Automated segmentation of lacunes of presumed vascular origin in brain MRI scans

Literature Study

by

E. Ooms

as partial fulfilment to obtain the degree of Master of Science
at the Delft University of Technology,
to be presented on 13 August 2020 at 2 PM.

Student number: 4749448
Thesis committee: Prof. dr. ir. C. Vuik,   TU Delft, supervisor
                   Prof. dr. ir. M. De Bruijne,   Erasmus MC, supervisor
                   Ir. K. Van Wijnen,   Erasmus MC, supervisor

TU Delft   **Erasmus MC**

Contents

1

Introduction

Cerebral small vessel disease (cSVD) is a disease that can result in different types of brain lesions, including lacunes of presumed vascular origin, perivascular spaces, white matter hyperintensities and microbleeds. It affects mostly elderly people and can have a large impact on psychological and physical abilities. However, the mechanisms underlying the disease are not fully known. cSVD itself is hard to detect with neuroimaging, but the lesions resulting from cSVD are better to image. Therefore, the disease is detected by inspecting brain MRI images for these lesions. This task is mostly performed by a radiologist and can take up a lot of time. Additionally, as the interpretation is executed by a human being, it is subject to bias and variations across interpreters.

In the past, other researchers have come up with several machine learning methods that tried to automate this manual process to assist the radiologist in making more accurate decisions and save time. For lacunes of presumed vascular origin these methods were mostly detection based, which means that the method detects the lesion with drawing a box around it. In order to get more precise information about the size and location, semantic segmentation should be applied as this procedure provides the exact outline of the lesion. Previously, one other method was designed to segment lacunes of presumed vascular origin. However, since the field of machine learning is still advancing, more recent developed techniques might be able to improve the segmentation task even more.

Therefore the aim of this thesis is to come up with a new automated method that is able to produce more accurate lacune of presumed vascular origin segmentation results using the more recent techniques. The automation of this task would reduce the time a radiologist has to spend analyzing the images and increases the accuracy of the diagnosis. Moreover, it also will benefit the patient, since potential diseases can be detected in an earlier stage and as a consequence can be treated more effectively. Lastly, it may help giving more insight in the exact cause of the disease.

This literature study will provide more medical background on lacunes of presumed vascular origin and cerebral small vessel disease in chapter 2. Next, an introduction on convolutional neural networks is given in chapter 3, after which chapter 4 describes how such a network can be trained. A promising architecture within the field of convolutional neural networks, U-Net, is explained in chapter 5. Chapter 6 discusses previously developed methods to automate the process of detecting and segmenting lacunes of presumed vascular origin. Potential challenges that can occur while developing a segmentation method in this thesis are mentioned in chapter 7. Subsequently, a conclusion of this literature study is given in chapter 8. Finally, chapter 9 contains the research proposal, in which the research questions of this thesis are stated.

2

Medical Background

2.1. Cerebral small vessel disease

The term cerebral small vessel disease (cSVD) refers to pathological changes in the small arteries, arterioles, venules and capillaries in the brain. All together these blood vessels are called the small vessels. As a consequence of these changes, lesions can occur in the brain parenchyma, which is the functional tissue in the brain consisting of neurons and glial cells. However, the mechanisms causing the pathological changes that lead to lesions are heterogeneous and not fully understood. It is shown that the parenchymal damages, such as lacunes of presumed vascular origin, white matter hyperintensities of presumed vascular origin and microbleeds, are generally located in the subcortical structures. Since alterations in small vessels are hard to detect with neuroimaging, whereas the resulting lesions are better to image, these lesions are often used as biomarkers for cerebral small vessel disease. As a result of this, the term cerebral small vessel disease often refers to the parenchymal lesions rather than the pathological small vessel changes [34].

There are several types of small vessel diseases which can be subdivided based on its pathology into the amyloidal form and the non-amyloidal form. This separation is based on whether aggregates of proteins, called amyloids, play a role in the disease or not. Within the amyloidal group, cerebral amyloid angiopathy (CAA) is a very common subtype [9]. CAA is a chronic degenerative disease which is predominantly associated with advanced age [5]. It is caused by progressive accumulation of β-amyloid in the walls of small arteries in the brain. The deposition of the β-amyloid leads to occlusion and rupture of the vessels and eventually to brain parenchymal injuries [6, 39]. The non-amyloidal form is also known as age-related and vascular risk-factor-related small vessel disease because it is often associated to ageing, diabetes and hypertension. This form of cSVD represents itself by an increase of the vessel wall, narrowed interior of the vessels, the occurrence of dilated and elongated vessels or the buildup of fats, cholesterol and other substances on the vessel walls [34].

Although the exact link between cSVD and parenchymal damage is still unclear, it is hypothesized that damage to the blood-brain barrier might be a possible cause of some cSVD types [51]. The blood-brain barrier (BBB) is a semipermeable barrier that separates the cardiovascular system from the extracellular fluid of the central nervous system. It prevents the pathogens and circulating immune cells from passively entering and as a result damaging the brain [1, 9]. In several studies it was found that in patients with cSVD, BBB leakage was present [42, 52, 56, 60]. The hypothesis is that with increasing age the permeability of the BBB increases and this damage process is enhanced by several risk factors such as hypertension, diabetes and inflammation. If the BBB is weakened, pathogens and immune cells can then invade into the brain and induce brain parenchymal lesions [9, 51, 56].

The consequences of small vessel disease are diverse. Elderly patients with cSVD experience cognitive decline [48], depressive symptoms [18] and physical disabilities such as gait disturbances [10] and urinary problems [37]. Furthermore, the disease is the cause of 25% of all ischaemic strokes [2], contributes to up to 45% of dementias [53] and appears to be associated with Alzheimer's Disease and Parkinson's Disease [9]. However the exact relation of cSVD with these neurodegenerative diseases is still unclear.

