasets, raising generalizability concerns. Some form of rotation and scaling equivariance is required, i.e. the segmentation needs to stay the same independent of the orientation and size. Classic CNNs are only invariant to translations and do not inherently address resolution and head variations. In training, external augmentations can approximate equivariance by exposing the network to different resolutions and positions. The VINN introduced in Chapter 4 uses internal scaling to normalize resolutions within the network itself [3]. However, to address the

increased variability of head positions, external augmentation is still needed. Hence, CNNs and VINNs require interpolation of both the intensity image and the associated ground truth, which can cause a loss of information specifically for small structures.

6.2.2 Network-Integrated Internal Augmentations

The paper introduces a new framework called VINNA – a voxel-size independent network with internal augmentations. VINNA features a network-integrated four degree of freedom (4-DOF) transform module that extends the resolution-normalization layer introduced in Chapter 4 by spatial augmentations, i.e. translation and rotation transformation. In essence, augmentations commonly applied externally on the intensity image and label map are shifted into the network.

At the first layer scale transition, the multi-dimensional feature maps are flexibly rescaled and randomly transformed to mimic position variations frequently encountered in newborns. The network receives a transformation matrix M with four degrees of freedom as an additional input parameter, encoding scaling, in-plane rotation, and translation. Akin to VINN, the scaling factor normalizes the images to a common fixed base resolution within the network. The randomly sampled rotation and translation parameters diversify the training distribution and facilitate the approximation of equivariance with respect to position changes by the network. The interpolation-based affine per-channel mapping $M : \mathcal{U} \rightarrow \mathcal{V}$ samples the feature maps to the respective new position based on M. In the end, the inverse affine transformation $M^{-1} : \mathcal{V} \rightarrow \mathcal{U}$ restores the original image input resolution and position.

While the implementation might seem straightforward, it represents a major paradigm shift (external versus internal augmentation) that has not been covered in the relevant literature so far. Akin to VINN, a priori knowledge about the image, such as the resolution, can be passed to the network via the transformation matrix M. The internal interpolation of continuous feature maps further avoids any type of lossy label interpolation. Hence, VINNA accommodates various resolutions and head positions while avoiding any type of label resampling.

6.2.3 Internal versus External Augmentations - VINNA as a Solution for Newborn MRI Segmentation

Ablative optimization under equal architecture, training, hardware, and dataset conditions, shows improved performance of VINNA compared to CNNs and VINNs with external augmentations. The 4-DOF transform module reduces loss of information directly translating to better segmentation results. Across three different resolutions and two modalities, VINNA achieves the highest DSC (95.33 on average), as well as the lowest ASD (0.102 mm on average). Specifically, the ASD at the high-resolution benefits from the new module with an improvement of 4.47 % on the cortex, 5.19 % on the WM, and 5.89 % on the subcortical structures compared to VINN + exA.

As state-of-the-art methods do not exist out of the box for infant segmentation, the nnU-Net [134] in 2D and 3D configuration with external augmentation is retrained under identical conditions as VINNA. Consistent with the ablation, results for DSC and ASD improve with the 4-DOF module, indicating the superiority of the internal augmentation scheme. Compared to the 3D nnU-Net + exA, ASD and DSC scores are significantly improved with the highest gain on the cortical structures (56 %, 35 %, and 45 % ASD and 3.8 %, 2.7 %, and 3.5 % DSC for 0.5 mm, 0.8 mm and 1.0 mm, respectively).

Finally, comparing VINNA to traditional tools iBEAT and infantFS also shows a closer emulation of the ground truth on the dHCP [20] cohort. It should be noted that this cross-protocol comparison includes atlas differences (i.e. label definitions differ and require mapping/masking to reach a common space) which likely introduces additional unquantifiable errors. While results represent an as-fair-as-possible comparison, results should be interpreted with the caveat in mind. A qualitative evaluation corroborates the accuracy results and shows that VINNA preserves structural details better. Small structures lost in models trained with external augmentation or traditional methods are recovered.

6.2.4 Conclusion

With VINNA, we provide the first fully automated neonatal brain MRI segmentation into 89 classes in under 1 minute. VINNA's 4-DOF module shifts spatial augmentations into the network thereby supporting segmentation of diverse head positions at the native resolution. Due to the network-integrated interpolation of continuous values, VINNA avoids label resampling and effectively regularizes the latent space. Both features result in VINNA outperforming external augmentation and better preservation of structural details. Implementing transformation operations for image and label pairs that are truly equivariant can be difficult. Hypothetically, the loss of information introduced by external augmentation due to image interpolation (lossy interpolation of the label map and image) is larger than previously assumed. The internal augmentations reduce this information loss by interpolating continuous values in multiple feature maps. This approach, therefore, holds potential benefits for areas outside newborn segmentation.

CHAPTER 7

Conclusion and Outlook

Non-invasive neuroimaging studies collect a myriad of imaging data across different populations and cohorts to advance our understanding of healthy development, disease effects, and age-related changes. The extraction and quantification of the underlying structural changes in the brain in terms of different morphometric measures (e.g. volume or thickness) require time-efficient, accurate, objective, and reproducible segmentation workflows. Yet, existing traditional neuroimaging pipelines suffer from extensive processing times which limits scalability. Faster, deep learning based segmentation algorithms, on the other hand, lack estimation of reliability and sensitivity to real neuroanatomical effects as well as utilities such as the creation of cortical surface models or support for different image resolutions. The translation of the tools into practice is therefore limited. The central contributions of this thesis address these concerns and provide well-validated, open-source next-generation deep learning solutions for the automated analysis of structural human brain MRI.

7.1 FastSurfer - The Deep Learning Based Neuroimaging Pipeline

The presented FastSurfer [2] is the first and to this date only framework integrating a CNN for brain MRI segmentation into a self-contained surface pipeline mimicking the traditional tool FreeSurfer [94]. FastSurferCNN, FastSurfer's volume-based segmentation part, creates a volumetric whole-brain segmentation into 95 classes in less than 1 minute. The introduction of local and global competition by replacement of concatenations with the efficient maxout operation [169] in the dense encoder and decoder blocks, as well as multi-slice information aggregation, specifically tailor performance towards accurate segmentation of both cortical and subcortical structures. Subsequent reconstruction of surfaces with FastSurfer's surface-based processing stream RECON-SURF creates a FreeSurfer-compatible output structure in around 60 minutes. While processing time is significantly reduced compared to traditional tools, FastSurfer maintains high levels of accuracy and generalizes across different data distributions. Sensitivity to disease effects as well as reliability of the segmentation is on par with the state-of-the-art FreeSurfer pipeline.

In contrast to other deep learning solutions for brain segmentation [119, 120, 134, 143–151], FastSurfer's outputs are not restricted to segmentations but include surface models of the cortex, as well as point- and ROI-wise thickness estimates. Consequently, seamless integration into existing tools for subsequent analysis tasks such as registration and statistical frameworks for group comparison [94,

103], relation of structural and functional data [185], or white matter tract extraction [186] becomes possible.

The developed deep learning network FastSurferCNN is on its own a highly competitive baseline and has been proven useful in several tasks such as cerebellum segmentation [187], early Alzheimer's Disease detection [188], or hippocampal segmentation in stroke patients [189]. Its predictions also show less numerical uncertainty than traditional image-processing results suggesting better reproducibility across execution environments [190]. Despite small numerical perturbations applied to all floating-point operations, edges of subcortical structures are drawn consistently with FastSurferCNN. FreeSurfer exhibits higher variability, probably caused by the use of numerical optimizations [190]. Thus, FastSurferCNN is a valuable tool to avoid reproducibility issues and subsequently improves trust in segmentation-derived measures and conclusions.

