Universidad de Concepción
Dirección de Postgrado
Facultad de Ingeniería - Programa de Magíster en Ciencias de la Computación

DEEP IMAGE INPAINTING FOR AUTOMATIC BRAIN TUMOR EXTRACTION USING WEAK LABELS.

Tesis para optar al grado de
MAGÍSTER EN CIENCIAS DE LA COMPUTACIÓN

POR
CONSTANZA PAZ VASQUEZ VENEGAS
CONCEPCIÓN, CHILE
Enero, 2020

Profesor guía: GUILLERMO FELIPE CABRERA VIVES
Departamento de Ingeniería Informática y Ciencias de la Computación
Facultad de Ingeniería
Universidad de Concepción

Profesor co-guía: DAVID SONDAK
Institute for Applied Computational Science
Harvard University

CONICYT-PFCHA/Magíster Nacional/2019-22190225

To mom, dad and brother...

ACKNOWLEDGMENTS

Foremost, I want to thank my friends and family for always being by my side and inspire me to give the best of me. I sincerely express my gratitude to my advisors for the guidance, support and motivation during this work. Thanks for the kind support of many others that have been part of this path.

I thankfully acknowledge the support from CONICYT through the grant CONICYT-PFCHA/Magíster Nacional/2019-22190225. Me and my advisors thank Pamela Guevara, PhD. (University of Concepción) for her useful suggestions at the beginning of this work. This research was partially supported by the supercomputing infrastructure of the NLHPC (ECM-02) and has been possible thanks to the use of AWS credits managed by the NLHPC.

Abstract

Brain tumors are one of the leading cancer-related causes of death in all ages. The diversity of tumor shapes and the varying degrees of the visibility of their edges makes the analysis of tumors complex. The development of automatic tools can enhance tumor visualization and improve understanding and support of tumor-focused tasks. We propose an automatic brain tumor extraction method based on image inpainting. Using weak labels containing the approximate shape of the tumor, we are able to successfully remove the tumor from a Magnetic Resonance Image (MRI) by replacing it with non-tumor tissue through a partial convolution neural network trained over non-tumor tissue regions. Brain tumor extraction is then performed by calculating the residual between the original MRI and the reconstructed image without the tumor. The isolated tumor in the extracted tumor image is amenable to further analysis. To demonstrate the extracted tumor image potential, we performed tumor delineation using an active contour method. By clearly showing the tumor, the proposed method is valuable in helping experts come to an agreement when segmenting biomedical images.

Contents

Abstract **v**

List of Tables **vii**

List of Figures **viii**

Chapter 1 Introduction **1**

Chapter 2 Hypothesis and Goals **4**

2.1 Hypothesis 4

2.2 Goals 4

2.2.1 General Goal 4

2.2.2 Specific Goals 4

Chapter 3 Related Work **5**

Chapter 4 Materials and Methods **10**

4.1 Dataset and medical imaging overview 10

4.1.1 Medical Imaging 10

4.1.2 Dataset 11

4.2 Proposed Method 12

4.2.1 Theoretical Background 12

4.2.2 Partial convolutions for image inpainting 15

4.2.3 Training process 20

4.2.4 Semi-supervised brain tumor extraction 21

Chapter 5 Results and Discussion **24**

Chapter 6 Conclusions and Future Work **31**

References **33**

List of Tables

4.1 T1-weighted CE MRI dataset composition. 12

5.1 Segmentation performance using MorphGAC. 26

List of Figures

4.1 Anatomical planes, brain MRI. 11
4.2 T1-weighted CE MRI dataset sample images for each anatomical plane (axial, coronal and sagittal) and tumor type. 13
4.3 Simple ANN architecture. 14
4.4 U-Net architecture. 15
4.5 Partial convolutions for image inpainting architecture 17
4.6 Brain tumor extraction pipeline. 23

5.1 Non-tumor tissue inpainting prediction 25
5.2 Tumor removal through PConv-net 26
5.3 Tumor segmentation performance on Glioma 28
5.4 Tumor segmentation performance on Meningioma 29
5.5 Tumor segmentation performance on Pituitary tumor 30

Chapter 1

Introduction

Cancer is the second cause of death worldwide. In 2018, 9.6 million deaths were reported and this number increases every year [1]. Brain tumors and other central nervous system tumors are the most common cancer-related cause of death in childhood (0-14 years old) and within the top ten cancer causes of death in people above 14 years old. Non malignant tumors (e.g meningioma and pituitary tumor) are the most common type and though they grow slow and most of the time do not spread, they negatively impact normal brain function. On the other hand, malignant tumors are cancerous (e.g glioma) and tend to grow faster, spread to other areas and be more aggressive and therefore harder to treat. The two-year survival rate of people with malignant tumors is limited, ranging between 15% and 85% depending on the malignancy grade [2][3].

Tumor delineation is a critical step for diagnosis, patient monitoring and planning of surgical resection and radiotherapy. It is typically performed manually by experienced specialists using one or more imaging modalities, such as computerized tomography (CT) scans or magnetic resonance imaging (MRI), in an effort to contour numerous 2D slides and determine the exact shape of the tumor. This task is time consuming and particularly challenging due to high variability in shapes, sizes and locations of the tumors, requiring a developed clinical expertise to accurately extract the needed information.

Automatic segmentation algorithms have gained relevance due to the cost of generating the delineations and the level of expertise needed to correctly perform the task. Supervised deep learning models are generally preferred when access to expert manual delineations is available. Supervised medical image segmentation approaches are commonly based on the U-Net, which is able to learn the most likely segmentation of the structure of interest using binary masks made by experts [4]. Although U-Net based models have shown reasonable results for brain tumors (e.g [5], [6], [7], [8], [9]), there is limited information about the annotation process, the protocols used and inter-observer agreement of the selected datasets, reducing the chance of incorporating these automatic segmentation models into clinical workflow.

Studies related to intra and inter agreement among experts are not conclusive for delineation tasks. Specifically for the delineation of brain tumors in MRIs, [10] reported an average intra-observer variability of 20%, having the highest variability when less time was employed to fulfill the task. The same study showed an inter-observer variability between 11% and 68%, which is consistent with the results reported by [11], [12] and [13], where it was demonstrated that the performance of each expert is influenced by their level of expertise, the type of tumor being observed and their interpretation of the extent of the tumor boundaries.

In the absence of reliable labels, unsupervised segmentation methods have demonstrated competitive performance when compared to supervised algorithms. Approaches based on clustering methods are one of the most common when dealing with brain tumors [14][15][16][17][18]. They rely on inherent features from the image (e.g voxel intensity or texture) to differentiate healthy tissue from tumors. According to [15], clustering methods achieve their best performance when the intensity values corresponding to the tumor are significantly different from the surrounding structures. On the other hand, unsupervised deep learning algorithms for brain tumor anomaly detection and tumor delineation, have been proposed by [19] and [20]. They use architectures based on variational autoencoders that are able to learn healthy brain representations. Thus, to determine anomalies, images containing brain tumors are reconstructed as healthy and then used to compute a pixel-wise difference between the original image and its reconstruction. In order for this method to work, it is necessary that both healthy and tumor images have the same acquisition protocols. A common point between unsupervised methods is the need of image preprocessing, particularly image enhancement and manual or automatic skull removal [14][15][17][21][19].

To overcome problems related to supervised and unsupervised segmentation, semi-supervised methods have been proposed to combine the best of both. These methods are varied and even though most of them have not outperformed supervised methods, they have been demonstrated to be especially helpful when there is a limited amount of labeled data. Most brain tumor segmentation models are characterized by the combination of unlabeled or weakly-annotated data (e.g presence or absence of tumor) with few labeled images (e.g manual segmentation labels) to exploit feature learning from both kinds of data and improve segmentation outcomes [22][23][24][25]. On the other hand, [26] proposed an active contouring approach that takes advantage of medical knowledge to define boundaries around the tumor and facilitate the delineation of the region of interest. Semi-supervised methods propose the incorporation of additional medical knowledge to improve tumor segmentation. Following that premise, we propose a brain tumor extraction method based on the inpainting of tumoral regions that only requires brain tumor MRIs and an approximation of the tumor shape. The approximations of the tumor shape are referred in this work as weak labels, due to the lower level of expertise required for its generation in comparison with manual delineations. Algorithms for image inpainting are able to either remove unwanted information or to complete missing regions. For this work, we use a partial convolution-based network (PConv-net), proposed by [27], for the inpainting stage given its ability to smoothly fill irregular missing regions on natural images and its demonstrated potential to remove artifacts and unwanted structures from biomedical images [28][29][30][31][32]. The PConv-net learns to reconstruct missing non-tumor tissue regions taken from the surroundings of the tumor. Using this network the tumor is removed from the MRI by replacing the region defined by the weak label with non-tumor tissue. The reconstructed and original image are used for tumor extraction computing the pixel-wise absolute difference between them. Finally, the extracted tumor image with improved tumor visibility can be used to achieve several clinical tasks such as tumor delineation. In this work we propose its use for automatic brain tumor segmentation employing an active contouring method. To the best of our knowledge, this is the first time this inpainting technique is used for automatic brain tumor extraction.

Chapter 2

Hypothesis and Goals

2.1 Hypothesis

A deep image inpainting model that uses weak labels of tumor shapes will be able to remove tumor tissue from an MRI and allow extraction of the tumor from the original image.

2.2 Goals

2.2.1 General Goal

Design an automatic brain tumor extraction pipeline through the use of partial convolutions for tumor inpainting.

2.2.2 Specific Goals

SG1 Train a deep image inpainting model based on partial convolutions for non-tumor tissue inpainting.

SG2 Predict non-tumor tissue over weak label of tumor shape.

SG3 Perform tumor extraction using the original and inpainted MRI.

SG4 Evaluate tumor extraction quality through brain tumor automatic delineation.

Chapter 3

Related Work

Semantic segmentation aims to predict labels for each pixel of an image, giving a better understanding about the spatial relationship between the elements of it. For biomedical images, when manual delineations are available, supervised deep learning models are commonly chosen to perform semantic segmentation. Given its capacity of working with few labels, most proposed approaches for medical image segmentation are based on the U-Net [4], an architecture consisting of a contraction path that captures the context in the images and an expansive path that is able to define precise localization of the structures of interest. Research around MR images provided for the Multimodal Brain Tumor Segmentation Challenge 2019, have demonstrated a Dice Similarity Coefficient above 80% for whole tumors on approaches based on the U-Net architecture [33]. Although supervised learning challenges, such as the Multimodal Brain Tumor Segmentation, promote research around the subject, when moving to clinical scenarios data collection with such quality becomes time consuming.

On the other hand, many unsupervised brain tumor segmentation models have been proposed in the literature. Most of them are based on clustering methods such as K-Means and Fuzzy C-means (FCM)[16][15] and perform skull stripping to facilitate tumor segmentation [14][15][17][21][19]. The main difference between them is the type of MR image modality used (e.g T1, T2, T3, T4 or FLAIR) and the applied preprocessing, including the use of wavelet multi-resolution to manage spatial context between the pixel [18] and intensity adjustments to highlight the tumor [17]. In [14] was proposed a combination of K-means with non-negative matrix factorization to improve the separation between the tissues, reporting a dice coefficient above 86% for whole tumor. Reaching a similar performance is the approach proposed by [21], an untrained method based on expectation minimization. From a deep learning perspective, unsupervised brain tumor segmentation has been addressed as an anomaly detection problem [19][20][34][35]. Emulating the discernment process performed by radiologists when identifying whether an image is healthy or not, autoencoder-based models are trained to learn feature distribution of healthy MRIs. Hence, if a brain tumor image is supplied to the model it will output its reconstruction following healthy features representation. Any deviation from the learned distribution will correspond to an abnormality, which can be visualized by the computation of residual errors between the input image and the reconstructed one.