In order to detect small vessel disease, neuroimaging and especially magnetic resonance imaging

(MRI) is used to visualize the different parenchymal lesions. To assess the severity of the disease a total cSVD score was proposed, which takes several of these imaged biomarkers into account instead of evaluating them separately [26, 55]. Besides this, the neuroimaging of biomarkers can also be important to get insight about the cause of the disease and its connection to other neurodegenerative diseases. It is shown that dysfunction of the blood-brain barrier is associated with the consequential parenchymal lesions of cSVD [56, 60]. However, if the exact relation and process between the leaking BBB and the disease can be unraveled, it might help to prevent the small vessel disease from originating.

The parenchymal lesions that are commonly used as cSVD biomarkers are recent small subcortical infarcts, lacunes of presumed vascular origin, white matter hyperintensities of presumed vascular origin, perivascular spaces, cerebral microbleeds and brain atrophy. These lesions vary in size, shape and location in the brain. Furthermore, their intensities on an MRI scan may vary as well from having a low intensity (white) to having a high intensity (dark gray/black) [54]. It is beyond the scope of this report to discuss all of the lesions here. However, since the goal of this research is to automate the segmentation of lacunes of presumed vascular origin, this lesion type will be discussed in more detail in the next section, as well as other lesions that are relevant due to their similarities in appearance.

2.2. Lacunes of presumed vascular origin

Lacunes of presumed vacular origin are small cavities filled with fluid and are mainly located in the deeper parts of the brain [15]. They are presumed to be a result of subcortical infarcts, which on their turn are expected to be caused by occlusion of small arteries [15, 54]. However, the lesions may also be caused by intracerebral haemorrhage as it was found to be associated with lacunes of presumed vascular origin [35, 40]

Since across papers there was little consistency in terminology and definitions for the biomarkers of small vessel disease, Wardlaw et al. proposed a consensus term and definition for each marker. In this report we will use the recommended standards for lacunes of presumed vascular origin as stated by Wardlaw et al. [54]. Therefore, a lacune of presumed vascular origin is defined as "a round or ovoid, subcortical, fluid-filled cavity of between 3 mm and about 15 mm in diameter, consistent with a previous acute small deep brain infarct or haemorrhage in the territory of one perforating arteriole". Lesions can be recognized on MRI sequences, which are combinations of particular settings of radiofrequency pulses and field gradients that influence the appearance of an image. On all MRI sequences the lacunes of presumed vascular origin can be recognized by its intensity which is similar to the cerebrospinal fluid (CSF) signal, that is, hypointense or hyperintense depending on the sequence used. In addition, images of the fluid-attenuated inversion recovery sequence (FLAIR) sequence can in some cases also display a hyperintense rim around the central CSF-like hypointensity. However, on these FLAIR images

Figure 2.1: A lacune on an FLAIR image having a hyperintense ring around a center of hypointensity [54].

the hypointensity is not always present and the lacunes of presumed vascular origin appear entirely hyperintense [54]. An example of a lacune imaged with a FLAIR sequence is depicted in figure 2.1.

Lacunes of presumed vascular origin need to be distinguished from perivascular spaces as they can look very similar. According to Wardlaw et al. [54], perivascular spaces are defined as "fluid-filled spaces that follow the typical course of a vessel as it goes through grey or white matter". On all MRI sequences these lesions also have a CSF-like intensity. They can have an elongated shape if imaged parallel to the course of the vessel, but they can also appear round or ovoid in shape if imaged perpendicular to the course of the vessel. When they pass through an area of white matter hyperintensity, they can even mimic the hyperintense rim of a lacune of presumed vascular origin on FLAIR images [54]. Several studies differentiate between these two lesions based on the diameter size and the presence of the hyperintense rim [4, 16, 27–29, 35]. In these articles, FLAIR imaged hypointense lesions with a diameter between 3 mm and 15 mm with a hyperintense rim are classified as lacunes of presumed vascular origin. Moreover, when the lesion is not round or ovoid in shape but somewhat elongated, it is more likely to be a perivascular space.

In the remainder of this report, the term lacunes of presumed vascular origin will be abbreviated by 'lacunes'. However, in all cases this refers to the consensus term and definition of lacunes of presumed vascular origin.

2.3. Data: Rotterdam Scan Study

For the segmentation of the lacunes, brain MRI scans from the Rotterdam Scan Study will be used. The Rotterdam scan study (RSS) is a study with the aim to examine causes and consequences of neurological diseases among the elderly by imaging the underlying pathological changes in the brain [20]. Its participants originate from the related Rotterdam study, a population-based study with the goal of unraveling causes and consequences of chronic disease in mid-life and late-life. The group of participants in this Rotterdam study consists of around 18,000 inhabitants of the district Ommoord in the city of Rotterdam who are all aged 40 years or over. After every 3 to 6 years they undergo extensive physical examinations and are interviewed [21]. A selection of the same group of participants was approached to participate in the RSS to additionally undergo MRI-exams. More specifically, within the group of participants of the Rotterdam study only the people without dementia, MRI contra-indications, claustrophobia and the persons that did give consent were considered for the Rotterdam scan study. As a result, by 2015 the amount of 12,147 brain MRI scans have been collected of over 5,800 different participants of 40 years and over [20].