The code for FastSurfer and all its components is openly available on GitHub (`https://github.com/deep-mi/fastsurfer`) and actively used, as evidenced by more than 390 stars, 165 issues, 200 monthly clones, 7k views, and 4.3k downloads of the accompanying Docker containers (`https://hub.docker.com/r/deepmi/fastsurfer`). Taken together, FastSurfer has been well-received by the community and is on its way to becoming a major player in neuroimaging analysis.

7.2 Parameter Space CNN for Cortical Surface Segmentation

The p^3CNN [5] is an efficient network for cortical surface segmentation using a polar parameterization approach to create 2D image-like representations of spherical signals. The view aggregation across different pole axis orientations is a useful tool to alleviate distortions introduced by the non-isometric mapping and subsequently reduces prediction errors. The final p^3CNN emulates the ground truth better than the spherical segmentation network UGSCNN [174], the only available geometric deep learning network for segmentation at the time of publication. While the performance of the UGSCNN is inferior, subsequent optimization of the spherical network design via DenseBlocks and an efficient pooling operation increases segmentation accuracy and highlights the potential for improvement. In theory, approaches acting directly on the mesh representation should outperform parameterization approaches. First, the brain is a highly folded, convoluted structure, and mesh representations stay true to the underlying anatomy [175]. In traditional pipelines like FreeSurfer, analysis of signals on the surface itself has been shown to improve reliability and detectability of effects [161–166]. Second, distortions induced by imperfect mappings between the sphere and the plane can be completely avoided. Third, in contrast to spherical CNNs, the polar projection will result in different visualizations depending on the orientation of the sphere. For the task of cortical surface segmentation, this point has less effect as the head positions during scanning are relatively consistent. The resulting spherical representations of the cortex generated with FreeSurfer are therefore in a similar orientation as well. Hence, the projected image-like representations are congruent across participants and the missing rotation equivariance property has less impact on the performance. The p^3CNN is, therefore, probably best equipped for tasks where assumed orientation information is crucial and a pre-alignment naturally exists (or can be created).

Overall, geometric deep learning was (and still is) in its infancy, and methods have predominantly been developed for object classification, mainly focusing on optimization of the convolution operations [191–194]. Their high computational cost in combination with the complexity of the cortex and the correspondingly large number of vertices further limits their applicability in the neuroimaging domain.

Spherical CNNs for cortical segmentation therefore still predominantly rely on the discretization of the sphere to the icosahedron for scale transitions and efficient processing [195, 196]. This, in turn, requires resampling of the original signal and may introduce errors. Newer work for cortical parcellation relies on graph attention modules to directly operate on the mesh representation without resampling to a spherical representation [197]. The authors report an accuracy of 89.00 on the Mindboggle data which is the second-best result after the Spherical Deformable U-Net, another recently introduced spherical CNN [195]. Interestingly, the evaluations in both papers do not include parameterization approaches, and a comparison of the p^3CNN to the newer architectures is an interesting direction for future work.

7.3 VINN - Resolution-Independence in Deep Learning

The VINN is the first brain segmentation tool for resolution-independent brain MRI processing. Combining and leveraging both the increased structural information from submillimeter 3T brain scans and the generalizability to intensity, scanner, disease, and other variations from standard 1.0 mm MRIs, benefits both high- and low-resolution datasets. The data imbalance problem in high-resolution images is circumvented while structural information translates to the lower resolutions. In contrast to scaling augmentations, the network-integrated resolution-normalization layer explicitly uses a priori knowledge about image resolution to internally transition between scales. Shifting the scaling into the network architecture disentangles perceived voxel volume changes from anatomically relevant size differences. Consequently, network capacity becomes available for different tasks, improving segmentation results. Further, VINN avoids lossy nearest-neighbor interpolation and operates on continuous multi-dimensional feature maps as opposed to image slices. Hence, neighborhood information is available and allows a smooth transition while assuring spatial consistency. Compared to external scaling augmentation, segmentation accuracy consequently improves across all evaluated resolutions, including domains outside the training corpus.

After the publication, FastSurferVINN has been integrated into the FastSurfer pipeline. Therefore, high-resolution images can now be analyzed end-to-end from the original intensity image, over a highly accurate whole-brain segmentation into 95 structures, and up to reconstructed cortical surface models and morphological measures. The high-resolution adaptation of the FastSurfer pipeline including integration of the FastSurferVINN framework was officially released with version v2.0.0 and is openly available on the aforementioned Github repository.

7.4 Biases in Multi-Resolution Settings

The assessment of biases in medical imaging is of utmost importance. Considering the rise of CNNs for the automated analysis of MRI in research studies, the limitations and preconceptions introduced by the training data have to be known to avoid biased findings. Undiscovered segmentation biases in neural networks may heavily impact group separations and subsequent diagnosis or treatment decisions. The results presented in Chapter 5 show that the VINN architecture is highly effective in combating methodological biases in multi-resolution settings. Structural information translates well across the resolution border and VINN consistently achieves good results on both, low- and high-resolution cohorts. External augmentation can sustain the same effect. Fixed, single-resolution networks, on the other hand, fail to generalize across groups with different resolutions. Importantly, this bias translates to downstream tasks. While no other paper has so far assessed bias in multi-resolution settings,

studies have been conducted on demographic biases (sex and race) [198–200] as well as motion bias [201]. The congruent findings demonstrate the importance of appropriately balanced training sets: disparity in the distributions translates to significant bias effects in segmentation model performance. A key advantage of VINN is its ability to include datasets independent of their resolution, mitigating potential data imbalance problems.

While the results presented here are based on artificially constructed examples, they do have relevance for ongoing research projects. As an example, several initiatives currently collect data on the African continent to alleviate demographic imbalances in neuroimaging collections [202, 203]. Due to the limited availability of high-field scanners, the MRI are predominantly recorded with lower resolutions [202, 203]. Enriching training sets of high-resolution MRI from Caucasian cohorts with these scans would result in a scenario identical to the simulated conditions: two groups at different resolutions. Despite the demographic balance, training under the single-resolution paradigm could therefore still translate to biased predictions when applied in practice (e.g. scanning of an individual of the respective group at a higher or lower resolution). Overall, processing with multi-resolution deep learning based analysis pipelines should be the preferred mode of operation.

7.5 VINNA - Internal versus External Augmentation in Newborn Segmentation

VINNA presents the first network for native-resolution neonatal brain MRI segmentation including subcortical segmentation, GM and WM parcellation. The large variability of head morphometry and resolution non-uniformity encountered in the newborn cohort are addressed via a network-integrated 4-DOF transform module. This module extends the resolution-normalization layer introduced in VINN by spatial augmentations, i.e. translation and rotation transformation. At the first layer scale transition, the multi-dimensional feature maps are flexibly rescaled and randomly transformed to diversify the training distribution. Shifting the spatial augmentations into the network allows VINNA to outperform state-of-the-art external augmentation approaches in CNN, VINN, and nnU-Net [134] on the dHCP cohort. Metric evaluation combined with qualitative inspection indicates that the internal 4-DOF transform module in VINNA retains high-level details across all resolutions and age groups. The presented module is also easy to integrate into other network architectures and might prove useful in different areas dealing with strong orientation variations.