The preferred approach for the latter is the Variational Autoencoder (VAE) [36], given its capacity to encode the main features on a low dimensional generative latent space. Modifications such as the addition of the structural similarity constraint [34] or restrictions on the latent space [20] are necessary to enhance reconstruction over brain images. Although abnormality detection models for brain tumor segmentation seem promising, no metrics comparable to other methods have been reported. Furthermore, healthy images used for training must have similar acquisition features as brain tumor images in order to have proper reconstructions. These types of images may not be available, especially for contrast-enhanced modalities.

Non-learning methods such as decision forest [22] or active contouring [26] are also present in the literature, being the latter the most relevant to our work. [26] proposes the addition of medical guidance to rigorously define the tumor region and improve active contour models performance by asking the user to draw the smallest possible circle around the tumor to then apply contouring methods to find the segmentation. Despite the improvement, their process is tied to a high medical expertise level. On the other hand, deep learning semi-supervised approaches consider the use of both labeled and unlabeled data as well as different network architectures [23][25]. While [23] proposed a Stacked Denoising Autoencoder trained with unlabeled patches and then fined-tuned with few labeled samples for pixel-wise classification, [25] proposed the combination of a U-Net like architecture with an autoencoder. The U-Net is trained to generate the tumor segmentation from the input image and the autoencoder is trained to reconstruct synthetic tumor segmentation using an attention mechanism, adding value to the supervised segmentation model. Similarly, [37] proposed a Generative Adversarial Network (GAN)-based approach where a U-Net model benefits from the adversarial relation between a Generator, that synthesizes auxiliary segmentation maps, and a discriminator that tries to determine if its input segmentation map comes from the U-Net or the generator. Also following a deep learning approach, [24] reported a Dice Similarity Coefficient above 80% in whole tumor segmentation by the use of mixed-supervision, which includes a pixel-wise classification method for segmentation and a weak supervision branch for image classification based on tumor presence or absence, extracting relevant features from both kind of labels.

Inspired by anomaly detection models and the use of weak labels of the region of interest, we propose the use of an image inpainting model to remove the brain tumor from the image and subtract both images (the inpainted image and the image with the tumor) to produce a clear image of the tumor. Image inpainting aims to reconstruct missing portions of an image or to remove any distortion that could be damaging it, such as noise, scratches, watermarks or text. Classical local methods, such as PatchMatch [38], search for the missing pixels among the available information on the input image, failing when the region to complete is large or too complex. On the other hand, computer vision community taking advantage of the performance of convolutional layers [39] on the extraction of high-level features from images, has proposed several learning-based methods for natural images inpainting outperforming classical methods. In particular for medical images, available methods on the literature are generally based on GANs. For medical images, [40] performed a comparison between a context encoder model and a contextual attention network. They reported that although the contextual attention network gave better results in terms of edges blending with the original pixels, its predictions for the missing region were not as sharp as the ones produced by the context encoder. On the other hand, among GAN-based approaches there is the ip-MedGAN [41]. A conditional GAN (cGAN) architecture that through a cascade of multiple U-Net networks [42] as generator and two discriminators, global and local, is able to realistically complete missing information from brain CTs and MRIs, outperforming other approaches meant for natural images. Despite the reported performance, the model only completes squared regions of fixed size and the location of the target region must be given to the local discriminator. To overcome those limitations, a second version of the model was proposed in [43] called ipA-MedGAN. The ipA-MedGAN replaces the generator by a MultiRes-UNet [44] cascade architecture to improve the U-Net performance and changes the patch-based local and global discriminators for two discriminators with different receptive fields that accomplish the same goal than before, but without the need of having a fixed region shape and location information. Though the ipA-MedGAN has outperformed most models in the inpainting of non-tumor brain tissue, it has not been tested on tumor removal. Its applications are constrained to remove image distortions produced during the acquisition process. On the other hand, the CT-GAN [45] was proposed with the goal of tampering medical images. It consists of two cGANS: GAN_{inj} for tumor injection and GAN_{rem} for tumor removal. Both networks are deep 3D convolutional cGANs trained to perform inpainting in a cube in the center of the input sample. In order to inject a tumor, the GAN_{inj} is trained on tumor images, and, to remove tumors, GAN_{rem} is trained using healthy samples. The tampered images were able to fool both medical experts and a deep learning classification model. Although the quality of the reconstructions, an extensive pre and post processing is required in order to obtain results in which the inpainted region blends smoothly with its surroundings. Additionally, only squared masks were used. Current methods for medical image inpainting are commonly based on squared missing regions, which restricts its application to tumor inpainting given that they might not correctly adapt to the high variety of possible shapes, resulting in over-inpainting of healthy tissue. Furthermore, given the fact that to study abnormalities most brain pathologies require the use of a gadolinium contrast medium, the proposed methods may not work on contrast enhanced images since they are trained over healthy images with different acquisition protocols. Finally, even though GANs are able to fulfill the task in simpler scenarios they could suffer of mode collapse [46][47] and overfitting, leading to wrong inpainting results.

Given its performance on irregular region inpainting over natural images, for the present work, we chose the PConv-net [27]. It consist of a U-Net to which each standard convolutional layer is replaced by a partial convolution layer. This layer, besides receiving the feature map as input, also receives the binary mask corresponding to the missing region, which is used to condition the output to the unmasked values and then updated for the next layer. In relation to medical images, this model has been mostly used to remove artifacts and unwanted external elements. Such is the case of [28] that proposed its use to remove metal artifacts from CT images, [29] that incorporated it to remove occlusions from skin lesion images aiming to improve lesion segmentation, and [30] that employed it to remove from an intraprocedural CT image the probe used in thermal ablation of liver tumors, with the final goal to improve the registration of pre-procedural MR images to its intraprocedural CT image. For a different purpose, [31] proposed the use of partial convolutions to replace partially-obstructed prostate regions with negative cancer tissue with goal of finding a prostate suspiciousness cancer map. Specifically for brain images, [32] proposed the use of partial convolutions to improve brain registration.

Chapter 4

Materials and Methods

In this chapter the selected dataset and the proposed method are presented. First, basic concepts of medical imaging are introduced to support the description of the type of data chosen for this work. Afterwards, the proposed method is introduced along with the theory needed for its application, going through artificial neural networks, convolutional neural networks and partial convolutions for image inpainting, to end up with the proposed pipeline for automatic brain tumor extraction.

4.1 Dataset and medical imaging overview

4.1.1 Medical Imaging

Anatomical planes

When acquiring a medical image a standard anatomical position is used as reference for posterior image analysis. In order to correctly define this position the subject must be standing with the head facing forward. The basic anatomical planes are the axial plane, sagittal plane and coronal plane, which are orthogonal to each other [48]. The axial plane is parallel to the ground and the images are taken from the top of the head to the feet (Fig. 4.1 (a)). On the other hand, the sagittal plane is parallel to the median plane, that divides the body in half, and its images are taken from one ear to the other (Fig. 4.1 (b)). Finally, the coronal plane is perpendicular to the median plane and the images are taken from the front to the back of the body (Fig. 4.1 (c)).

Brain tumor imaging

CT scans and MRI are the most common imaging techniques used for brain tumor diagnosis, treatment planning and monitoring. While CT scans use a motorized X-Ray

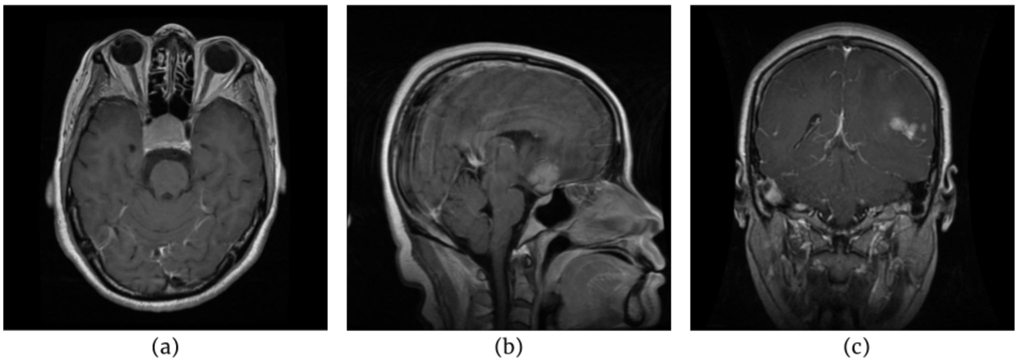

Figure 4.1: Anatomical planes, brain MRI.

(a) Axial plane (b) Sagittal plane (c) coronal plane.

tube that rotates around the patient to collect images [49], MRI employs powerful magnets [50] and radiofrequency waves that stimulate the protons of the body to create a diagnostic image. The conventional MRI protocol for brain tumor assessment includes the following MRI sequences: T1-weighted imaging, T2-weighted imaging, FLAIR, diffusion weighted imaging, postcontrast sequences, perfusion and spectroscopy [51]. Each MRI sequence is a particular configuration of radiofrequency pulses and gradients that result in a specific type of image showing different tissue properties [52]. Although, for brain tumors, CT scans are commonly used as screening method, MRIs are preferred for detailed tumor analysis. MRIs provide sensitive information about brain tissue condition and the relation of the tumor with its surroundings [53]. Furthermore, given the difficulty of observing some tumors, paramagnetic contrast agents are frequently injected to the patient to enhance the visualization of certain tumor characteristics and the demarcation of the boundaries [53].

4.1.2 Dataset

Due to the variety of brain tumor types and anatomical planes the T1-weighted contrast-enhanced (CE) MRI dataset was selected for this work, which has been made publicly available in [54]. The dataset consists of 3,064 images, representing slices from 233 patients collected from the Nanfang Hospital (Guangzhou, China) and the General Hospital (Tianjing Medical University, China). The images have an in-plane resolution of 512×512

Table 4.1: T1-weighted CE MRI dataset composition.

Type/Plane	Axial	Coronal	Sagittal	Total
Meningioma	223	227	258	708
Glioma	447	509	470	1,426
Pituitary tumor	327	306	297	930
Total	997	1,042	1,025	3,064

with a pixel size of 0.49×0.49 mm^2. Furthermore, the slice thickness is 6 mm and the slice gap is 1 mm.

The images are composed of the three anatomical planes (axial, sagittal and coronal) and contain three types of tumors: meningiomas, gliomas and pituitary tumors. Table 4.1 shows the dataset distribution. Samples for each type of tumor and plane are illustrated in Fig. 4.2. Additionally, each image in the dataset contains a single tumor delineation which was manually performed by three experienced radiologists. There is no further information about the chosen fusion technique to generate the final delineation and the inter-observer agreement. Finally, as a preprocessing step we normalized the pixel intensity values to $[0, 255]$ and resized the images to an in-plane resolution of 256×256, the latter with the aim of reducing computational time.