The MRI scans were all performed by a 1.5 Tesla scanner with an 8-channel head coil (General Electric Healthcare). The Tesla value and the number of channels are indicators of the image quality and the examination times. In general, the higher the Tesla and channel number, the quicker the images can be made and the higher their quality. Scanners can have a channel number of up to 32 and often have a Tesla value of up to 7.0, but can also go beyond. The examinations were all executed by trained radiology technicians according to a standardized protocol. Several high-resolution MRI sequences were performed of which, for the purpose of this research, the images of the FLAIR sequence and T1-weighted sequence will be used. The slice thickness of the FLAIR sequence and the T1-weighted sequence are 2.5 mm and 1.6 mm respectively and the slices were contiguous [20]. Furthermore, the scans have a dimension of $512x512x192$ voxels, where a voxel is the 3D equivalent of a pixel and have a size of 0.5 mm in both height and width.
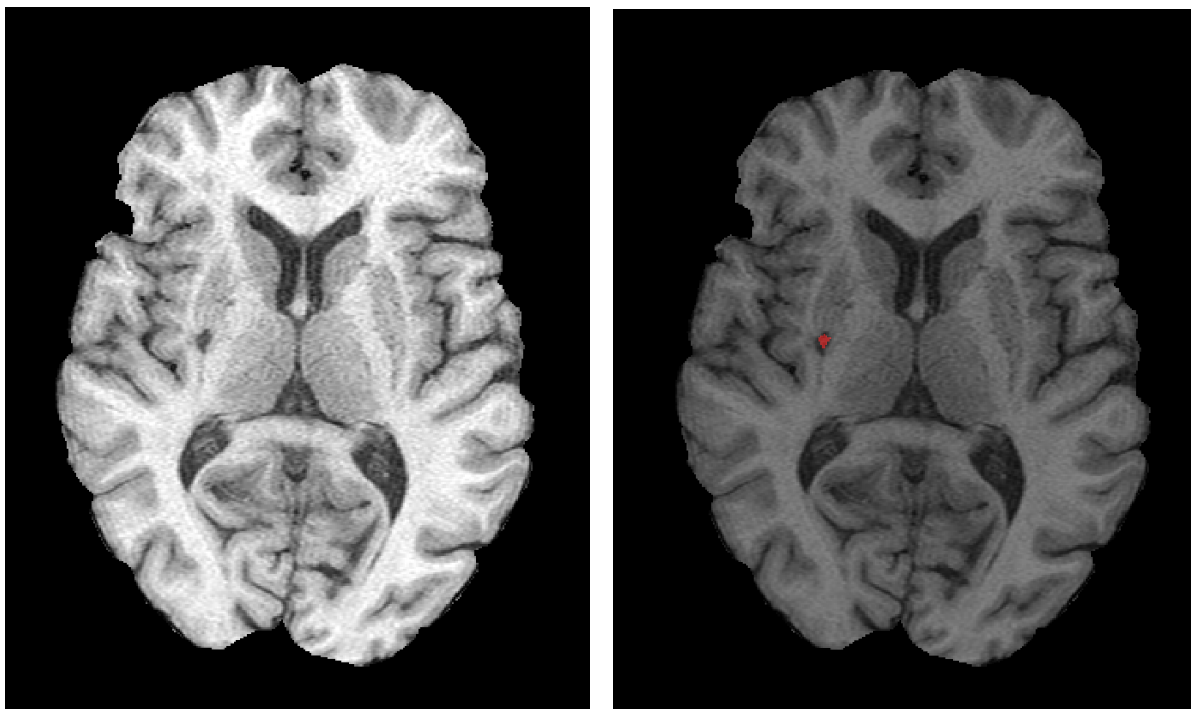
(a) Original T1-weighted scan.   (b) Annotated T1-weighted scan. Here, the lacune is given a red color.

Figure 2.2: Brain MRI T1-weighted scan containing a lacune.

In order to obtain a scan with only brain tissue, the brain is extracted from the image: the cerebellum, eyes and skull are removed. Subsequently, scans are corrected for non-uniformity in intensity. Furthermore, lacunes in the Rotterdam Scan Study are defined as focal lesion of ≥ 3 and < 15 mm in size with the same signal characteristics as CSF on all sequences, and with a hyperintense rim on the FLAIR sequence [20]. This means that on a T1-weighted image they look hypointense (see figure 2.2a), while on a FLAIR images they look hypointense with sometimes a hyperintense rim. Around 5,000 of the 12,147 MRI scans have been examined for the presence of these lesions. If lacunes were identified, they were annotated on the T1-weighted image, that is, they were provided with an overlaying mask as is seen in figure 2.2b. These lacune annotated scans are also called the ground truth. As a result, we have 734 annotated scans and around 4,000 scans without annotation.

3

Convolutional neural networks

Artificial intelligence gained a lot of attention in the last couple of years. In [7], artificial intelligence is defined as "the effort to automate intellectual tasks normally performed by humans". A part of the field of artificial intelligence consists of non-learning approaches, which means that with these approaches programmers manually have to define a large set of explicit rules to automate a process. However, although this appears to be suitable for simple well-defined and logical problems, these approaches fail to solve the more complex problems.

A subfield of artificial intelligence that is able to address those more complex problems, such as language translation, speech recognition and image classification, is called machine learning. Instead of being manually specified, in machine learning the set of rules are learned by a computer when looking at the data. When a machine learning system is provided with many examples, which include both the problems and their answers, and with features it needs to pay attention to, the system can find a statistical structure and can come up with rules to automate the task. These learned rules can then be applied to unseen examples of the same task to predict the answer of the new examples. For example, if we want to let a machine learning system learn to classify whether a picture contains a cat or not, we have to feed the system with a lot of example pictures of cats and pictures without cats. Additionally, as the system needs to know how a cat can be recognized, we also need to tell the system which features are characteristic for a cat, such as a tail, ears, eyes and whiskers. With these example pictures and features, it defines rules to automate the classification process such that when the learned system is provided with a new image it can categorize it as a cat or a non-cat image.
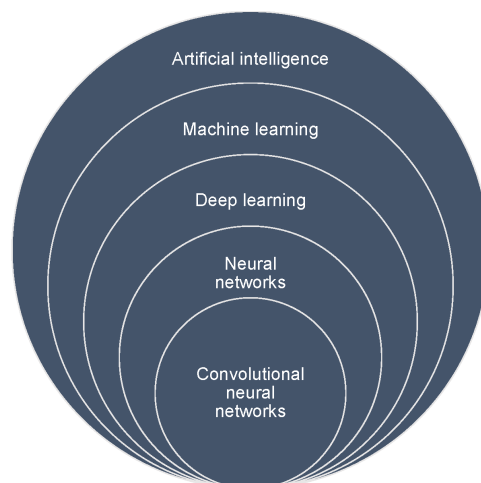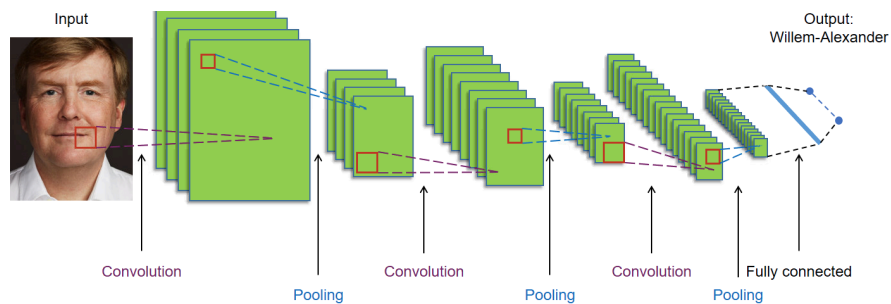

Figure 3.1: Graphical explanation of the relationships between artificial intelligence, machine learning, deep learning, neural networks and convolutional neural networks.