Currently, VINNA is limited to volume-based analysis. The surface reconstruction in adults is distinct from newborns due to the large differences in contrast and predominant focus on T1w MRI. FastSurfer's RECON-SURF pipeline is therefore not applicable to this cohort. Adaptation of infantFreeSurfer (infantFS) [184] is a potential solution and interesting avenue for future work. This adaptation does, however, come with several non-trivial challenges. infantFS accepts only T1w images at a resolution of 1.0 mm. In newborns, the T2w MRIs provides better contrast between GM and WM. Subsequently, infantFS often fails to accurately capture the cortical surface in younger age groups [4, 183]. Additionally, T1w images are not always available. A complete pipeline would therefore require a transformation approach to create good contrast T1w images. Alternatively, direct mesh reconstruction from the generated segmentation with the marching-cube algorithm [204] could provide an intensity-independent approach. Similarly, deep learning approaches for surface reconstruction like Topofit [205] or Vox2Cortex [206] may prove useful to create a modality-independent pipeline. It should, however, be noted that these approaches have been developed for adults only. Transfer to

newborns is not trivial [167] and may prove more difficult than expected.

Bibliography

[1] J. Stiles and T. L. Jernigan, *The basics of brain development*, Neuropsychology review **20** (2010) 327–348 (cit. on pp. 1, 2).

[2] L. Henschel et al., *FastSurfer - A fast and accurate deep learning based neuroimaging pipeline*, NeuroImage **219** (2020) 117012 (cit. on pp. 1, 11, 17, 21, 38, 51).

[3] L. Henschel, D. Kügler and M. Reuter, *FastSurferVINN: Building resolution-independence into deep learning segmentation methods—A solution for HighRes brain MRI*, NeuroImage **251** (2022) 118933 (cit. on pp. 1, 8, 11, 37, 44, 48).

[4] L. Henschel, D. Kügler, L. Zöllei and M. Reuter, *VINNA for Neonates - Orientation Independence through Latent Augmentations*, Imaging Neuroscience **2** (2024) 1–26 (cit. on pp. 1, 47, 54).

[5] L. Henschel and M. Reuter, *Parameter Space CNN for Cortical Surface Segmentation*, Informatik Aktuell (2020) 216–221, ed. by T. Tolxdorff et al. (cit. on pp. 1, 27, 28, 52).

[6] L. Henschel, D. Kügler, D. S. Andrews, C. W. Nordahl and M. Reuter, *Identifying and Combating Bias in Segmentation Networks by Leveraging Multiple Resolutions*, Medical Image Computing and Computer Assisted Intervention - MICCAI 2022, Lecture Notes in Computer Science **13435** (2022) 350–359, ed. by L. Wang, Q. Dou, P. T. Fletcher, S. Speidel and S. Li (cit. on pp. 1, 43, 44).

[7] E. R. Kandel et al., *Principles of Neuroscience*, McGraw-hill New York, 2012, chap. Cell and Molecular Biology of Cells of the Nervous System 133–165 (cit. on p. 2).

[8] S. Ackerman, *Discovering the Brain*, ed. by I. of Medicine and N. A. of Sciences, Washington, DC: The National Academies Press, 1992, ISBN: 978-0-309-46799-5 (cit. on p. 2).

[9] T. Bui and J. M. Das, *Neuroanatomy, cerebral hemisphere*, en, StatPearls Publishing, 2023 (cit. on p. 2).

[10] P. Chauhan, A. Rathawa, K. Jethwa, S. Mehra et al., *The anatomy of the cerebral cortex*, Exon Publications (2021) 1–16 (cit. on p. 2).

[11] B. Fischl and A. Dale, *Measuring the thickness of the human cerebral cortex from magnetic resonance images*, Proceedings of the National Academy of Sciences of the United States of America **97** (2000) 11050–11055 (cit. on p. 2).

[12] E. R. Kandel et al., *Principles of Neuroscience*, McGraw-hill New York, 2012, chap. The cells of the Nervous system 133–165 (cit. on p. 2).

[13] Q. R. Morell P, *Basic Neurochemistry: Molecular, Cellular and Medical Aspects*, 6th, Lippincott-Raven, 1999, chap. The Myelin Sheath (cit. on p. 2).

[14] A. Klein and J. Tourville, *101 Labeled Brain Images and a Consistent Human Cortical Labeling Protocol*, Frontiers in Neuroscience **6** (2012) 171 (cit. on pp. 2, 8, 10, 22, 31).

[15] R. S. Desikan et al., *An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest*, NeuroImage **31** (2006) 968–980 (cit. on pp. 2, 10, 22, 31).

[16] C. Destrieux, B. Fischl, A. Dale and E. Halgren, *Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature*, NeuroImage **53** (2010) 1–15 (cit. on pp. 2, 10).

[17] D. N. Kennedy et al., *Gyri of the human neocortex: an MRI-based analysis of volume and variance*, Cerebral Cortex **8** (1998) 372–384 (cit. on p. 2).

[18] P. M. Thompson, C. Schwartz, R. T. Lin, A. A. Khan and A. W. Toga, *Three-dimensional statistical analysis of sulcal variability in the human brain*, Journal of Neuroscience **16** (1996) 4261–4274 (cit. on p. 2).

[19] P. M. Thompson et al., *Genetic influences on brain structure*, Nature neuroscience **4** (2001) 1253–1258 (cit. on p. 2).

[20] A. Makropoulos et al., *The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction*, NeuroImage **173** (2018) 88–112 (cit. on pp. 3, 4, 6, 48, 50).

[21] D. C. Van Essen et al., *The Human Connectome Project: a data acquisition perspective*, NeuroImage **62** (2012) 2222–2231 (cit. on pp. 3, 4, 6, 18, 24).

[22] D. S. Marcus et al., *Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults*, Journal of cognitive neuroscience **19** (2007) 1498–1507 (cit. on pp. 3, 6, 25, 31).

[23] K. Hayakawa, Y. Konishi, M. Kuriyama, K. Konishi and T. Matsuda, *Normal brain maturation in MRI*, European Journal of Radiology **12** (1991) 208–215 (cit. on p. 2).

[24] R. Dietrich and W. Bradley Jr, *Normal and abnormal white matter maturation*, Seminars in Ultrasound, CT, and MR **9** (1988) 192–200 (cit. on p. 2).

[25] A. Barkovich, B. Kjos, D. Jackson Jr and D. Norman, *Normal maturation of the neonatal and infant brain: MR imaging at 1.5 T*, Radiology **166** (1988) 173–180 (cit. on p. 2).

[26] K. L. Mills et al., *Structural brain development between childhood and adulthood: Convergence across four longitudinal samples*, NeuroImage **141** (2016) 273–281 (cit. on p. 2).

[27] A. M. Hedman, N. E. van Haren, H. G. Schnack, R. S. Kahn and H. E. Hulshoff Pol, *Human brain changes across the life span: A review of 56 longitudinal magnetic resonance imaging studies*, Human Brain Mapping **33** (2012) 1987–2002 (cit. on p. 2).

[28] S. Frangou et al., *Cortical thickness across the lifespan: Data from 17,075 healthy individuals aged 3–90 years*, Human Brain Mapping **43** (2022) 431–451 (cit. on p. 2).

[29] H. Braak and E. Braak, *Staging of alzheimer's disease-related neurofibrillary changes*, Neurobiology of Aging **16** (1995) 271–278 (cit. on p. 2).