4.2 Proposed Method

4.2.1 Theoretical Background

Deep Learning attempts to learn different feature representations through multiple layers that are typically Artificial Neural Networks (ANNs). The ANNs (Fig. 4.3) are simplified mathematical models that aim to emulate the electrical activity of the brain and nervous system. They are composed of multiple neurons organized in layers in which given a certain input ($\mathbf{x}$) it is expected to get an output ($\hat{y}$). This process is called forward propagation and starts when each component of $\mathbf{x}$ is multiplied with a set of weights ($\mathbf{w}$) in each neuron of the layer. Afterwards, the product resulting from each multiplication is summed up and passed by a non-linear activation function which determines the activation state of the neuron. In order to measure the quality of the generated prediction, the error

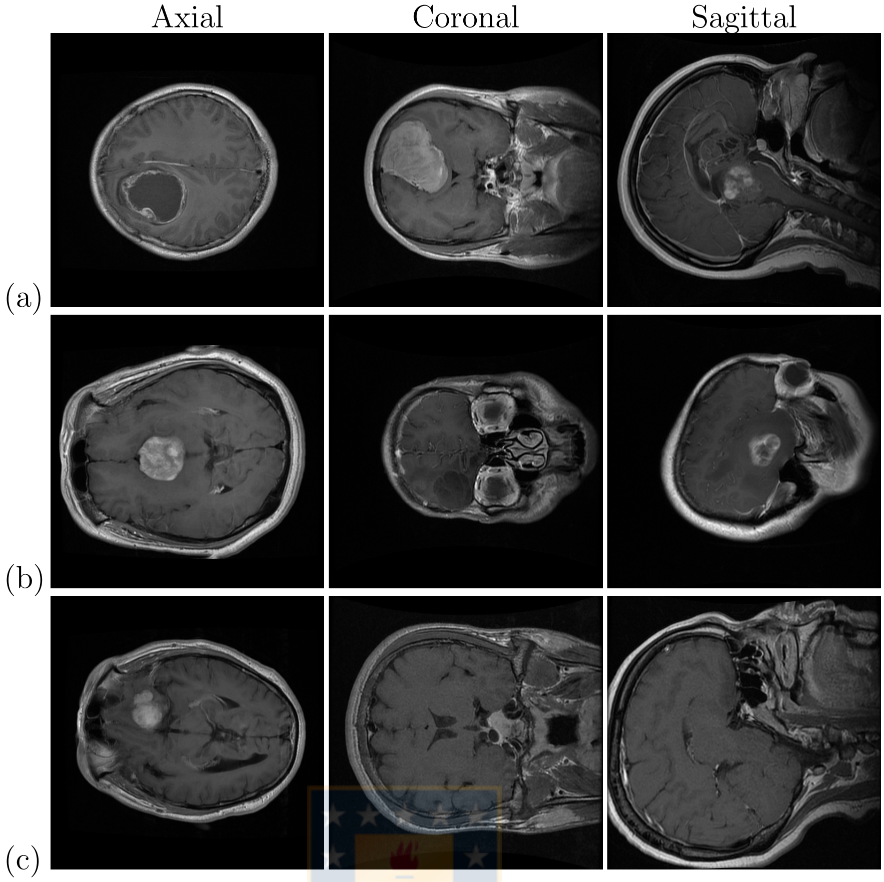

Figure 4.2: T1-weighted CE MRI dataset sample images for each anatomical plane (axial, coronal and sagittal) and tumor type.
(a) Glioma, (b) Meningioma and (c) Pituitary tumor.

(J) of $\hat{y}$ with respect to the ground truth label (y) is computed using a loss function chosen according to the task aiming to solve. Finally, to improve the performance during training the loss function is minimized over the weights using the backpropagation algorithm [55].

In computer vision, convolutional layers [39] are widely used for image-based tasks. They are able to extract features from input data through convolution operations. The convolution operations is an element-wise multiplication of the input image with a kernel of size (K_x, K_y) that slides through the input according to a constant called stride. In Convolutional Neural Networks (CNNs), convolutional layers are often accompanied by a non-linear activation layer (e.g rectified linear unit, ReLU) and a pooling layer. The pooling layer is used to reduce the spatial dimension of the convolution operation output by taking the main characteristics of the feature map within a pool of fixed size (e.g

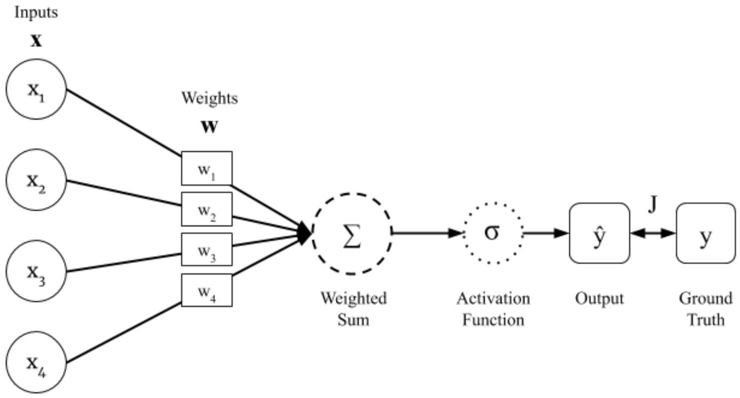

Figure 4.3: Simple ANN architecture.

2 × 2) that moves through the feature map [56]. Among the pooling functions, the most common is max-pooling which chooses the maximum activation among the values in the pool. Other examples of pooling functions include: average or L2 norm of the rectangular region [57].

Built upon the Fully Convolutional Network (FCN) proposed by [58] for semantic segmentation of natural images, the U-Net [4] became the base for most biomedical image segmentation deep learning approaches due to its performance using a few annotated samples for training. The U-Net (Fig. 4.4) is a symmetric architecture consisting of a contraction path that captures the context in the images and an expansive path that is able to precisely define the location of the region of interest. Each block on the contraction path is made up of two convolutional layers with a 3×3 kernel, followed by ReLU activation and a max pooling layer to perform down-sampling. Information about the content of the image extracted through the contraction path is transferred to the expansive path using skip connections, allowing the network to propagate the content information through the expansive path layers. On the other hand, each block from the expansive path consists of an up-convolution layer (up-sampling operation followed by a convolutional layer with a 2 × 2 kernel) whose output feature map is then concatenated with its corresponding feature map from the contraction path. This is followed by two convolutional layers with a 3 × 3 kernel and ReLU activation. Finally, a convolutional layer with a 1 × 1 kernel is added to map the feature maps to the corresponding number of classes for the task.

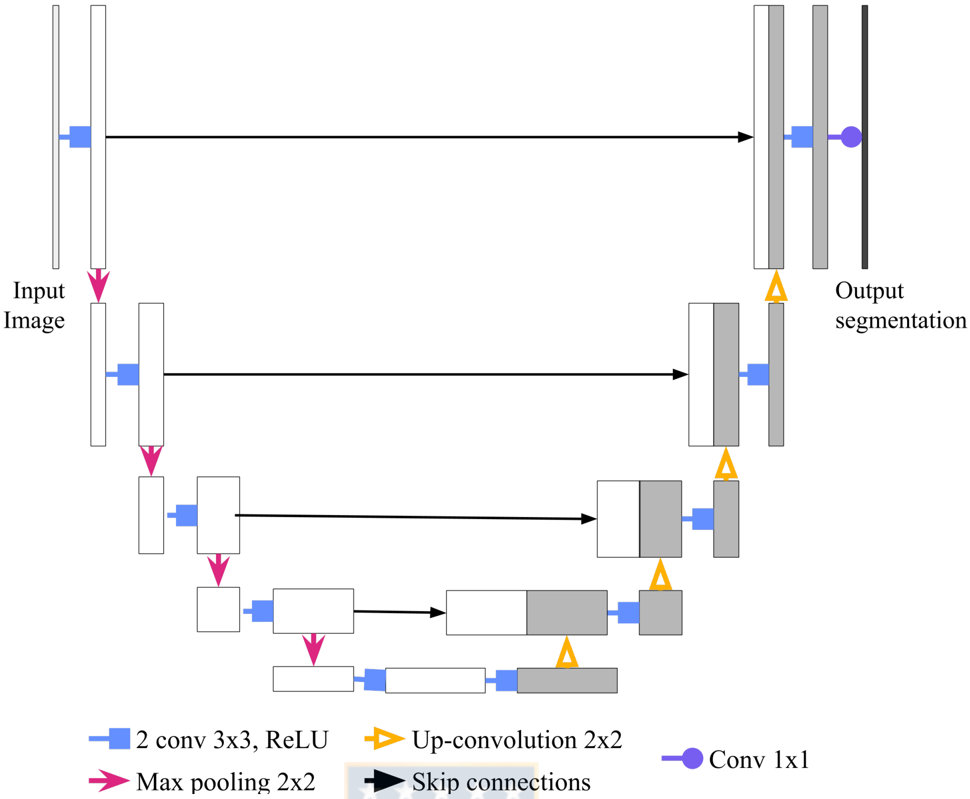

Figure 4.4: U-Net architecture.

Contraction path: feature maps represented by white boxes. Expansive path: feature maps represented by gray boxes. Each arrow color represents a different operation described at the bottom of the illustration above.

4.2.2 Partial convolutions for image inpainting

Aiming to remove brain tumors from MRIs, partial convolutions ([27]) are employed to replace missing areas with non-tumor tissue. The output of the partial convolution layer on the current sliding window (x') is given by,

$$x' = \begin{cases} \mathbf{W}^T (\mathbf{X} \odot \mathbf{M}) \dfrac{\text{sum}(\mathbf{1})}{\text{sum}(\mathbf{M})} + b, & \text{if } \text{sum}(\mathbf{M}) > 0 \\ 0, & \text{otherwise.} \end{cases}, \tag{4.1}$$

where $\mathbf{W}$ are the weights, b the bias, $\mathbf{X}$ the pixel values from the feature map in the current sliding window, $\mathbf{M}$ a binary mask corresponding to the current missing region and $\mathbf{1}$ a tensor of ones with the same shape as $\mathbf{M}$. Element-wise multiplication is denoted by $\odot$ and $\operatorname{sum}(\mathbf{1})/\operatorname{sum}(\mathbf{M})$ corresponds to a scaling factor that adjusts the amount of unmasked pixels of the input feature map. Once the partial convolution operation is applied, the binary mask $\mathbf{M}$ is updated only if at least one masked pixel was reconstructed. The reconstructed pixels are then unmasked in the binary mask via,

$$m' = \begin{cases} 1, & \text{if } \operatorname{sum}(\mathbf{M}) > 0 \\ 0, & \text{otherwise} \end{cases}, \tag{4.2}$$

where m' is the unmasked pixel.