Deep learning is a subgroup of the field of machine learning (see figure 3.1). The main difference between deep learning techniques and other techniques within the machine learning field, is that deep

learning techniques learn to identify characteristic features of an object automatically, while for other techniques these features have to be manually specified. Because of this automated feature extraction, more complex and an increased number of features can be extracted which leads to better performance on many problems [7, 17].

A convolutional neural network (CNN) is a deep learning algorithm that is mostly used for analyzing images and image classification problems. It falls within the group of neural networks, which is a part of the deep learning field. A general CNN architecture consists of convolution layers, activation functions, pooling layers and a fully connected layer with an end activation function. When it is provided with an image, the information will be passed on to a first convolution layer (figure 3.2a). This convolution layer transforms the input with a so called kernel and outputs the new information to the next convolution layer. The kernel acts as a kind of filter that is able to detect patterns. As can be seen in figure 3.2b, in the first layers these patterns are simple features, like edges [7]. The kernels in later layers can detect more sophisticated features such as ears, eyes and mouths. In the deepest layers, even entire objects can be detected using these kernels, and the network is able to tell you that on the input image a human is portrayed: Willem-Alexander, King of the Netherlands (figure 3.2a). In the following sections the components of a CNN will be explained in more detail.

(a) Graphical explanation of the process within a convolutional neural network. An image goes through several operation layers after which a prediction of the output is given.
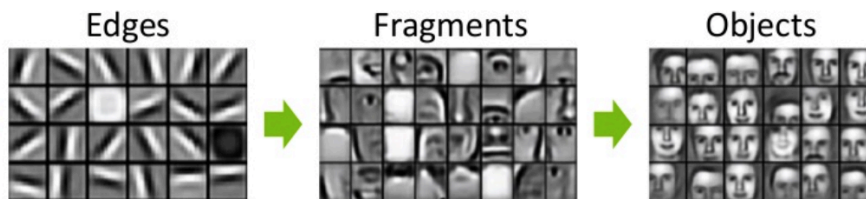
(b) Examples of features through a neural network, starting with simple edges and ending with entire objects.

Figure 3.2: An overview of a convolutional neural network (CNN).

3.1. Convolution layer

In a convolution layer, a so called feature map is given as input and transformed using a kernel into a new feature map. The input feature maps, output feature maps and kernels are usually multi-dimensional arrays. In the first convolution layer, the input feature map can be an image, but in later layers these input maps are the output feature maps of previous layers. A kernel is a matrix with weights that is able to find one specific pattern, for example a horizontal edge. It often has the size of $3x3$ or $5x5$ and as these are small matrices compared to the input, a convolution is sparse. By applying the same transformation kernel to all of the patches within the input feature map, it can detect all horizontal edges in the given input. If we want the network to find even more features, such as vertical and diagonal edges, more distinct kernels have to be applied to find these patterns as well [7, 13, 17].

A convolution is executed by sliding the kernel over the input feature map, stopping at all possible positions. At each location, every entry of the kernel is multiplied by its corresponding overlapping entry of the input feature map. The results of all kernel entry multiplications are summed to obtain the value of the output feature map at that position [7, 13]. Figure 3.3 shows how the process works for a 2-dimensional problem. In the top of the figure we see that each entry of the $2x2$ kernel is multiplied by the corresponding value of the upper left $2x2$ window that is part of the $4x4$ input feature map. The

Figure 3.3: A kernel of $2x2$ slides over the input feature map of $4x4$. At each position it calculates the sum of the multiplications of each kernel entry with their corresponding overlapping window entry. The result of the calculation is positioned in the output feature map at the same location. The kernel slides with one stride (stepsize) to the right and below. Only the first four and the last windows and their calculations are given in this figure [13, 17].

result of the first window appears in the output feature map at the same location. In the second step, the window has slid one stepsize, which is called a stride, to the right to apply the same transformation to a new region. After three slides, the fourth window begins one stride lower, again at the left. This procedure of sliding and calculating continues until the kernel transformed every patch of the input feature map.

The entire process can be repeated by applying as many different kernels as needed. However, when more than one kernel is used, the output feature map will contain a third dimension. As a consequence of this, cuboid, 3-dimensional, kernels should be applied to slide over the width, height and depth of the input feature map [13].

The convolution arithmetic from figure 3.3 can also be represented by one generalized formula. Let $I \in \mathbb{R}^{K \times L}$ be the input feature map, $K \in \mathbb{R}^{M \times N}$ be the kernel and the output feature map S with dimensions of $\left(\frac{K - M}{\text{pooling stride}} + 1 \right) \times \left(\frac{L - N}{\text{pooling stride}} + 1 \right)$. An entry of the output map can then be calculated as follows [17],

$$S(i, j) = (K * I)(i, j) = \sum_{m}^{M} \sum_{n}^{N} I(i + m, j + n) K(m, n). \tag{3.1}$$

3.2. Activation function

In the convolution layer, only linear operations are used. This means that the network would only be able to learn the linear transformations of the input. If we want the model to learn complex mapping functions as well, a non-linearity, also referred to as activation function, is needed [7]. An activation function is applied elementwise such that for $X \in \mathbb{R}^{m \times n}, f : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}^{m \times n}$ [19]

$$(f(x))_{ij} = f(x_{ij}). \tag{3.2}$$

The most common activation function for neural networks is the rectified linear unit (ReLU) [17]. If the input is positive it "activates", if the input is negative it will become 0 [22, 33]:

$$f(x_{ij}) = \begin{cases} 0 & \text{for } x_{ij} \leq 0, \\ x_{ij} & \text{for } x_{ij} > 0. \end{cases} \tag{3.3}$$

Other often used activation functions are leaky ReLU [19] and exponential linear unit (ELU) [7]. The leaky ReLU is given by [31],

$$f(x_{ij}) = \begin{cases} 0.01 x_{ij} & \text{for } x_{ij} \leq 0, \\ x_{ij} & \text{for } x_{ij} > 0, \end{cases} \tag{3.4}$$

and the ELU activation function is given by [8],

$$f(x_{ij}) = \begin{cases} \alpha(e^{x_{ij}} - 1) & \text{for } x_{ij} \leq 0, \\ x_{ij} & \text{for } x_{ij} > 0, \end{cases} \tag{3.5}$$

where $\alpha > 0$.