[30] Y. Grignon, C. Duyckaerts, M. Bennecib and J.-J. Hauw, *Cytoarchitectonic alterations in the supramarginal gyrus of late onset Alzheimer's disease*, Acta neuropathologica **95** (1998) 395–406 (cit. on p. 2).

[31] O. Querbes et al., *Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve*, Brain **132** (2009) 2036–2047 (cit. on pp. 2, 3).

[32] E. B. Johnson et al., *The impact of occipital lobe cortical thickness on cognitive task performance: An investigation in Huntington's Disease*, Neuropsychologia **79** (2015) 138–146 (cit. on p. 2).

[33] H. Rosas et al., *Regional and progressive thinning of the cortical ribbon in Huntington's disease*, Neurology **58** (2002) 695–701 (cit. on p. 2).

[34] G. Halliday et al., *Regional specificity of brain atrophy in Huntington's disease*, Experimental neurology **154** (1998) 663–672 (cit. on p. 2).

[35] L. Nanetti et al., *Cortical thickness, stance control, and arithmetic skill: An exploratory study in premanifest Huntington disease*, Parkinsonism & Related Disorders **51** (2018) 17–23 (cit. on p. 2).

[36] N. E. Van Haren et al., *Changes in cortical thickness during the course of illness in schizophrenia*, Archives of general psychiatry **68** (2011) 871–880 (cit. on p. 2).

[37] Y. Zhao et al., *Cortical thickness abnormalities at different stages of the illness course in schizophrenia: a systematic review and meta-analysis*, JAMA psychiatry **79** (2022) 560–570 (cit. on p. 2).

[38] L. M. Rimol et al., *Cortical thickness and subcortical volumes in schizophrenia and bipolar disorder*, Biological psychiatry **68** (2010) 41–50 (cit. on p. 2).

[39] C. M. Wannan et al., *Evidence for network-based cortical thickness reductions in schizophrenia*, American Journal of Psychiatry **176** (2019) 552–563 (cit. on p. 2).

[40] J. Baron et al., *In Vivo Mapping of Gray Matter Loss with Voxel-Based Morphometry in Mild Alzheimer's Disease*, NeuroImage **14** (2001) 298–309 (cit. on p. 2).

[41] J. A. Kaye et al., *Volume loss of the hippocampus and temporal lobe in healthy elderly persons destined to develop dementia*, Neurology **48** (1997) 1297–1304 (cit. on p. 2).

[42] A.-T. Du et al., *Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia*, Brain **130** (2007) 1159–1166 (cit. on p. 2).

[43] M. Zarei et al.,
Cortical thinning is associated with disease stages and dementia in Parkinson's disease,
Journal of Neurology, Neurosurgery & Psychiatry **84** (2013) 875–882 (cit. on p. 2).

[44] G. Wenk, *Neuropathologic Changes in Alzheimer's Disease*,
The Journal of clinical psychiatry **64** (2003) 7–10 (cit. on p. 2).

[45] J. P. Lerch et al., *Focal Decline of Cortical Thickness in Alzheimer's Disease Identified by Computational Neuroanatomy*, Cerebral Cortex **15** (2004) 995–1001 (cit. on p. 2).

[46] S. P. Poulin, R. Dautoff, J. C. Morris, L. F. Barrett and B. C. Dickerson,
Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity,
Psychiatry Research: Neuroimaging **194** (2011) 7–13 (cit. on pp. 2, 3).

[47] C. R. Jack et al.,
Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease,
Neurology **49** (1997) 786–794 (cit. on p. 2).

[48] Y. Noh et al.,
Anatomical heterogeneity of Alzheimer disease: based on cortical thickness on MRIs,
Neurology **83** (2014) 1936–1944 (cit. on p. 2).

[49] M. S. Albert, *Cognitive and neurobiologic markers of early Alzheimer disease*,
Proceedings of the National Academy of Sciences **93** (1996) 13547–13551 (cit. on p. 2).

[50] K. Double et al., *Topography of brain atrophy during normal aging and Alzheimer's disease*,
Neurobiology of aging **17** (1996) 513–521 (cit. on p. 2).

[51] L. W. de Jong et al.,
Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study,
Brain **131** (2008) 3277–3285 (cit. on p. 3).

[52] W. Henneman et al., *Hippocampal atrophy rates in Alzheimer disease*,
Neurology **72** (2009) 999–1007 (cit. on p. 3).

[53] N. Schuff et al., *MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers*, Brain **132** (2009) 1067–1077 (cit. on p. 3).

[54] J. P. Aggleton, A. Pralus, A. J. D. Nelson and M. Hornberger,
Thalamic pathology and memory loss in early Alzheimer's disease: moving the focus from the medial temporal lobe to Papez circuit, Brain **139** (2016) 1877–1890 (cit. on p. 3).

[55] L. Pini et al., *Brain atrophy in Alzheimer's Disease and aging*,
Ageing Research Reviews **30** (2016) 25–48, Brain Imaging and Aging (cit. on p. 3).

[56] M. Lehmann et al., *Cortical thickness and voxel-based morphometry in posterior cortical atrophy and typical Alzheimer's disease*, Neurobiology of Aging **32** (2011) 1466–1476
(cit. on p. 3).

[57] M. R. Sabuncu et al.,
The Dynamics of Cortical and Hippocampal Atrophy in Alzheimer Disease,
Archives of Neurology **68** (2011) 1040–1048 (cit. on p. 3).

[58] K. Im et al., *Variations in cortical thickness with dementia severity in Alzheimer's disease*, Neuroscience Letters **436** (2008) 227–231 (cit. on p. 3).

[59] B. R. Howell et al., *The UNC/UMN Baby Connectome Project (BCP): An overview of the study design and protocol development*, NeuroImage **185** (2019) 891–905 (cit. on pp. 3, 6, 48).

[60] N. D. Volkow, J. A. Gordon and M. P. Freund, *The Healthy Brain and Child Development Study—Shedding Light on Opioid Exposure, COVID-19, and Health Disparities*, JAMA Psychiatry **78** (2021) 471–472 (cit. on pp. 3, 6).

[61] S. G. Mueller et al., *Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI)*, Alzheimer's & Dementia **1** (2005) 55–66 (cit. on pp. 3, 6, 24, 31).

[62] C. J. Weber et al., *The Worldwide Alzheimer's Disease Neuroimaging Initiative: ADNI-3 updates and global perspectives*, Alzheimer's & Dementia: Translational Research & Clinical Interventions **7** (2021) e12226 (cit. on p. 3).

[63] M. W. Weiner et al., *Increasing participant diversity in AD research: Plans for digital screening, blood testing, and a community-engaged approach in the Alzheimer's Disease Neuroimaging Initiative 4*, Alzheimer's & Dementia **19** (2023) 307–317 (cit. on p. 3).

[64] D. S. Marcus, A. F. Fotenos, J. G. Csernansky, J. C. Morris and R. L. Buckner, *Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults*, Journal of cognitive neuroscience **22** (2010) 2677–2684 (cit. on pp. 3, 6, 31).

[65] P. J. LaMontagne et al., *P3-083: OASIS-3: Longitudinal Neuroimaging, clinical and cognitive dataset for normal aging and Alzheimer's Disease*, Alzheimer's & Dementia **14** (2018) P1097–P1097 (cit. on p. 3).

[66] L. N. Koenig et al., *Select Atrophied Regions in Alzheimer disease (SARA): An improved volumetric model for identifying Alzheimer disease dementia*, NeuroImage: Clinical **26** (2020) 102248 (cit. on p. 3).