By stacking partial convolution layers, as the missing region from the input image is reconstructed, the number of masked values from the binary mask is reduced until the masked values eventually disappear. To successfully achieve the inpainting of natural images, [27] designed a U-Net like architecture (Fig. 4.5) whose main differences with the original network ([4] and Figure 4.4) are the replacement of convolutional layers by partial convolutions and the use of both the image and its binary mask as input for each layer of the network. Additionally, no down-sampling is applied in the contraction path, the up-sampling on the expansive path is performed through nearest-neighbor interpolation instead of using the up-convolution layer and a Leaky ReLU is chosen as the activation function for the expansive path. Finally, aiming at copying non-masked pixels, the last partial convolution layer concatenates the outputs from the last up-sampling layer and the original inputs of the network. We modify this architecture in order to adapt it for grayscale images and add dropout for generalization purposes. Specific details on the architecture used in the current work are presented in Table 4.2.

The network training is governed by a loss function that accounts for both the per-pixel reconstruction quality and the smoothness of the transition from the predicted values to their surrounding context. An L^1 loss is used for the per-pixel reconstruction loss for pixels inside and outside the masked area,

$$\mathcal{L}_{\text{hole}} = \frac{1}{N_{\mathrm{I}_{\text{gt}}}} \left\| (\mathbf{1} - \mathbf{M}) \odot (\mathbf{I}_{\text{out}} - \mathbf{I}_{\text{gt}}) \right\|_1 , \tag{4.3}$$

$$\mathcal{L}_{\text{valid}} = \frac{1}{N_{\mathrm{I}_{\text{gt}}}} \left\| \mathbf{M} \odot (\mathbf{I}_{\text{out}} - \mathbf{I}_{\text{gt}}) \right\|_1 , \tag{4.4}$$

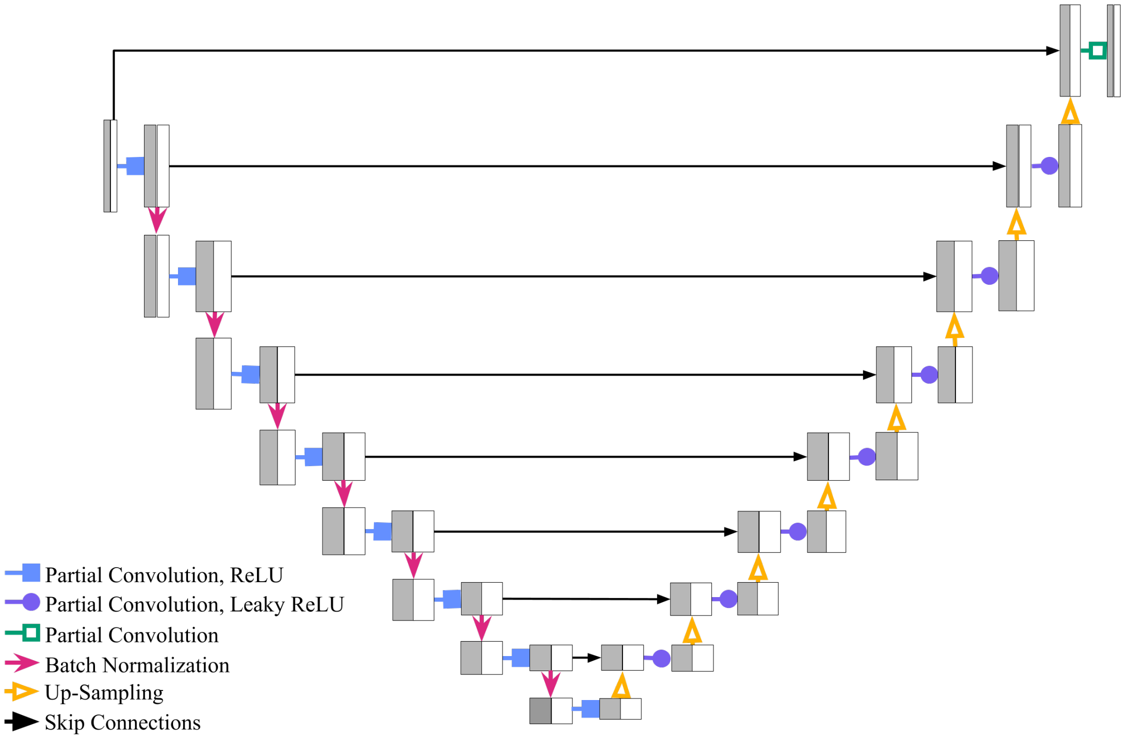

Figure 4.5: Partial convolutions for image inpainting architecture

Partial convolutions, batch normalization, concatenation through skip connections and up-sampling are applied to both masked images (gray block) and masks (white box). Each color of arrow represents a different operation summarized in the figure legend.

where $\mathcal{L}_{\text{hole}}$ and $\mathcal{L}_{\text{valid}}$ are the losses inside and outside the masked area, respectively, $\mathbf{I}_{\text{out}}$ is the prediction, $\mathbf{I}_{\text{gt}}$ is the ground truth, and $N_{\mathbf{I}_{\text{gt}}}$ is the number of elements in $\mathbf{I}_{\text{gt}}$.

The content of the reconstructed missing regions is addressed by the perceptual loss ($\mathcal{L}_{\text{perceptual}}$),

$$\mathcal{L}_{\text{perceptual}} = \sum_{p=0}^{P-1} \frac{\left\| \Psi_p^{\mathbf{I}_{\text{out}}} - \Psi_p^{\mathbf{I}_{\text{gt}}} \right\|_1}{N_{\Psi_p^{\mathbf{I}_{\text{gt}}}}} + \sum_{p=0}^{P-1} \frac{\left\| \Psi_p^{\mathbf{I}_{\text{comp}}} - \Psi_p^{\mathbf{I}_{\text{gt}}} \right\|_1}{N_{\Psi_p^{\mathbf{I}_{\text{gt}}}}}, \tag{4.5}$$

where $\mathbf{I}_{\text{comp}}$ is the network prediction with the valid pixels set to the ground truth and $\Psi_p^{\mathbf{I}_{\text{gt}}}$, $\Psi_p^{\mathbf{I}_{\text{out}}}$ and $\Psi_p^{\mathbf{I}_{\text{comp}}}$ is the pth projection for $\mathbf{I}_{\text{gt}}$, $\mathbf{I}_{\text{out}}$ and $\mathbf{I}_{\text{comp}}$, respectively. The $\mathcal{L}_{\text{perceptual}}$ robustly measures image similarities by computing the difference (L^1 distance) between $\mathbf{I}_{\text{gt}}$ and both $\mathbf{I}_{\text{out}}$ and $\mathbf{I}_{\text{comp}}$. Before computing L^1, $\mathbf{I}_{\text{gt}}$, $\mathbf{I}_{\text{out}}$ and $\mathbf{I}_{\text{comp}}$ are projected

Table 4.2: Partial convolutions for image inpainting architecture detail

PC stands for a partial convolution layer and Concat represents the skip connections that concatenates the output of Up-sampling with the corresponding PC of the contraction path (both masks and feature maps). Additionally, after each PC Batch Normalization is performed excepting the last PC layer and Dropout is added in PC5-13.

Layer	**Size**	**Channels**	**Stride**	**Activation Function**
		Contraction path		
PC1	7x7	64	2	ReLU
PC2	5x5	128	2	ReLU
PC3	5x5	256	2	ReLU
PC4	3x3	512	2	ReLU
PC5	3x3	512	2	ReLU
PC6	3x3	512	2	ReLU
PC7	3x3	512	2	ReLU
PC8	3x3	512	2	ReLU
		Expansive path		
UpSampling1	2x2	512	2	-
Concat w/ PC7	-	512+512	-	-
PC9	3x3	512	1	LeakyReLU
UpSampling2	2x2	512	2	-
Concat w/ PC6	-	512+512	-	-
PC10	3x3	512	1	LeakyReLU
UpSampling3	2x2	512	2	-
Concat w/ PC5	-	512+512	-	-
PC11	3x3	512	1	LeakyReLU
UpSampling4	2x2	512	2	-
Concat w/ PC4	-	512+512	-	-
PC12	3x3	512	1	LeakyReLU
UpSampling5	-	512	2	-
Concat w/ PC3	-	512+256	-	-
PC13	3x3	256	1	LeakyReLU
UpSampling6	2x2	256	2	-
Concat w/ PC2	-	256+128	-	-
PC14	3x3	128	1	LeakyReLU
UpSampling7	2x2	128	2	-
Concat w/ PC1	-	128+64	-	-
PC15	3x3	64	1	LeakyReLU
UpSampling8	2x2	64	2	-
Concat w/ Input	-	64+1	-	-
PC16	3x3	1	1	Sigmoid

to a higher level feature space using the first three pooling layers from a VGG-16 network pre-trained with ImageNet data, whose weights were adapted for grayscale images following [59].

Third, to tackle appearance similarity between the predicted reconstruction and the ground truth image, the style loss ($\mathcal{L}_{\text{style}}$) is included,

$$\mathcal{L}_{\text{style}_{out}} = \sum_{p=0}^{P-1} \frac{1}{C_p C_p} \left\| K_p \left(\left(\Psi_p^{\mathbf{I}_{\text{out}}} \right)^{\mathrm{T}} \left(\Psi_p^{\mathbf{I}_{\text{out}}} \right) - \left(\Psi_p^{\mathbf{I}_{\text{gt}}} \right)^{\mathrm{T}} \left(\Psi_p^{\mathbf{I}_{\text{g t}}} \right) \right) \right\|_1, \tag{4.6}$$

$$\mathcal{L}_{\text{style}_{comp}} = \sum_{p=0}^{P-1} \frac{1}{C_p C_p} \left\| K_p \left(\left(\Psi_p^{\mathbf{I}_{\text{comp}}} \right)^{\mathrm{T}} \left(\Psi_p^{\mathbf{I}_{\text{comp}}} \right) - \left(\Psi_p^{\mathbf{I}_{\text{gt}}} \right)^{\mathrm{T}} \left(\Psi_p^{\mathbf{I}_{\text{g t}}} \right) \right) \right\|_1, \tag{4.7}$$

where $\mathcal{L}_{\text{style}_{out}}$ and $\mathcal{L}_{\text{style}_{comp}}$ are the losses for $\mathbf{I}_{\text{out}}$ and $\mathbf{I}_{\text{comp}}$, respectively, $\mathrm{C}_p \times \mathrm{C}_p$ is the Gram matrix considering that Ψ_p^* are of shape $(\mathrm{H_p W_p}) \times \mathrm{C_p}$ (with $\mathrm{H_p}$, $\mathrm{W_p}$ and $\mathrm{C_p}$ corresponding to height, width and channels, respectively), and K_p is the normalization factor $1/\mathrm{C}_p \mathrm{H_p W}_p$ for the p-th pooling layer.

Finally, prediction smoothness both inside the mask and in its boundary with the original pixels are addressed by the total variation loss ($\mathcal{L}_{\text{TV}}$),

$$\mathcal{L}_{tv} = \sum_{(i,j)\in R,(i,j+1)\in R} \frac{\left\| \mathbf{I}_{\text{comp}}^{i,j+1} - \mathbf{I}_{\text{comp}}^{i,j} \right\|_1}{N_{\mathbf{I}_{\text{comp}}}} + \sum_{(i,j)\in R,(i+1,j)\in R} \frac{\left\| \mathbf{I}_{\text{comp}}^{i+1,j} - \mathbf{I}_{\text{comp}}^{i,j} \right\|_1}{N_{\mathbf{I}_{\text{comp}}}}, \tag{4.8}$$

where R is the masked region dilated by 1-pixel and $N_{\mathbf{I}_{\text{comp}}}$ is the number of elements in $\mathbf{I}_{\text{comp}}$.