3.3. Pooling layer

After the output of the convolution layer is nonlinearized by the activation function, a pooling layer is applied. The output feature map is downsampled by a pooling function, which maps patches of the output feature map to one summary statistic of the nearby outputs. This is often done by taking the maximum value from the outputs of the patch, max pooling, or by averaging the outputs of that particular subimage [19], average pooling. Similar to the window sliding procedure in the convolution layer, the patches are also chosen by sliding a window over the feature map. However, the pooling window is mostly smaller, usually $2x2$, than a convolutional window. Additonally, a stride of two is often used, which means that the window slides with two steps across the input. As a result of the stride of two, the feature map is down sampled with a factor of 2 [7]. In figures 3.4 and 3.5 the mapping of the max pooling and the average pooling respectively are shown.

Figure 3.4: Example of max pooling. A window of $2x2$ slides over the input feature map of $4x4$ with a stride of 2. At each position it takes the maximum of the entries within that window. The result is positioned in the output at the same location [13].
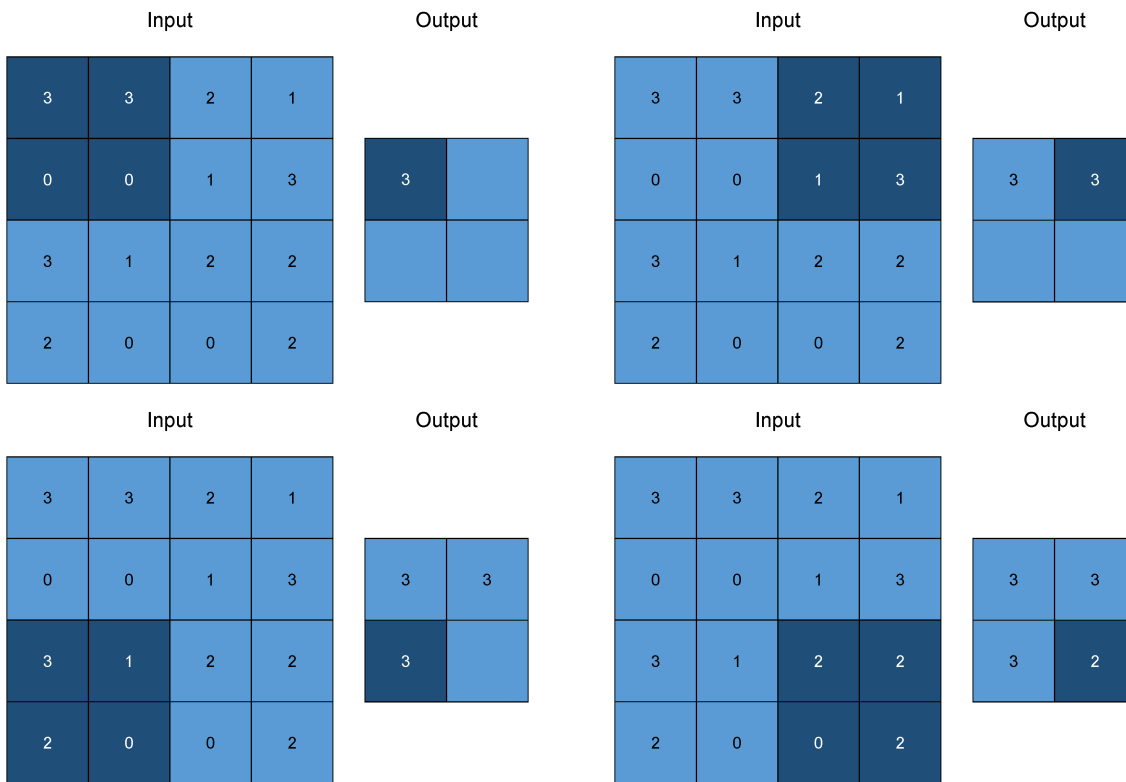