[67] R. A. Poldrack et al., *A phenome-wide examination of neural and cognitive function*, Scientific data **3** (2016) 160110 (cit. on pp. 4, 6, 31).

[68] S. C. Tanaka et al., *A multi-site, multi-disorder resting-state magnetic resonance image database*, Scientific data **8** (2021) 227 (cit. on p. 4).

[69] J. R. Bustillo et al., *Glutamatergic and neuronal dysfunction in gray and white matter: a spectroscopic imaging study in a large schizophrenia sample*, Schizophrenia bulletin **43** (2017) 611–619 (cit. on p. 4).

[70] D. Chyzhyk, A. Savio and M. Graña, *Computer aided diagnosis of schizophrenia on resting state fMRI data by ensembles of ELM*, Neural Networks **68** (2015) 23–33 (cit. on p. 4).

[71] K. Alpert, A. Kogan, T. Parrish, D. Marcus and L. Wang,
The northwestern university neuroimaging data archive (NUNDA),
NeuroImage **124** (2016) 1131–1136 (cit. on p. 4).

[72] A. Di Martino et al., *Enhancing studies of the connectome in autism using the autism brain imaging data exchange II*, Scientific data **4** (2017) 170010 (cit. on pp. 4, 6, 31).

[73] A. Di Martino et al., *The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism*, Mol Psychiatry **19** (2013) 659–67 (cit. on p. 4).

[74] L. Schmaal et al., *ENIGMA MDD: seven years of global neuroimaging studies of major depression through worldwide data sharing*, Translational psychiatry **10** (2020) 172 (cit. on p. 4).

[75] M. M. Breteler, T. Stöcker, E. Pracht, D. Brenner and R. Stirnberg,
MRI in the Rhineland study: a novel protocol for population neuroimaging,
Alzheimer's & Dementia: The Journal of the Alzheimer's Association **10** (2014) P92 (cit. on pp. 4, 7, 10, 18).

[76] C. Sudlow et al., *UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age*, PLOS Medicine **12** (2015) 1–10 (cit. on p. 4).

[77] L. H. Somerville et al., *The Lifespan Human Connectome Project in Development: A large-scale study of brain connectivity development in 5–21 year olds*,
NeuroImage **183** (2018) 456–468 (cit. on p. 4).

[78] S. Y. Bookheimer et al., *The Lifespan Human Connectome Project in Aging: An overview*,
NeuroImage **185** (2019) 335–348 (cit. on p. 4).

[79] H. Kabasawa, *MR imaging in the 21st century: technical innovation over the first two decades*,
Magnetic Resonance in Medical Sciences **21** (2022) 71–82 (cit. on p. 4).

[80] V. M. Runge and J. T. Heverhagen, *The Physics of Clinical MR Taught Through Images*,
Cham: Springer International Publishing, 2022, ISBN: 978-3-030-85413-3 (cit. on pp. 4–6).

[81] R. Bitar et al.,
MR Pulse Sequences: What Every Radiologist Wants to Know but Is Afraid to Ask,
RadioGraphics **26** (2006) 513–537 (cit. on pp. 4, 6).

[82] H. Schild, *MRI Made Easy (...well Almost)*, Schering-Diagnostics, Schering AG, 1990,
ISBN: 9783921817414 (cit. on pp. 4–6).

[83] I. B. Malone et al.,
MIRIAD—Public release of a multiple time point Alzheimer's MR imaging dataset,
NeuroImage **70** (2013) 33–36 (cit. on pp. 6, 31).

[84] R. T. Constable and R. M. Henkelman, *Contrast, resolution, and detectability in MR imaging*,
Journal of computer assisted tomography **15** (1991) 297–303 (cit. on p. 6).

[85] W. Edelstein, G. Glover, C. Hardy and R. Redington,
The intrinsic signal-to-noise ratio in NMR imaging,
Magnetic resonance in medicine **3** (1986) 604–618 (cit. on p. 6).

[86] D. L. Parker and G. T. Gullberg, *Signal-to-noise efficiency in magnetic resonance imaging*, Medical Physics **17** (1990) 250–257 (cit. on p. 6).

[87] K. Setsompop, D. A. Feinberg and J. R. Polimeni, *Rapid brain MRI acquisition techniques at ultra-high fields*, NMR in biomedicine **29** (2016) 1198–1221 (cit. on p. 7).

[88] N. Zaretskaya, B. Fischl, M. Reuter, V. Renvall and J. R. Polimeni, *Advantages of cortical surface reconstruction using submillimeter 7 T MEMPRAGE*, NeuroImage **165** (2018) 11–26 (cit. on pp. 7, 38, 40).

[89] M. F. Glasser et al., *The minimal preprocessing pipelines for the Human Connectome Project*, NeuroImage **80** (2013) 105–124 (cit. on pp. 7, 40).

[90] F. Luesebrink, A. Wollrab and O. Speck, *Cortical thickness determination of the human brain using high resolution 3T and 7T MRI data*, NeuroImage **70** (2013) 122–131 (cit. on p. 7).

[91] M. Ono, S. Kubik and C. D. Abernathey, *Atlas of the cerebral sulci*, G. Thieme Verlag ; Thieme Medical Publishers, 1990, ISBN: 9780865773622 (cit. on p. 8).

[92] M. Petrides, *The human cerebral cortex: an MRI atlas of the sulci and gyri in MNI stereotaxic space*, 1st ed, Elsevier Academic, 2012, ISBN: 9780123869388 (cit. on p. 8).

[93] M. L. McHugh, *Interrater reliability: the kappa statistic*, Biochemia medica **22** (2012) 276–282 (cit. on p. 8).

[94] B. Fischl, *FreeSurfer*, NeuroImage **62** (2012) 774–781 (cit. on pp. 8–10, 18, 22, 31, 51).

[95] A. M. Dale, B. Fischl and M. I. Sereno, *Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction*, NeuroImage **9** (1999) 179–194 (cit. on pp. 8, 9).

[96] B. Fischl et al., *Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain*, Neuron **33** (2002) 341–355 (cit. on pp. 8, 9).

[97] J. Talairach et al., *Atlas d'anatomie stereotaxique du telencephale*, (1967) (cit. on p. 9).

[98] J. Talairach and P. nd Tournoux, *Co-planar stereotaxic atlas of the human brain*, Thieme Medical Publishers (1988) 1–122 (cit. on p. 9).

[99] B. Fischl, A. K. Liu and A. M. Dale, *Automated Manifold Surgery: Constructing Geometrically Accurate and Topologically Correct Models of the Human Cerebral Cortex*, IEEE Trans. Med. Imaging **20** (2001) 70–80 (cit. on p. 9).

[100] B. Fischl, M. I. Sereno, R. Tootell and A. Dale, *High-resolution intersubject averaging and a coordinate system for the cortical surface*, Human brain mapping **8** (1999) 272–84 (cit. on p. 9).

[101] B. Fischl et al., *Automatically Parcellating the Human Cerebral Cortex*, Cerebral cortex (New York, N.Y. : 1991) **14** (2004) 11–22 (cit. on p. 9).

[102] K. Friston, J. Ashburner, S. Kiebel, T. Nichols and W. Penny, *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, Academic Press, 2007 (cit. on pp. 10, 18, 22).

[103] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich and S. M. Smith, *FSL*, NeuroImage **62** (2012) 782–790 (cit. on pp. 10, 22, 51).

[104] J. Ashburner and K. J. Friston, *Unified segmentation*, NeuroImage **26** (2005) 839–851 (cit. on p. 10).