The total loss ($\mathcal{L}_{\text{ total }}$) is the combination of the loss terms previously described,

$$\mathcal{L}_{\text{ total }} = \mathcal{L}_{\text{valid}} + 6\mathcal{L}_{\text{hole}} + 0.05\mathcal{L}_{\text{perceptual}} + 120\left(\mathcal{L}_{\text{style}_{\text{out}}} + \mathcal{L}_{\text{style}_{\text{comp}}}\right) + 0.1\mathcal{L}_{tv}. \tag{4.9}$$

Each individual loss is weighted by a factor, the values of which were obtained from [27].

4.2.3 Training process

Independent training was performed for each anatomical plane. Visual inspection of the results revealed superior performance on the coronal plane. This was found to be true even in the reconstruction of tissue in the axial and sagittal plane images. Based on these observations, we used the coronal plane images for the model. We randomly selected 686 samples for training and 250 for validation out of 1042 total images. The test set included 330 samples from the three anatomical planes.

Given that we aim at replacing tumor tissue with non-tumor tissue, it is necessary to teach the PConv-net to precisely reconstruct the non-tumor tissue. The lack of publicly available healthy images with the same acquisition protocols as the T1-weighted CE MRIs selected for this work limit the possibilities for non-tumor tissue reconstruction. We tackle this issue by selecting arbitrary non-tumor tissue regions from the original tumor images for its reconstruction. Multiple irregular binary masks over non-tumor tissue were generated. To obtain the masks we combined circles and irregular shapes in random locations based on manual delineations from the dataset. The manual delineations were modified by applying random rotations and adding or subtracting zero to ten pixels from the edges of the irregular shape. To avoid tumor tissue we use the tumor location determined by the manual delineation available for each MRI. Between twenty and twenty five independent binary masks were generated offline for each training and validation image. The number of masks varies due to the high variability of tumor shapes and available non-tumor tissue. Although the process was performed offline to avoid extending training time, it can also be performed as part of the training pipeline if desired. Once the binary masks were generated, we performed a two step inspection to ensure that they were located over non-tumor tissue. First, masks located in the background, i.e outside the subject, were removed. Second, masks located over regions outside the brain, e.g the neck, were identified by visual inspection and manually removed. Additionally, twenty one irregular binary masks over the tumor were generated for the test set images by randomly adding from one to fifteen pixels around different edge sections of the manual delineation provided on the dataset. These binary masks were then used to perform brain tumor removal with the trained model.

The PConv-net was implemented in Keras Tensorflow [1] and trained using the Adam optimizer and batch size of 16 on four V100 GPUs. Extensive data augmentation was applied to the training set in order to improve generalization of the model, including rotations, flipping and shifting. The progressive filling of the missing regions from the image affects the computation of mean and variance in the batch normalization layer. Consequently, [27] proposed training in two stages. In the first stage, batch normalization is included in both paths, contraction and expansive, using a learning rate of 2×10^{-4}. In the second stage the learning rate is adjusted to 5×10^{-5} and the batch normalization layer is disabled on the contraction path, allowing to apply batch normalization on the filled masks and facilitate convergence.

4.2.4 Semi-supervised brain tumor extraction

Once the PConv-net is trained using non-tumor tissue, the binary masks over tumor tissue ($\mathbf{M}_{\text{tumor}}$) generated for the test set are used to mask its corresponding original image ($\mathbf{I}_{ori}$). The masked image ($\mathbf{I}_{\text{masked}}$) serves as input for the inpainting model which, in inference mode, predicts non-tumor tissue over the missing region ($\mathbf{M}_{\text{pred}}$). The tumor is extracted from $\mathbf{I}_{ori}$ by computing a pixel-wise absolute difference ($\mathbf{I}_{res}$) between $\mathbf{I}_{ori}$ and $\mathbf{I}_{pred}$,

$$\mathbf{I}_{res} = \mid \mathbf{I}_{ori} - \mathbf{I}_{pred} \mid . \tag{4.10}$$

To ensure tumor extraction quality, the binary mask used to remove the tumor from the MRI can be used to remove noise around the extracted tumor.

The residual image ($\mathbf{I}_{res}$) containing the extracted tumor can be then used to obtain the delineation of the tumor ($\mathbf{I}_{segm}$) itself. Given the variability of intensity values among the extracted tumors, we selected the Morphological Geodesic Active Contour[2] (MorphGAC) proposed in [60] to perform tumor delineation. Active Contouring aim to detect the boundaries of an object through the use of evolving contours. In each iteration, in order to minimize parameters related to the target's boundary, it is solved the partial differential equation (PDE) $\mathcal{C}_t = L(\mathcal{C})$, where $\mathcal{C}_t$ is a parameterized 2D curve over time t and

---

[1] Code available at: https://github.com/covasquezv/brain-tumor-inpainting

[2] Morphological Snakes, https://github.com/pmneila/morphsnakes

L is a differential operator that defines the curve evolution. While an internal component maintains smoothness of the delineation, an external component adjusts the curve to the object. Given that the original active contour is purely parametric and does not consider image features, several approximations have been proposed in the literature. MorphGAC [60] is a fast and stable algorithm that is able to achieve the same performance as other methods by replacing terms in the PDE with morphological operators (e.g dilation and erosion). The use of these operators allow to obtain an approximation of the numerical result of the PDE and simplifies the implementation of the algorithm, reducing the computational cost. Furthermore, MorphGAC is able to accurately contour objects through the use of a balloon whose force is determined by the inclusion of relevant image features given by the inverse of gradient magnitude from the image ($g(I)$). Thus, according to a fixed threshold (θ), regions far from the boundaries of the object of interest are ignored by the balloon when $g(I)$ is higher than θ, otherwise, morphological operators are applied to adjust the curve to the boundary. For this work, the chosen value of θ is determined by the method proposed in [61] and reaffirmed by [60] and the size of the initial value is set as three quarters of the input image.

In order to measure segmentation performance, two positively correlated metrics are used: the Dice Similarity Coefficient ([62]),

$$DSC = \frac{2 \mid A \cap B \mid}{\mid A \mid + \mid B \mid}, \tag{4.11}$$

and the Jaccard Similarity Coefficient ([63]),

$$JSI = \frac{\mid A \cap B \mid}{\mid A \cup B \mid}, \tag{4.12}$$

where A is the manual delineation and B the generated delineation. Even though both metrics range from 0 to 1 (with 0 no significant similarity and 1 total overlap) and are similar, they differ in the approach taken to define the level of similarity. While the Dice Similarity Coefficient tends to report values closer to the average performance, the Jaccard Similarity Score reports values that are closer to the worst case.

Image inpainting through partial convolutions is used to remove brain tumors from a T1-weighted CE MRI by replacing a wide delineation of the tumor region with non-tumor tissue. Afterwards, through the computation of the absolute difference between the original image and the inpainted one, we are able to extract the tumor in a residual image, which can then be used to obtain the tumor delineation through MorphGAC. An overall illustration of the pipeline can be found in Fig. 4.6.

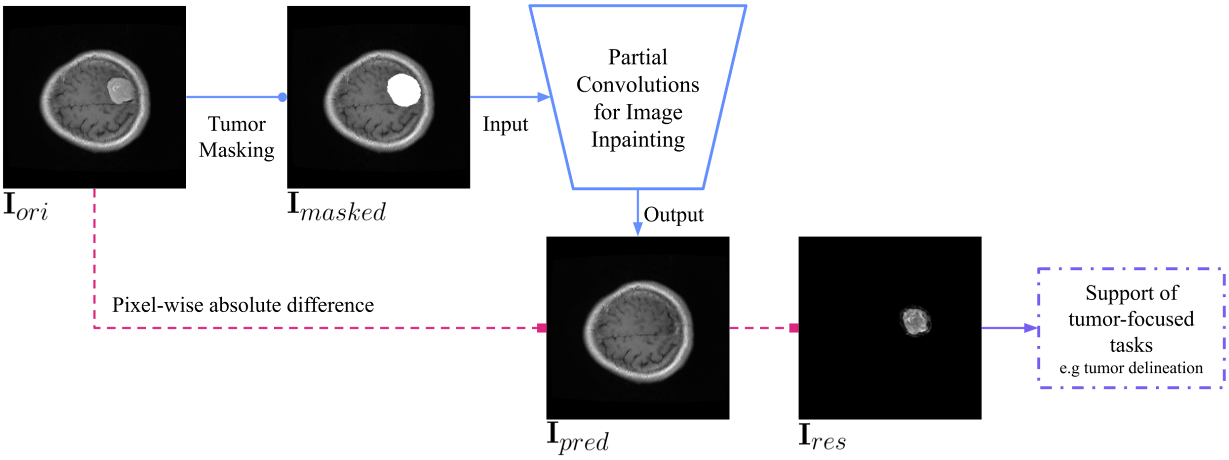

Figure 4.6: Brain tumor extraction pipeline.

Brain tumor extraction pipeline.
Using an approximation of the brain tumor area $\mathbf{I}_{ori}$ is masked ($\mathbf{I}_{masked}$) to then be provided to the PConv-net, which was previously trained using non-tumor tissue missing regions. The output of the network ($\mathbf{I}_{pred}$) is afterwards used for tumor extraction ($\mathbf{I}_{res}$) through the computation of a pixel-wise difference against $\mathbf{I}_{ori}$. Finally, $\mathbf{I}_{res}$ can be employed to support tumor-focused tasks, such us tumor delineation.

Chapter 5

Results and Discussion

A PConv-net was trained to learn non-tumor tissue reconstruction, accounting for both reconstruction accuracy and soft transition with boundary areas while training. Given the amount of correct possibilities when inpainting a missing region, as far it is known, there is no quantitative metric that ensures correct evaluation of the task [27][40]. Therefore, for visualization purposes, results for each anatomical plane are provided in Fig. 5.1. Although the model was trained using only samples from the coronal plane, it is capable of generating correct non-tumor tissue reconstruction for all three anatomical planes, even when the missing region is located on the edge of the brain (Fig. 5.1(a)). There are no significant visible edges of the masked region on the reconstructed image and the predicted pixels follows the appearance of its surrounding pixels.

When removing tumor tissue, the model inpaints the missing region with healthy-like tissue rather than abnormal brain tissue (Fig. 5.2). There is no significant difference among the three anatomical planes for tumor removal. However, when the missing region covers most of the brain area the model fails to successfully remove the tumor. This situation is evident in Fig. 5.2(a) where, although the tumor has been removed, the intensity pixel values from the reconstructed area do not match the surrounding tissue, making it more obvious that the image has been manipulated.