Figure 3.5: Example of average pooling. A window of $2x2$ slides over the input feature map of $4x4$ with a stride of 2. At each position it takes the average of the entries within that window. The result is positioned in the output at the same location [13].
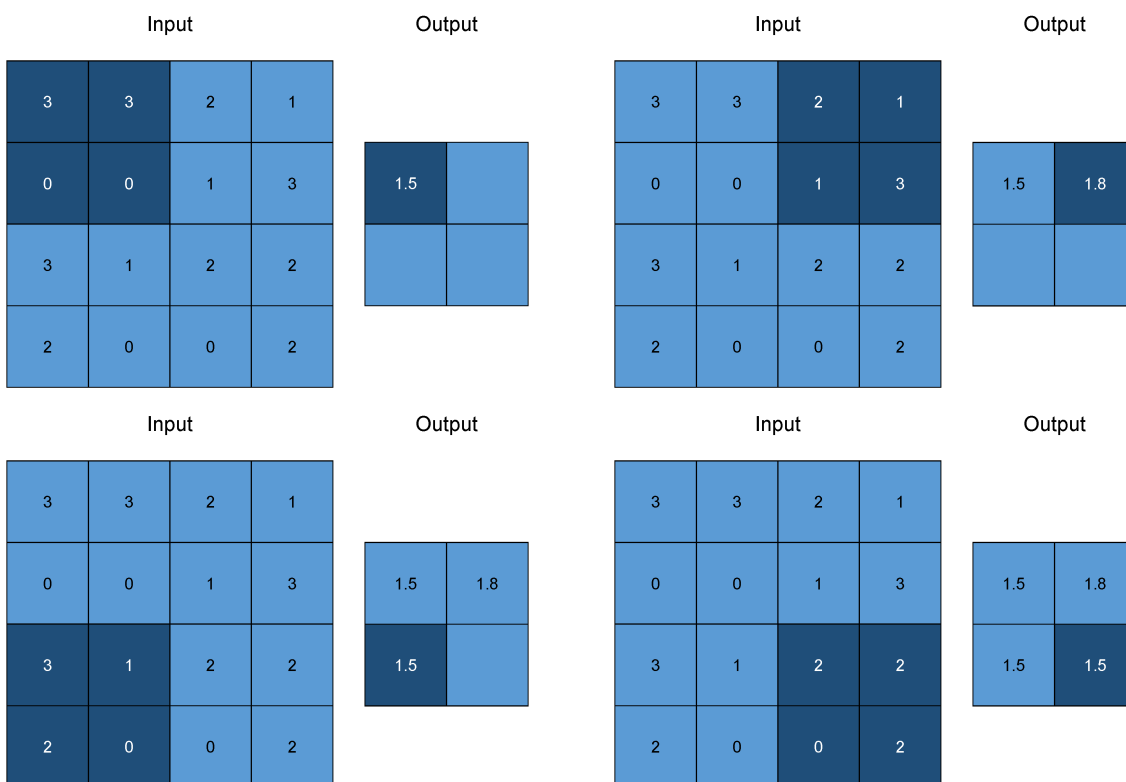
A consequence of the size reduction of the feature maps is that there are less parameters to process for the next convolution layer, which will make it more computational efficient. Furthermore, because the results of the convolution layer will be positioned at the same location in the output feature map as their input features, it can be sensitive to translations in the input. For example, if the object in the input image would be slightly shifted, the output feature maps created by later convolution layers would look different. By using a pooling function, the output layers become approximately invariant to these small movements [17].

3.4. Fully connected layer with end activation function

In order to give a classification to the input image, in the last part of the architecture one or more fully connected layers are applied. These fully connected layers are able to classify the image by combining the features extracted in the previous convolution layers. First, the output of the last activation function or pooling layer $S \in \mathbb{R}^{k \times l \times m}$ is stretched into a $k \cdot l \cdot m$ sized vector. Second, the stretched outputs will be connected via weights and an activation function to all of the nodes in the fully connected layer, where the amount of nodes can be any preferred number. Finally, to output a probability that gives an indication whether the image belongs to a certain class or not, a final activation function is applied after the last layer. For this final transformation it is common to use a sigmoid function or a softmax function. The sigmoid function is especially suitable for binary classification tasks (for example, is there a cat in the input image "yes" or "no") and if we let $\mathbf{x} \in \mathbb{R}^2$ it can be written as follows,

$$f(x_i) = \frac{1}{1 + e^{-x_i}}. \tag{3.6}$$

For a multiclass classification task (for example, is it a cat, a dog, a cow, a pig, a goat or a goose that is displayed on the image?), the softmax function is preferred. In this function, the exponent of every entry is taken and divided by the sum of all exponentiated entries. Letting $\mathbf{x} \in \mathbb{R}^c$, where c represents the number of classes, the softmax function is given by,

$$f(x_i) = \frac{e^{x_i}}{\sum_j^c e^{x_j}}. \tag{3.7}$$

Putting this together, the last part of the convolutional neural network may look like the network that is shown in figure 3.6.
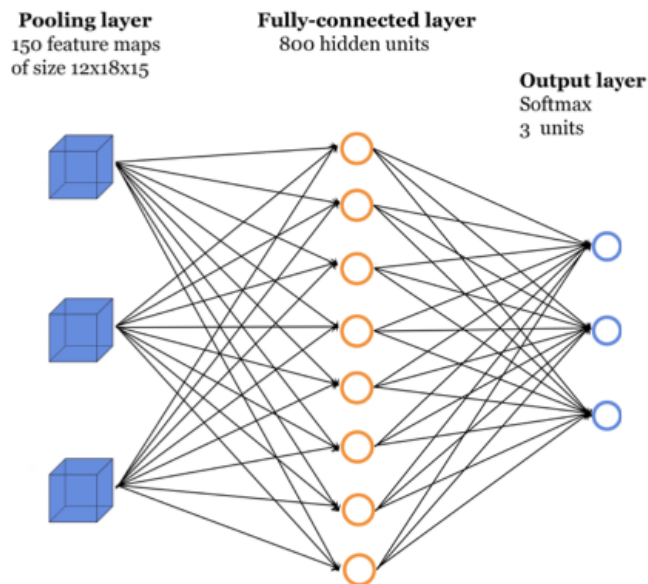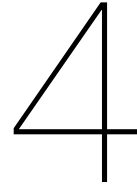

Figure 3.6: Example of the final part of a convolutional neural network [36].

4

Training a convolutional neural network

By training a network we want to obtain that the network can predict the true labels as accurate as possibles. This can be accomplished by minimizing the distance between the predicted outcomes and the true labels. The distance is calculated by the loss, and this loss is minimized using an optimizer.