[105] H. Wang and P. Yushkevich, *Multi-atlas segmentation with joint label fusion and corrective learning—an open source implementation*, Frontiers in neuroinformatics **7** (2013) 27 (cit. on p. 10).

[106] J. Schmidhuber, *Deep learning in neural networks: An overview*, Neural Networks **61** (2015) 85–117 (cit. on p. 11).

[107] Y. LeCun, Y. Bengio and G. Hinton, *Deep learning*, Nature **521** (2015) 436–444 (cit. on pp. 11, 13, 14, 16).

[108] A. Krizhevsky, I. Sutskever and G. E. Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*, Commun. ACM **60** (2017) 84–90 (cit. on pp. 11, 16).

[109] W. S. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, The Bulletin of Mathematical Biophysics **5** (1943) 115–133 (cit. on p. 11).

[110] F. Rosenblatt, *The perceptron: A probabilistic model for information storage and organization in the brain*, Psychological Review **65** (1958) 386–408 (cit. on p. 11).

[111] D. Hebb, *The Organization of Behavior*, 0th ed., Psychology Press, 2005 (cit. on p. 11).

[112] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning representations by back-propagating errors*, Nature **323** (1986) 533–536 (cit. on pp. 11, 13).

[113] K. Fukushima, *Cognitron: A self-organizing multilayered neural network*, Biological Cybernetics **20** (1975) 121–136 (cit. on p. 11).

[114] V. Nair and G. E. Hinton, *Rectified linear units improve restricted boltzmann machines*, Proceedings of the 27th international conference on machine learning (ICML-10) (2010) 807–814 (cit. on p. 11).

[115] D. C. Ciresan, U. Meier, L. M. Gambardella and J. Schmidhuber, *Deep Big Simple Neural Nets For Handwritten Digit Recogntion*, Neural Computation **22** (2010) 3207–3220 (cit. on p. 11).

[116] K.-S. Oh and K. Jung, *GPU implementation of neural networks*, Pattern Recognition **37** (2004) 1311–1314 (cit. on p. 11).

[117] K. Chellapilla, S. Puri and P. Simard, *High Performance Convolutional Neural Networks for Document Processing*, International Workshop on Frontiers in Handwriting Recognition (2006) (cit. on p. 11).

[118] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella and J. Schmidhuber, *Flexible, High Performance Convolutional Neural Networks for Image Classification*, Intl. Joint Conference on Artificial Intelligence IJCAI (2011) 1237–1242 (cit. on pp. 11, 16).

[119] A. G. Roy, S. Conjeti, N. Navab, C. Wachinger, A. D. N. Initiative et al., *QuickNAT: A fully convolutional network for quick and accurate segmentation of neuroanatomy*, NeuroImage **186** (2019) 713–727 (cit. on pp. 11, 17, 18, 24, 38, 51).

[120] B. Billot, E. Robinson, A. V. Dalca and J. E. Iglesias, *Partial Volume Segmentation of Brain MRI Scans of Any Resolution and Contrast*, Medical Image Computing and Computer Assisted Intervention - MICCAI 2020, Lecture Notes in Computer Science (2020) 177–187 (cit. on pp. 11, 17, 18, 38, 51).

[121] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press, 2016, URL: http://www.deeplearningbook.org (cit. on pp. 13, 16).

[122] S. Ioffe and C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, Proceedings of the 32nd International Conference on Machine Learning, ICML **37** (2015) 448–456 (cit. on p. 14).

[123] G. Huang, Z. Liu, K. Q. Weinberger and L. van der Maaten, *Densely connected convolutional networks*, Proceedings of the IEEE conference on computer vision and pattern recognition **1** (2017) 3 (cit. on pp. 14, 22, 28, 34).

[124] Y.-L. Boureau, J. Ponce and Y. LeCun, *A theoretical analysis of feature pooling in visual recognition*, ICML 2010 - Proceedings, 27th International Conference on Machine Learning (2010) 111–118 (cit. on p. 16).

[125] J. Weng, N. Ahuja and T. Huang, *Cresceptron: A Self-Organizing Neural Network Which Grows Adaptively*, Proceedings - 1992 International Joint Conference on Neural Networks, IJCNN 1992 (1992) 576–581 (cit. on p. 16).

[126] D. Scherer, A. Müller and S. Behnke, *Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition*, Artificial Neural Networks – ICANN 2010 (2010) 92–101, ed. by K. Diamantaras, W. Duch and L. S. Iliadis (cit. on p. 16).

[127] D. Ciresan, A. Giusti, L. Gambardella and J. Schmidhuber, *Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images*, Advances in Neural Information Processing Systems **25** (2012), ed. by F. Pereira, C. Burges, L. Bottou and K. Weinberger (cit. on p. 16).

[128] C. Farabet, C. Couprie, L. Najman and Y. LeCun, *Learning Hierarchical Features for Scene Labeling*, IEEE Transactions on Pattern Analysis and Machine Intelligence **35** (2013) 1915–1929 (cit. on p. 16).

[129] E. Shelhamer, J. Long and T. Darrell, *Fully Convolutional Networks for Semantic Segmentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence **39** (2017) 640–651 (cit. on pp. 16, 17).

[130] O. Ronneberger, P. Fischer and T. Brox,
U-Net: Convolutional Networks for Biomedical Image Segmentation,
Medical Image Computing and Computer Assisted Intervention - MICCAI 2015, Lecture Notes in Computer Science **9351** (2015) 234–241 (cit. on pp. 16, 17, 22, 34).

[131] H. Noh, S. Hong and B. Han, *Learning deconvolution network for semantic segmentation*,
Proceedings of the IEEE international conference on computer vision (2015) 1520–1528 (cit. on p. 16).

[132] V. Badrinarayanan, A. Kendall and R. Cipolla,
Segnet: A deep convolutional encoder-decoder architecture for image segmentation,
IEEE transactions on pattern analysis and machine intelligence **39** (2017) 2481–2495 (cit. on p. 16).

[133] P. Simard, D. Steinkraus and J. Platt,
Best practices for convolutional neural networks applied to visual document analysis,
Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings (2003) 958–963 (cit. on pp. 16, 17).

[134] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen and K. H. Maier-Hein,
nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation,
Nature Methods **18** (2021) 203–211 (cit. on pp. 16, 18, 49, 51, 54).

[135] C. Zhang, S. Bengio, M. Hardt, B. Recht and O. Vinyals,
Understanding Deep Learning (Still) Requires Rethinking Generalization,
Commun. ACM **64** (2021) 107–115 (cit. on p. 16).

[136] E. Goceri, *Medical image data augmentation: techniques, comparisons and interpretations*,
Artificial Intelligence Review (2023) 12561–12605 (cit. on pp. 16, 17).

[137] J. A. Parker, R. V. Kenyon and D. E. Troxel,
Comparison of Interpolating Methods for Image Resampling,
IEEE Transactions on Medical Imaging **2** (1983) 31–39 (cit. on pp. 17, 38, 40).

[138] P. Thevenaz, *Image Interpolation and Resampling*,
Handbook of Medical Image Processing and Analysis (2009) 465–493 (cit. on pp. 17, 38, 40).

[139] A. Schaum, *Theory and Design of Local Interpolators*,
CVGIP: Graphical Models and Image Processing **55** (1993) 464–481 (cit. on pp. 17, 38, 40).