Using the methodology presented in section 4.2.4 we performed brain tumor extraction for each type of tumor and anatomical plane. Results are shown in the third row of Fig. 5.3, Fig. 5.4 and Fig. 5.5, for glioma, meningioma and pituitary tumor, respectively. Through the use of pixel-wise absolute difference between the original image with the one without the tumor (Eq. (4.10)), we are able to isolate the tumor even when the weak label used to define tumor area does not follow a similar shape with the tumor itself (Fig. 5.3(c), Fig. 5.4(b) and Fig. 5.5(a)). Other methods, such us [20] and [35], need to remove the skull to potentially detect anomalies on the brain. Conversely, the extracted tumor image

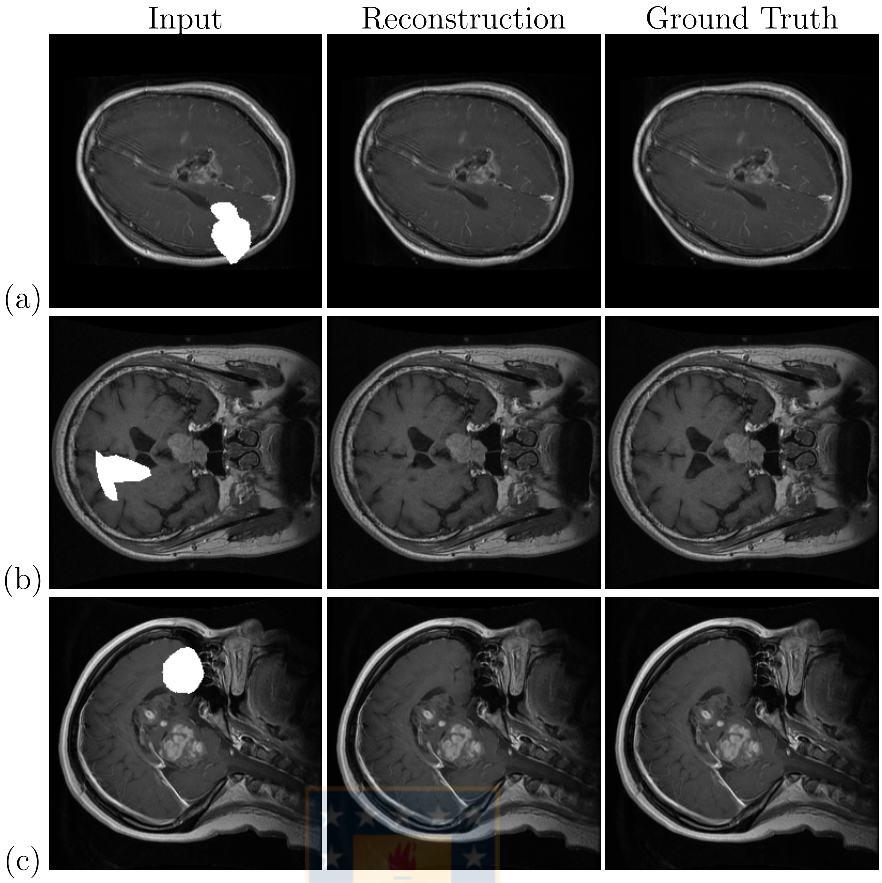

Figure 5.1: Non-tumor tissue inpainting prediction

(a) Axial, (b) Coronal and (c) Sagittal plane.

by the proposed method only contains the tumor facilitating its visualization, especially when the tumor's pixel intensity are close to non-tumor tissue and its boundaries are not clear.

The proposed method is able to successfully extract brain tumors from the input MRIs. Unlike anomaly detection models available on the literature, the approach proposed in this work requires only one kind of image, removing the need of having healthy and tumor datasets with the same acquisition protocols. This is especially relevant when dealing with sensible data with limited availability. Furthermore, since the tumor can be clearly visualized in the extracted tumor image, it can potentially be used for any task related to the pathology, including tumor delineation and tumor measuring.

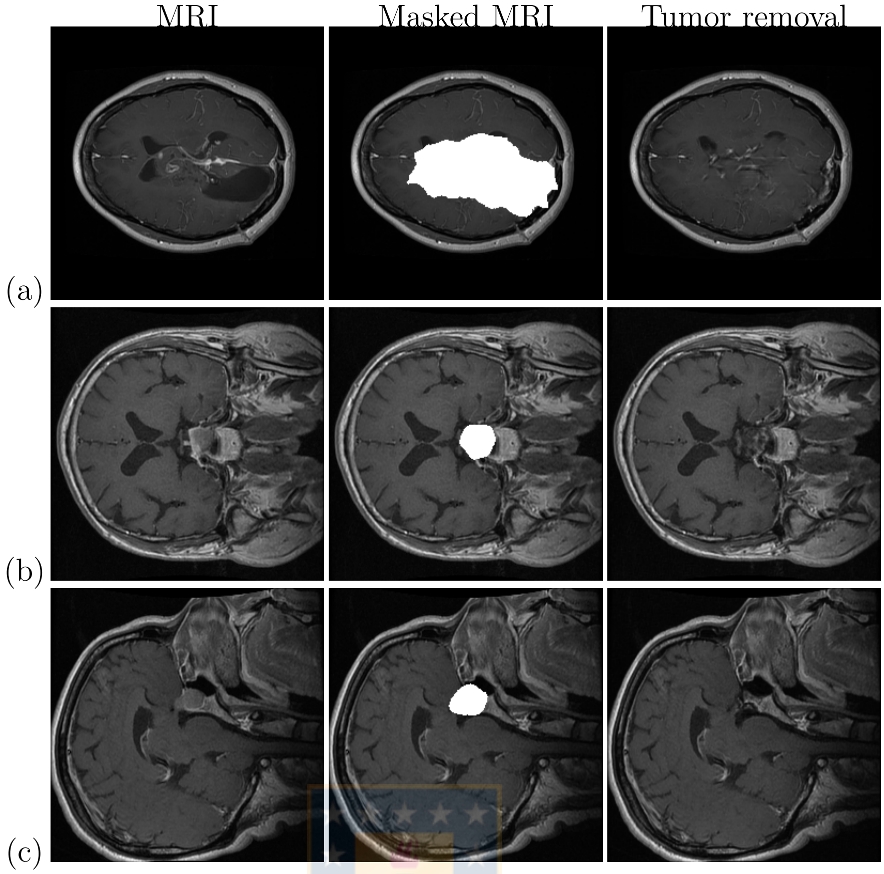

Figure 5.2: Tumor removal through PConv-net

(a) Axial, (b) Coronal and (c) Sagittal plane.

Table 5.1: Segmentation performance using MorphGAC.

Type	Extracted Tumor Image		MRI [26]
	DSC	JSI	JSI
Meningioma	0.8371 ± 0.0629	0.7267 ± 0.0885	0.7794
Glioma	0.8421 ± 0.0568	0.7336 ± 0.0812	0.6631
Pituitary tumor	0.8437 ± 0.0676	0.7375 ± 0.0932	0.7374

To demonstrate the potential use of the extracted tumor images for brain tumor automatic delineation, we employed MorphGAC over $\mathbf{I}_{res}$ to obtain the brain tumor segmentation. The resulting tumor delineations ($\mathbf{I}_{segm}$) for glioma, meningioma and pituitary tumor, along with its corresponding manual delineation, are shown in the last row of Fig. 5.3, Fig. 5.4 and Fig. 5.5, respectively. Better performance is observed when the intensity value from the edges of the tumor are evidently brighter or darker than the surrounding structures (e.g Fig. 5.3(a), Fig. 5.4(c) and Fig. 5.5(c)). Brain tumor segmentations obtained from $\mathbf{I}_{res}$, achieve comparable performance with other approaches using the same dataset and similar delineation method. Evaluation in terms of DSC and JSI are shown in Table 5.1. We were able to outperform [26] in the segmentation of glioma tumors and reach close performance in pituitary tumor and meningioma.

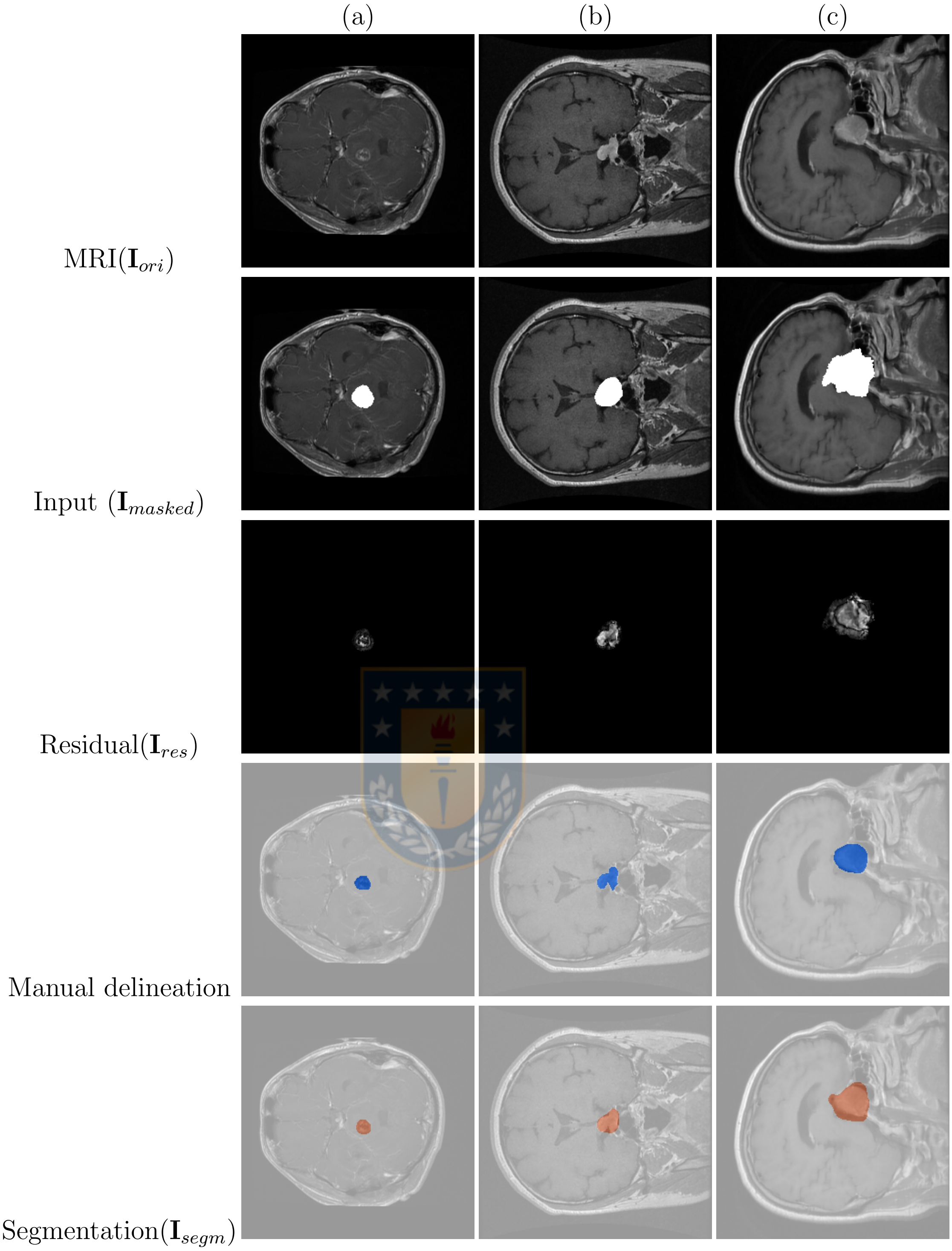

Figure 5.3: Tumor segmentation performance on Glioma

(a) Axial, (b) Coronal and (c) Sagittal plane.