4.1. Loss function

The objective of a loss function is to measure the distance between the predictions and their corresponding actual outcomes. This can be done in several ways, but the most common used loss function for classification problems is the cross-entropy loss. Where the binary cross-entropy loss is most often used for binary classification tasks, and categorical cross-entropy loss is most often used for tackling multiclass classification problems [7]. Suppose we want to train a dataset of m examples $\{X^{(i)}\}_{i=1}^m \in \mathbb{R}^{V \times W}$ with labeled outputs $\{Y^{(i)}\}_{i=1}^m \in \mathbb{R}^{V \times W}$ and the predictions computed by a network $\{\hat{P}(X^{(i)}; \Theta)\}_{i=1}^m \in \mathbb{R}^{V \times W}$ where Θ are the weights in the network, then the categorical cross-entropy loss is given by,

$$L(\Theta) = -\frac{1}{m} \sum_{i=1}^m \sum_{k=1}^K Y_k^{(i)} \log\left(\hat{P}(X^{(i)}; \Theta)\right), \tag{4.1}$$

where K is the total number of classes. If we take $K = 2$, the binary cross-entropy loss follows from this:

$$L(\Theta) = -\frac{1}{m} \sum_{i=1}^m Y^{(i)} \log\left(\hat{P}(X^{(i)}; \Theta)\right) + \left(1 - Y^{(i)}\right) \log\left(1 - \hat{P}(X^{(i)}; \Theta)\right). \tag{4.2}$$

4.2. Optimizer

To decrease the distance between the predicted output and the true output, the loss can be minimized by an optimizer which adjusts the weights of the network. At the start of the training, the weights of the network are initialized with random values. As a consequence, the predictions will be very far from what the actual output should be and therefore the value of the loss function will be high. To decrease this value, the optimizer updates the weights in the convolution and fully connected layers using backpropagation. This backpropagation operates from the end of the network, the loss function, through all of the layers to the beginning of the network, the first convolution layer. It computes the contribution of every weight to the loss function by applying the chain rule. Each time an example is processed by the network, the weights will be updated in the right direction and the predictions will slowly become more accurate [7]. For the minimization of the loss function,

$$\underset{\Theta}{\text{minimize}} \quad L\left(\hat{P}(X^{(i)}; \Theta), Y^{(i)}\right), \tag{4.3}$$

stochastic gradient descent (SGD) and its variants are the most common optimizers used in deep learning [17].

Stochastic gradient descent proceeds iteratively: a sequence of matrices is computed with the goal of converging to a matrix that minimizes the loss function. SGD executes this task as follows: in the model the weights are stored in matrices, Θ. However, for ease of the explanation we will assume that the weights that need to be updated are now represented by a vector, $\boldsymbol{\theta} \in \mathbb{R}^R$. Let the update of this vector $\Delta\boldsymbol{\theta}$ be small, then when the terms of order $||\Delta\boldsymbol{\theta}||^2$ are left out, the Taylor series expansion can be written as

$$L(\boldsymbol{\theta} + \Delta\boldsymbol{\theta}) \approx L(\boldsymbol{\theta}) + \sum_{r=1}^{R} \frac{\partial L(\boldsymbol{\theta})}{\partial \theta_r} \Delta\theta_r, \tag{4.4}$$

where $\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_r}$ is the partial derivative of the loss function with respect to weight r. We can simplify this equation by using the gradient, with $(\nabla L(\theta))_r = \frac{\partial L(\boldsymbol{\theta})}{\partial \theta_r}$, such that

$$L(\boldsymbol{\theta} + \Delta\boldsymbol{\theta}) \approx L(\boldsymbol{\theta}) + \nabla L(\boldsymbol{\theta})^T \Delta\boldsymbol{\theta}.$$

Now, to minimize the expression we need to take the step $\Delta\boldsymbol{\theta}$ in such a manner that $\nabla L(\boldsymbol{\theta})^T \Delta\boldsymbol{\theta}$ is as negative as possible. Using the Cauchy-Schwartz inequality where for any $f, g \in \mathbb{R}^R$ it holds that

$$|f^T g| \leq ||f||_2 ||g||_2.$$

As a consequence, $-||\Delta\boldsymbol{\theta}||_2 ||\nabla L(\boldsymbol{\theta})^T||_2$ is the most negative as $\nabla L(\boldsymbol{\theta})^T \Delta\boldsymbol{\theta}$ can get, which is true if $\Delta\boldsymbol{\theta} = -\nabla L(\boldsymbol{\theta})^T$. Therefore, the update step will be as follows

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \epsilon \nabla L(\boldsymbol{\theta})^T,$$

where ϵ is called the learning rate, which is a small stepsize as we assumed that $\Delta\boldsymbol{\theta}$ should be small [19]. However, when we need to deal with a large number of examples and many weights that need to be updated, updating can become extremely computationally expensive. In these cases it is common to take a subset of the examples for some $n << m$, a so called minibatch. Instead of calculating the mean of gradients over all examples, with a minibatch the mean of gradients will be calculated over only a few examples [19]. If we write the loss function as

$$L(\boldsymbol{\theta}) = \frac{1}{m} \sum_{i=1}^{m} C_{X_i}(\boldsymbol{\theta})$$

then the algorithm for SGD with using minibatches can be written as given in algorithm 1. In this algorithm we can see that the learning rate changes with every iteration, k. Generally, it decays linearly until iteration τ after which ϵ stays constant:

$$\epsilon_k = \left(1 - \frac{k}{\tau}\right)\epsilon_0 + \frac{k}{\tau}\epsilon_\tau.$$

Variants of the SGD optimizer that are often used are SGD with momentum, RMSProp, RMSProp with momentum, AdaDelta and Adam. However, "the choice of which algorithm to use, at this point, seems to depend largely on the user's familiarity with the algorithm" [17].

Algorithm 1: Stochastic gradient descent

Require: Learning rate schedule $\epsilon_1, \epsilon_2, ...$

Require: Initial parameter $\boldsymbol{\theta}$

k ←1;

while *stopping criterion not met* **do**

    Sample a subset of n examples from the training set $\{X^{(1)}, ..., X^{(m)}\}$ with corresponding targets $Y^{(i)}$;

    Compute gradient estimate: $\hat{g} \leftarrow \frac{1}{n}\nabla_{\boldsymbol{\theta}} \sum_i C_{X_i}(\boldsymbol{\theta})$;

    Apply update: $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \epsilon_k \hat{g}$;