[140] J. P. Allebach, *7.1 - Image Scanning, Sampling, and Interpolation*,
Communications, Networking and Multimedia (2005) 895–XXVII, ed. by A. BOVIK (cit. on pp. 17, 38, 40).

[141] T. Falk et al., *U-Net: deep learning for cell counting, detection, and morphometry*,
Nature Methods **16** (2019) 67–70 (cit. on p. 17).

[142] N. Siddique, S. Paheding, C. P. Elkin and V. Devabhaktuni, *U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications*,
IEEE Access **9** (2021) 82031–82057 (cit. on p. 17).

[143] J. E. Iglesias et al., *Joint super-resolution and synthesis of 1 mm isotropic MP-RAGE volumes from clinical MRI exams with scans of different orientation, resolution and contrast*,
NeuroImage **237** (2021) 118206 (cit. on pp. 17, 18, 51).

[144] Y. Huo et al., *3D whole brain segmentation using spatially localized atlas network tiles*, NeuroImage **194** (2019) 105–119 (cit. on pp. 17, 18, 22, 38, 51).

[145] R. Mehta, A. Majumdar and J. Sivaswamy, *BrainSegNet: a convolutional neural network architecture for automated segmentation of human brain structures*, Journal of Medical Imaging **4** (2017) 1–11 (cit. on pp. 17, 18, 22, 38, 51).

[146] H. Chen, Q. Dou, L. Yu, J. Qin and P.-A. Heng, *VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images*, NeuroImage **170** (2018) 446–455 (cit. on pp. 17, 18, 38, 51).

[147] L. Sun et al., *A 3D Spatially-Weighted Network for Segmentation of Brain Tissue from MRI*, IEEE Transactions on Medical Imaging (2019) 1–1 (cit. on pp. 17, 18, 38, 51).

[148] R. Ito, K. Nakae, J. Hata, H. Okano and S. Ishii, *Semi-supervised deep learning of brain tissue segmentation*, Neural Networks **116** (2019) 25–34 (cit. on pp. 17, 18, 38, 51).

[149] P. McClure et al., *Knowing What You Know in Brain Segmentation Using Bayesian Deep Neural Networks*, Frontiers in Neuroinformatics **13** (2019) 67 (cit. on pp. 17, 18, 38, 51).

[150] P. Coupé et al., *AssemblyNet: A large ensemble of CNNs for 3D whole brain MRI segmentation*, NeuroImage **219** (2020) 117026 (cit. on pp. 17, 18, 22, 38, 51).

[151] C. Wachinger, M. Reuter and T. Klein, *DeepNAT: deep convolutional neural network for segmenting neuroanatomy*, NeuroImage **170** (2018) 434–445 (cit. on pp. 18, 22, 38, 51).

[152] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox and O. Ronneberger, *3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation*, Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016, Lecture Notes in Computer Science **9901** (2016) 424–432 (cit. on pp. 18, 24).

[153] P. Moeskops et al., *Automatic Segmentation of MR Brain Images With a Convolutional Neural Network*, IEEE Transactions on Medical Imaging **35** (2016) 1252–1261 (cit. on pp. 18, 22).

[154] A. G. Roy et al., *Error Corrective Boosting for Learning Fully Convolutional Networks with Limited Data*, Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017, Lecture Notes in Computer Science (2017) 231–239 (cit. on pp. 18, 22, 41).

[155] A. Jog, A. Hoopes, D. N. Greve, K. V. Leemput and B. Fischl, *PSACNN: Pulse sequence adaptive fast whole brain segmentation*, NeuroImage **199** (2019) 553–569 (cit. on pp. 18, 22).

[156] S. Roy, D. Kügler and M. Reuter, *Are 2.5D approaches superior to 3D deep networks in whole brain segmentation?*, Medical Imaging with Deep Learning (2022) (cit. on p. 18).

[157] R. A. Bethlehem et al., *Brain charts for the human lifespan*, Nature **604** (2022) 525–533 (cit. on p. 18).

[158] K. Kamnitsas et al., *Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation*, Medical Image Analysis **36** (2017) 61–78 (cit. on p. 18).

[159] S. Jadon, *A survey of loss functions for semantic segmentation*, 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB) (2020) 1–7 (cit. on p. 18).

[160] C. Gaser et al., *CAT – A Computational Anatomy Toolbox for the Analysis of Structural MRI Data*, bioRxiv (2022) (cit. on p. 18).

[161] M. Goto et al., *Advantages of Using Both Voxel- and Surface-based Morphometry in Cortical Morphology Analysis: A Review of Various Applications*, Magnetic Resonance in Medical Sciences **21** (2022) 41–57 (cit. on pp. 18, 52).

[162] D. N. Greve et al., *Cortical surface-based analysis reduces bias and variance in kinetic modeling of brain PET data*, NeuroImage **92** (2014) 225–236 (cit. on pp. 18, 52).

[163] S. S. Ghosh et al., *Evaluating the validity of volume-based and surface-based brain image registration for developmental cognitive neuroscience studies in children 4 to 11 years of age*, NeuroImage **53** (2010) 85–93 (cit. on pp. 18, 52).

[164] K. L. Hyde, F. Samson, A. C. Evans and L. Mottron, *Neuroanatomical differences in brain areas implicated in perceptual and other core features of autism revealed by cortical thickness analysis and voxel-based morphometry*, Human brain mapping **31** (2010) 556–566 (cit. on pp. 18, 52).

[165] A. Tessitore et al., *Cortical thickness changes in patients with Parkinson's disease and impulse control disorders*, Parkinsonism & related disorders **24** (2016) 119–125 (cit. on pp. 18, 52).

[166] J. B. Pereira et al., *Assessment of cortical degeneration in patients with Parkinson's disease by voxel-based morphometry, cortical folding, and cortical thickness*, Human brain mapping **33** (2012) 2521–2534 (cit. on pp. 18, 52).

[167] M. Korom et al., *Dear reviewers: Responses to common reviewer critiques about infant neuroimaging studies*, Developmental Cognitive Neuroscience **53** (2022) 101055 (cit. on pp. 19, 47, 55).

[168] A. de Brébisson and G. Montana, *Deep neural networks for anatomical brain segmentation*, 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2015) 20–28 (cit. on p. 22).

[169] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville and Y. Bengio, *Maxout networks*, Proceedings of the 30th International Conference on International Conference on Machine Learning-Volume 28 (2013) III–1319 (cit. on pp. 22, 51).

[170] S. Estrada, S. Conjeti, M. Ahmad, N. Navab and M. Reuter, *Competition vs. Concatenation in Skip Connections of Fully Convolutional Networks*, Machine Learning in Medical Imaging (2018) 214–222 (cit. on p. 22).

[171] S. Estrada et al., *FatSegNet : A Fully Automated Deep Learning Pipeline for Adipose Tissue Segmentation on Abdominal Dixon MRI*, Magnetic Resonance in Medicine **83** (2019) 1471–1483 (cit. on p. 22).

[172] M. Reuter, F.-E. Wolter and N. Peinecke,
Laplace-Beltrami spectra as "Shape-DNA" of surfaces and solids,
Computer-Aided Design **38** (2006) 342–366 (cit. on p. 23).

[173] M. Reuter, *Hierarchical Shape Segmentation and Registration via Topological Features of Laplace-Beltrami Eigenfunctions*,
International Journal of Computer Vision **89** (2010) 287–308 (cit. on p. 23).

[174] C. M. Jiang et al., *Spherical CNNs on Unstructured Grids*,
International Conference on Learning Representations (2019) (cit. on pp. 28, 33, 52).