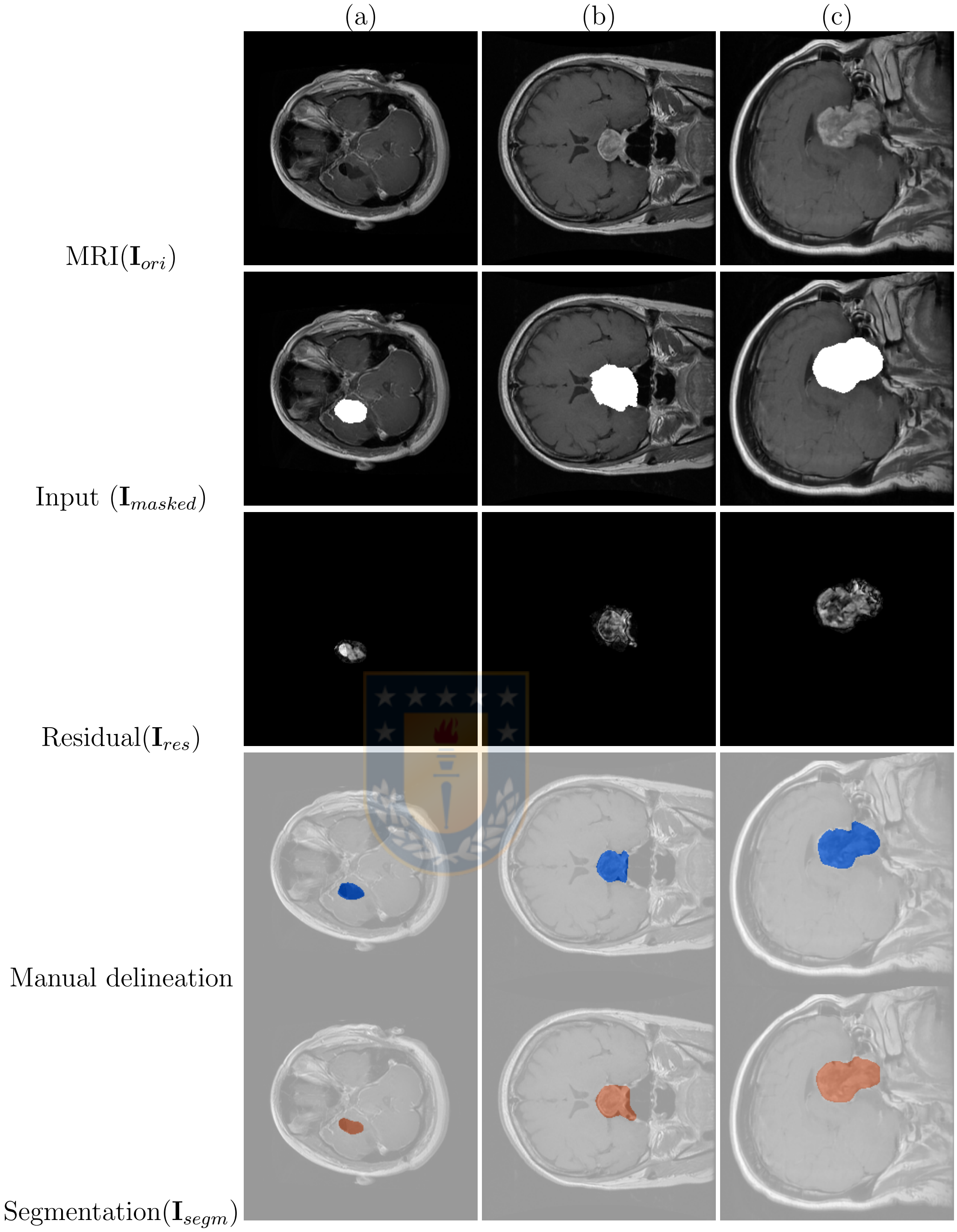

Figure 5.4: Tumor segmentation performance on Meningioma

(a) Axial, (b) Coronal and (c) Sagittal plane.

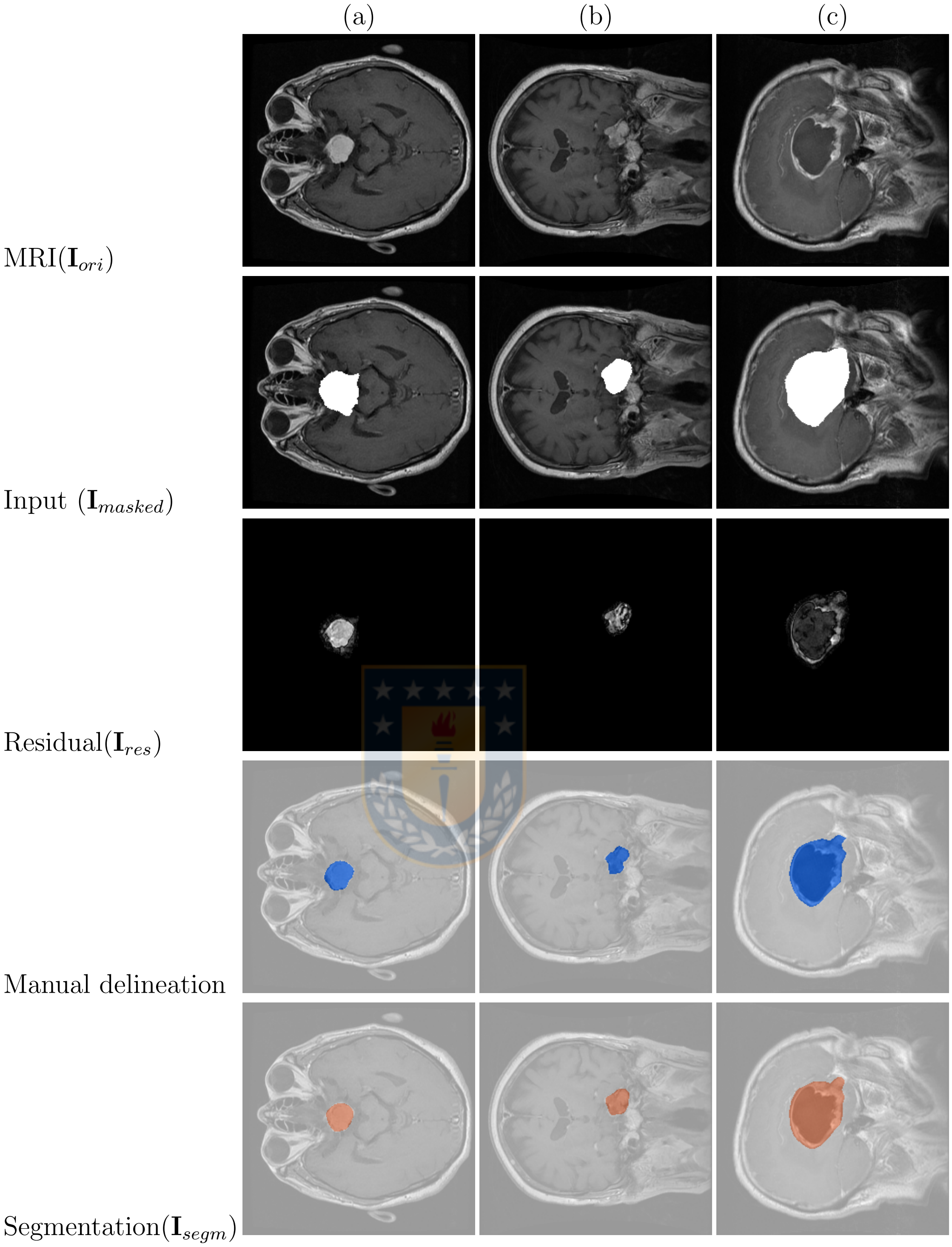

Figure 5.5: Tumor segmentation performance on Pituitary tumor

(a) Axial, (b) Coronal and (c) Sagittal plane.

Chapter 6

Conclusions and Future Work

A pipeline for brain tumor image extraction has been introduced. Based on the concept of abnormality detection, we aim at extracting the tumor from the original image by first removing the tumor from the MRI using image inpainting and then computing the residual between both images. The PConv-net for image inpainting was trained over non-tumor missing regions from coronal plane MRIs. Subsequently, to remove the tumor we used weak labels with an approximation of tumor shape and location to define the region to inpaint by the PConv-net. Once the tumor region is replaced by non-tumor tissue and a healthy-like representation of the MRI is obtained, we compute its pixel-wise absolute difference with the original MRI to extract the tumor.

Although the PConv-net was trained using multiple non-tumor missing regions from only 686 coronal plane MRIs, it is able to correctly fulfill the task of non-tumor tissue reconstruction (Fig. 5.1) and tumor tissue removal (Fig. 5.2) on samples from the three anatomical planes: axial, coronal and sagittal. Lower performance is obtained when the region to inpaint takes most of the brain area. Even in those cases, brain tumors were successfully extracted from the MRI with the computation of the residual. The extracted tumor image provides an improved tumor visibility which opens up the possibility of its use in tumor-centered tasks from a manual or automatic perspective. To demonstrate the potential of using the extracted tumor image, we performed tumor delineation using MorphGAC. Using a lower level of clinical expertise we obtained results comparable to a similar segmentation approach over the same dataset. To the best of our knowledge, this is the first time that partial convolutions have been used for tumor extraction. The use of weak labels is of special interest given the high cost of access to a high level of clinical expertise. On the other hand, the use of the extracted tumor image could potentially support clinical tasks by improving tumor visualization.

In future work, we aim to include automatic wide tumor segmentation to generate the weak label needed for the proposed tumor extraction pipeline. Furthermore, an automatic segmentation deep learning model can be attached at the end of the pipeline to obtain the tumor delineation from the residual image. Additionally, we expect to extend the proposed method to 3D data as well as implementing it for different kinds of tumor and image modalities, such as lung cancer CT images. Finally, the proposed method and the recently described extensions to the model can be unified in a single model able to automatically tackle each one of the pipeline stages. Further research must be done towards that direction.

References

[1] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L Siegel, Lindsey A Torre, and Ahmedin Jemal. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 68(6):394–424, 2018.

[2] Quinn T Ostrom, Gino Cioffi, Haley Gittleman, Nirav Patil, Kristin Waite, Carol Kruchko, and Jill S Barnholtz-Sloan. Cbtrus statistical report: Primary brain and other central nervous system tumors diagnosed in the united states in 2012–2016. *Neuro-oncology*, 21(Supplement_5):v1–v100, 2019.

[3] CP Wild, E Weiderpass, and BW Stewart. World cancer report: cancer research for cancer prevention. *Lyon: International Agency for Research on Cancer*, 2020.

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[5] Guotai Wang, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren. Automatic brain tumor segmentation using convolutional neural networks with test-time augmentation. In *International MICCAI Brainlesion Workshop*, pages 61–72. Springer, 2018.

[6] Wei Chen, Boqiang Liu, Suting Peng, Jiawei Sun, and Xu Qiao. S3d-unet: separable 3d u-net for brain tumor segmentation. In *International MICCAI Brainlesion Workshop*, pages 358–368. Springer, 2018.

[7] Dmitry Lachinov, Evgeny Vasiliev, and Vadim Turlapov. Glioma segmentation with cascaded unet. In *International MICCAI Brainlesion Workshop*, pages 189–198. Springer, 2018.

[8] Mobarakol Islam, VS Vibashan, V Jeya Maria Jose, Navodini Wijethilake, Uppal Utkarsh, and Hongliang Ren. Brain tumor segmentation and survival prediction using 3d attention unet. In *International MICCAI Brainlesion Workshop*, pages 262–272. Springer, 2019.

[9] Mehrdad Noori, Ali Bahri, and Karim Mohammadi. Attention-guided version of 2d unet for automatic brain tumor segmentation. In *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*, pages 269–275. IEEE, 2019.

[10] Gloria P Mazzara, Robert P Velthuizen, James L Pearlman, Harvey M Greenberg, and Henry Wagner. Brain tumor target volume determination for radiation treatment planning through automated mri segmentation. *International Journal of Radiation Oncology* Biology* Physics*, 59(1):300–312, 2004.

[11] Emily M Crowe, William Alderson, Jonathan Rossiter, and Christopher Kent. Expertise affects inter-observer agreement at peripheral locations within a brain tumor. *Frontiers in psychology*, 8:1628, 2017.

[12] M Visser, DMJ Müller, RJM van Duijn, M Smits, N Verburg, EJ Hendriks, RJA Nabuurs, JCJ Bot, RS Eijgelaar, M Witte, et al. Inter-rater agreement in glioma segmentations on longitudinal mri. *NeuroImage: Clinical*, 22:101727, 2019.

[13] S Growcott, T Dembrey, R Patel, D Eaton, and A Cameron. Inter-observer variability in target volume delineations of benign and metastatic brain tumours for stereotactic radiosurgery: results of a national quality assurance programme. *Clinical Oncology*, 32(1):13–25, 2020.

[14] Nicolas Sauwen, M Acou, S Van Cauter, DM Sima, J Veraart, Frederik Maes, Uwe Himmelreich, E Achten, and Sabine Van Huffel. Comparison of unsupervised classification methods for brain tumor segmentation using multi-parametric mri. *NeuroImage: Clinical*, 12:753–764, 2016.

[15] N Sri Sai Krishna Suraj, Vineeth Muppalla, Parita Sanghani, and Hongliang Ren. Comparative study of unsupervised segmentation algorithms for delineating glioblastoma multiforme tumour. In *2018 3rd International Conference on Advanced Robotics and Mechatronics (ICARM)*, pages 468–473. IEEE, 2018.

[16] B Srinivas and G Sasibhusana Rao. Unsupervised learning algorithms for mri brain tumor segmentation. In *2018 Conference on Signal Processing And Communication Engineering Systems (SPACES)*, pages 181–184. IEEE, 2018.

[17] Khurram Ejaz, Mohd Shafy Mohd Rahim, Usama Ijaz Bajwa, Nadim Rana, and Amjad Rehman. An unsupervised learning with feature approach for brain tumor segmentation using magnetic resonance imaging. In *Proceedings of the 2019 9th International Conference on Bioscience, Biochemistry and Bioinformatics*, pages 1–7, 2019.

[18] Sudheesh Kannur Vasudeva Rao and Basavaraj Lingappa. Image analysis for mri based brain tumour detection using hybrid segmentation and deep learning classification technique.

[19] Christoph Baur, Benedikt Wiestler, Shadi Albarqouni, and Nassir Navab. Deep autoencoding models for unsupervised anomaly segmentation in brain mr images. In *International MICCAI Brainlesion Workshop*, pages 161–169. Springer, 2018.

[20] Xiaoran Chen and Ender Konukoglu. Unsupervised detection of lesions in brain mri using constrained adversarial auto-encoders. *arXiv preprint arXiv:1806.04972*, 2018.

[21] Tom Haeck, Frederik Maes, and Paul Suetens. An untrained and unsupervised method for mri brain tumor segmentation. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 265–268. IEEE, 2016.

[22] Raphael Meier, Stefan Bauer, Johannes Slotboom, Roland Wiest, and Mauricio Reyes. Patient-specific semi-supervised learning for postoperative brain tumor segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 714–721. Springer, 2014.

[23] Varghese Alex, Kiran Vaidhya, Subramaniam Thirunavukkarasu, Chandrasekharan Kesavadas, and Ganapathy Krishnamurthi. Semisupervised learning using denoising autoencoders for brain lesion detection and segmentation. *Journal of Medical Imaging*, 4(4):041311, 2017.

[24] Pawel Mlynarski, Hervé Delingette, Antonio Criminisi, and Nicholas Ayache. Deep learning with mixed supervision for brain tumor segmentation. *Journal of Medical Imaging*, 6(3):034002, 2019.

[25] Shuai Chen, Gerda Bortsova, Antonio García-Uceda Juárez, Gijs van Tulder, and Marleen de Bruijne. Multi-task attention-based semi-supervised learning for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 457–465. Springer, 2019.

[26] Ahmad Habbie Thias, Abdullah Faqih Al Mubarok, Astri Handayani, Donny Danudirdjo, and Tati Erawati Rajab. Brain tumor semi-automatic segmentation on mri t1-weighted images using active contour models. In *2019 International Conference on Mechatronics, Robotics and Systems Engineering (MoRSE)*, pages 217–221. IEEE, 2019.

[27] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 85–100, 2018.

[28] Artem Pimkin, Alexander Samoylenko, Natalia Antipina, Anna Ovechkina, Andrey Golanov, Alexandra Dalechina, and Mikhail Belyaev. Multi-domain ct metal artifacts reduction using partial convolution based inpainting. *arXiv preprint arXiv:1911.05530*, 2019.

[29] Devansh Bisla, Anna Choromanska, Jennifer A Stein, David Polsky, and Russell Berman. Skin lesion segmentation and classification with deep learning system. *CoRR*, 2019.

[30] Dongming Wei, Sahar Ahmad, Jiayu Huo, Wen Peng, Yunhao Ge, Zhong Xue, Pew-Thian Yap, Wentao Li, Dinggang Shen, and Qian Wang. Synthesis and inpainting-based mr-ct registration for image-guided thermal ablation of liver tumors. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 512–520. Springer, 2019.

[31] Ruiming Cao, Xinran Zhong, Fabien Scalzo, Steven Raman, and Kyunghyun Sung. Prostate cancer inference via weakly-supervised learning using a large collection of negative mri. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 0–0, 2019.

[32] Zhongqiang Liu. Automatic segmentation of non-tumor tissues in glioma mr brain images using deformable registration with partial convolutional networks. *arXiv preprint arXiv:2007.05224*, 2020.

[33] Alessandro Crimi and Spyridon Bakas. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part I*, volume 11992. Springer Nature, 2020.

[34] Kazuma Kobayashi, Ryuichiro Hataya, Yusuke Kurose, Amina Bolatkan, Mototaka Miyake, Hirokazu Watanabe, Masamichi Takahashi, Naoki Mihara, Jun Itami, Tatsuya Harada, et al. Unsupervised brain abnormality detection using high fidelity image reconstruction networks. *arXiv preprint arXiv:2005.12573*, 2020.

[35] Christoph Baur, Benedikt Wiestler, Shadi Albarqouni, and Nassir Navab. Scale-space autoencoders for unsupervised anomaly segmentation in brain mri. *arXiv preprint arXiv:2006.12852*, 2020.

[36] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[37] Yi Sun, Chengfeng Zhou, Yanwei Fu, and Xiangyang Xue. Parasitic gan for semi-supervised brain tumor segmentation. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1535–1539. IEEE, 2019.

[38] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B Goldman. Patchmatch: A randomized correspondence algorithm for structural image editing. In *ACM Transactions on Graphics (ToG)*, volume 28, page 24. ACM, 2009.

[39] Kunihiko Fukushima. Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural networks*, 1(2):119–130, 1988.

[40] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5505–5514, 2018.

[41] Karim Armanious, Youssef Mecky, Sergios Gatidis, and Bin Yang. Adversarial inpainting of medical image modalities. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3267–3271. IEEE, 2019.

[42] Sohil Shah, Pallabi Ghosh, Larry S Davis, and Tom Goldstein. Stacked u-nets: a no-frills approach to natural image segmentation. *arXiv preprint arXiv:1804.10343*, 2018.

[43] Karim Armanious, Vijeth Kumar, Sherif Abdulatif, Tobias Hepp, Sergios Gatidis, and Bin Yang. ipa-medgan: Inpainting of arbitrarily regions in medical modalities. *arXiv preprint arXiv:1910.09230*, 2019.

[44] Nabil Ibtehaz and M Sohel Rahman. Multiresunet: rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121:74–87, 2020.

[45] Yisroel Mirsky, Tom Mahler, Ilan Shelef, and Yuval Elovici. Ct-gan: Malicious tampering of 3d medical imagery using deep learning. In *28th USENIX Security Symposium (USENIX Security 19)*, 2019.

[46] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[47] Luke Metz, Ben Poole, David Pfau, and Jascha Sohl-Dickstein. Unrolled generative adversarial networks. *arXiv preprint arXiv:1611.02163*, 2016.

[48] Dr Bahman Rasuli and Assoc Prof Craig Hacking et al. Anatomic position.

[49] National Institute of Biomedical Imaging and Bioengineering. Computed tomography.

[50] National Institute of Biomedical Imaging and Bioengineering. Magnetic resonance imaging.

[51] KSA Meteb Nasser Debsh and Dr Bruno Di Muzio et al. Brain tumor protocol (mri).

[52] Andrew Murphy and Assoc Prof Frank Gaillard et al. Brain tumor protocol (mri).

[53] Kamil Zeleňák, Cisáriková Viera, and Poláček Hubert. Radiology imaging techniques of brain tumours. *Clinical Management and Evolving Novel Therapeutic Strategies for Patients with Brain Tumors*, page 77, 2013.

[54] Jun Cheng, Wei Huang, Shuangliang Cao, Ru Yang, Wei Yang, Zhaoqiang Yun, Zhijian Wang, and Qianjin Feng. Enhanced performance of brain tumor classification via tumor region augmentation and partition. *PloS one*, 10(10):e0140381, 2015.

[55] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[56] Dominik Scherer, Andreas Müller, and Sven Behnke. Evaluation of pooling operations in convolutional architectures for object recognition. In *International conference on artificial neural networks*, pages 92–101. Springer, 2010.

[57] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning. book in preparation for mit press. *URL¡ http://www. deeplearningbook. org*, 2016.

[58] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[59] Limin Wang, Yuanjun Xiong, Zhe Wang, and Yu Qiao. Towards good practices for very deep two-stream convnets. *arXiv preprint arXiv:1507.02159*, 2015.

[60] Pablo Marquez-Neila, Luis Baumela, and Luis Alvarez. A morphological approach to curvature-based evolution of curves and surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1):2–17, 2013.

[61] Luis Álvarez, Luis Baumela, Pedro Henríquez, and Pablo Márquez-Neila. Morphological snakes. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2197–2202. IEEE, 2010.

[62] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[63] Raimundo Real and Juan M Vargas. The probabilistic basis of jaccard's index of similarity. *Systematic biology*, 45(3):380–385, 1996.