[175] B. Fischl, M. I. Sereno and A. M. Dale, *Cortical Surface-Based Analysis: II: Inflation, Flattening, and a Surface-Based Coordinate System*, NeuroImage **9** (1999) 195–207
(cit. on pp. 27, 52).

[176] M. Reuter, *Laplace spectra for shape recognition*, eng,
Norderstedt: Books on Demand GmbH, 2006 160–161, ISBN: 9783833450716
(cit. on pp. 31, 33).

[177] P.-L. Bazin et al.,
A computational framework for ultra-high resolution cortical segmentation at 7Tesla,
NeuroImage **93** (2014) 201–209 (cit. on p. 38).

[178] P. A. Yushkevich et al., *Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment*,
Human Brain Mapping **36** (2014) 258–287 (cit. on p. 38).

[179] J. M. Huntenburg, C. J. Steele and P.-L. Bazin,
Nighres: processing tools for high-resolution neuroimaging, GigaScience **7** (2018)
(cit. on p. 38).

[180] C. Gaser et al.,
CAT-A Computational Anatomy Toolbox for the Analysis of Structural MRI Data,
bioRxiv (2023) (cit. on p. 38).

[181] F. Wilcoxon, *Individual Comparisons by Ranking Methods*,
Biometrics Bulletin **1** (1945) 80–83 (cit. on p. 40).

[182] Y. Qin et al., *Autofocus Layer for Semantic Segmentation*,
Medical Image Computing and Computer Assisted Intervention - MICCAI 2018, Lecture Notes in Computer Science **11072** (2018) 603–611 (cit. on p. 40).

[183] L. Wang et al., *iBEAT V2.0: a multisite-applicable, deep learning-based pipeline for infant cerebral cortical surface reconstruction*, Nature Protocols **18** (2023) 1488–1509
(cit. on pp. 48, 54).

[184] L. Zöllei, J. E. Iglesias, Y. Ou, P. E. Grant and B. Fischl,
Infant FreeSurfer: An automated segmentation and surface extraction pipeline for T1-weighted neuroimaging data of infants 0–2 years, NeuroImage **218** (2020) 116946
(cit. on pp. 48, 54).

[185] O. Esteban et al., *fMRIPrep: a robust preprocessing pipeline for functional MRI*,
Nature methods **16** (2019) 111–116 (cit. on p. 52).

[186] A. Yendiki et al., *Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy*, Frontiers in neuroinformatics **5** (2011) 23 (cit. on p. 52).

[187] J. Faber et al., *CerebNet: A fast and reliable deep-learning pipeline for detailed cerebellum sub-segmentation*, NeuroImage **264** (2022) 119703 (cit. on p. 52).

[188] L. Bloch and C. M. Friedrich, *Comparison of Automated Volume Extraction with FreeSurfer and FastSurfer for Early Alzheimer's Disease Detection with Machine Learning*, 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS) (2021) 113–118 (cit. on p. 52).

[189] M. Schell, M. Foltyn-Dumitru, M. Bendszus and P. Vollmuth, *Automated hippocampal segmentation algorithms evaluated in stroke patients*, Scientific Reports **13** (2023) 11712 (cit. on p. 52).

[190] I. G. Pepe, V. Sivakolunthu, H. L. Park, Y. Chatelain and T. Glatard, *Numerical Uncertainty of Convolutional Neural Networks Inference for Structural Brain MRI Analysis*, Uncertainty for Safe Utilization of Machine Learning in Medical Imaging (2023) 64–73, ed. by C. H. Sudre et al. (cit. on p. 52).

[191] T. Cohen, M. Weiler, B. Kicanaoglu and M. Welling, *Gauge Equivariant Convolutional Networks and the Icosahedral CNN*, Proceedings of Machine Learning Research **97** (2019) 1321–1330, ed. by K. Chaudhuri and R. Salakhutdinov (cit. on p. 52).

[192] R. Kondor, Z. Lin and S. Trivedi, *Clebsch–Gordan Nets: a Fully Fourier Space Spherical Convolutional Neural Network*, Advances in Neural Information Processing Systems **31** (2018), ed. by S. Bengio et al. (cit. on p. 52).

[193] C. Esteves, C. Allen-Blanchette, A. Makadia and K. Daniilidis, *Learning SO(3) Equivariant Representations with Spherical CNNs*, International Journal of Computer Vision **128** (2020) 588–600 (cit. on p. 52).

[194] J. Gerken et al., *Equivariance versus augmentation for spherical images*, International Conference on Machine Learning (2022) 7404–7421 (cit. on p. 52).

[195] F. Zhao et al., *Spherical deformable u-net: Application to cortical surface parcellation and development prediction*, IEEE transactions on medical imaging **40** (2021) 1217–1228 (cit. on p. 53).

[196] P. Parvathaneni et al., *Cortical Surface Parcellation Using Spherical Convolutional Neural Networks*, Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 (2019) 501–509, ed. by D. Shen et al. (cit. on p. 53).

[197] X. Li et al., *Anatomically constrained squeeze-and-excitation graph attention network for cortical surface parcellation*, Computers in Biology and Medicine **140** (2022) 105113 (cit. on p. 53).

[198] E. Puyol-Anton et al., *Fairness in Cardiac MR Image Analysis: An Investigation of Bias Due to Data Imbalance in Deep Learning Based Segmentation*, Medical Image Computing and Computer Assisted Intervention - MICCAI 2021 (2021) 413–423, ed. by M. de Bruijne et al. (cit. on p. 54).

[199] E. Puyol Anton et al., *Fairness in AI: are deep learning-based CMR segmentation algorithms biased?*, European Heart Journal **42** (2021) (cit. on p. 54).

[200] S. Ioannou, H. Chockler, A. Hammers and A. P. King, *A Study of Demographic Bias in CNN-Based Brain MR Segmentation*, Machine Learning in Clinical Neuroimaging (2022) 13–22, ed. by A. Abdulkadir et al. (cit. on p. 54).

[201] T. S. Mathai, Y. Wang and N. Cross, *Assessing Lesion Segmentation Bias of Neural Networks on Motion Corrupted Brain MRI*, Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries (2021) 143–156, ed. by A. Crimi and S. Bakas (cit. on p. 54).

[202] U. C. Anazodo et al., *A framework for advancing sustainable magnetic resonance imaging access in Africa*, NMR in Biomedicine **36** (2023) e4846 (cit. on p. 54).

[203] M. Adewole et al., *The Brain Tumor Segmentation (BraTS) Challenge 2023: Glioma Segmentation in Sub-Saharan Africa Patient Population (BraTS-Africa)*, 2023 (cit. on p. 54).

[204] W. E. Lorensen and H. E. Cline, *Marching Cubes: A High Resolution 3D Surface Construction Algorithm*, SIGGRAPH Comput. Graph. **21** (1987) 163–169 (cit. on p. 54).

[205] A. Hoopes, J. E. Iglesias, B. R. Fischl, D. N. Greve and A. V. Dalca, *TopoFit: Rapid Reconstruction of Topologically-Correct Cortical Surfaces*, Proceedings of machine learning research **172** (2022) 508–520 (cit. on p. 54).

[206] A.-M. Rickmann, F. Bongratz and C. Wachinger, *Vertex Correspondence in Cortical Surface Reconstruction*, Medical Image Computing and Computer Assisted Intervention - MICCAI 2023 (2023) 318–327, ed. by H. Greenspan et al. (cit. on p. 54).

List of Figures