    $k \leftarrow k + 1$;

end

5

U-Net architecture

For the segmentation of the lacunes, we have to deal with medical images, namely MRI images. Within the field of medical image analysis for segmentation problems, the U-Net architecture developed by Ronneberger et al. [38] is the most-well known architecture [30]. Several other papers have successfully applied this U-Net for their segmentation problems as well [11, 12, 14, 24, 49, 57]. Rather than just getting a classification output, with this architecture it is possible to input an image and get a segmented image as output as well.
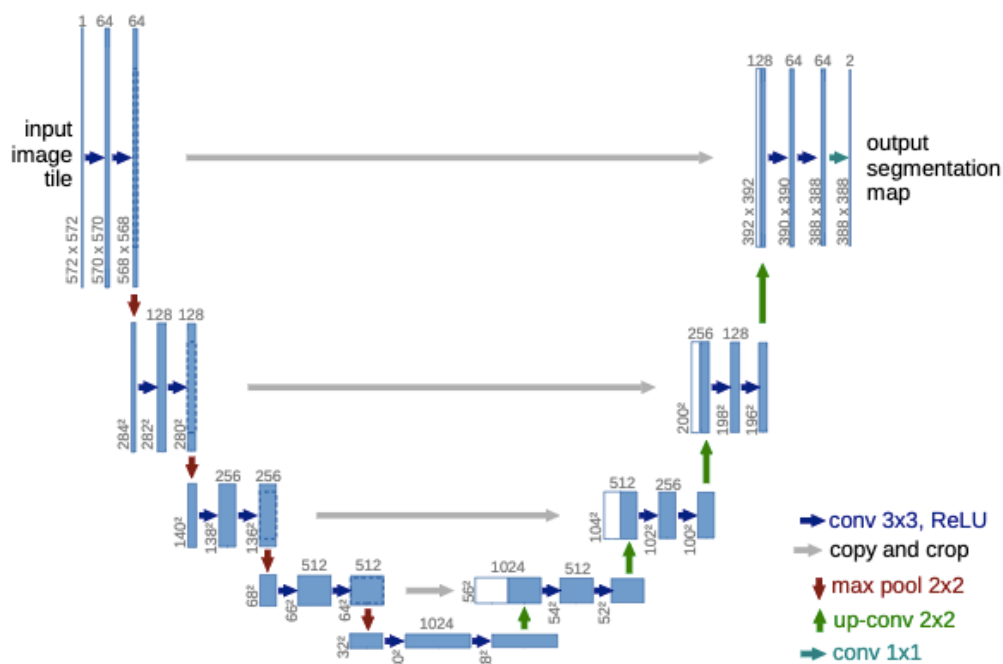

Figure 5.1: U-Net architecture in which convolutions, max pooling, up-convolutions and copy and crop operations are used. The blue bars represent the feature maps with on top the number of feature maps and at its left lower corner the x and y sizes of the feature maps [38]

The U-Net consists of a contracting part, in which feature maps are downsampled to analyze the image, and an expanding part, where using up-sampling a full-resolution segmentation is produced (see figure 5.1). The contracting part consists of 4 blocks containing two $3x3$ convolutions with stride 1, which are both followed by a ReLU activation function. Each block ends with one $2x2$ max pooling operation with stride 2. During the first block of operations 64 kernels are used and with every following contracting block the number of kernels is doubled. The expanding part contains 4 blocks of two $3x3$

17

convolutions with stride 1, each followed by a ReLU. In the second part of the block the feature map is doubled in size by upsampling, after which a $2x2$ up-convolution with stride 1 is applied. Finally, each block ends with a copy and crop operation. With every up-convolution the number of feature maps is halved. At the end of the network another two $3x3$ convolutions with stride 1 are applied, which are each again followed by a ReLU function. At the very end a final $1x1$ convolution with stride 1 is performed to output the segmentation map.

In the upsampling operation the output feature map of the last convolution is taken and expanded with a factor of two. This is done by taking for every pixel its intensity value and copy and paste it to three attached neighboring pixels, as can be seen in figure 5.2. Then, the result is used for an up-convolution of $2x2$ to reduce the feature maps. The copy and crop operations, sometimes also referred to as skip connections, are used to transfer information between the contracting path and the expanding path. As by downsampling localization information can be lost, while it is needed for the segmentation task, a part of the feature maps containing high resolution features from the contracting path is copied and concatenated to its corresponding layer in the expanding path (see figure 5.1). As a consequence, the expanded path is able to better localize pixels, which results in a more accurate segmentation.
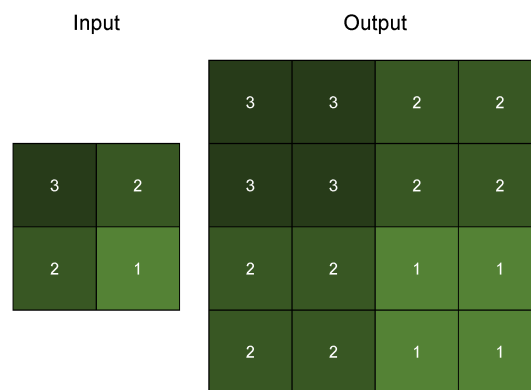

Figure 5.2: Example of upsampling. The value of a pixel is duplicated to three attached neighboring pixels.

6

Previous automated lacune detection and segmentation methods

The first automated lacune detection method was developed by Yokoyama et al. [58], which consists of two steps. In the first step all possible lacune candidates are detected using T2-weighted sequence images of the brain. For this, the images are binarized: every voxel is evaluated based on its intensity and mapped to either 0 or 1. In the paper it is mentioned that the intensity of lacunes can change according to its phase, which can be acute, sub-acute or chronic. To cope with these differences in intensity, the binarization technique is executed 15 times with 15 different thresholds. From the binarized images, candidates are selected based on their area (number of pixels), circularity (measure of how closely the candidate approaches a perfect circle) and gravity center. As this detection step is a very rough procedure, other lesions, brain structures and tissues will be incorrectly identified as candidates as well. Candidates that are misclassified as lacunes are called false positives. In the second phase of the model these false positives are reduced using the T1-weighted brain images. The cerebral parenchyma is extracted as lacunes can never occur outside the cerebral parenchyma or at its edge. Furthermore, most lacunes have high intensity differences compared to their surrounding area while this difference may not be present for false positives. Therefore, the intensity difference between the candidate and its peripheral area is measured and evaluated using the T1-weighted images. The method was able to detect 90.1% of the lacunes with an average of 1.4 false positives per image. Despite of the misclassified candidate reduction step, the results still showed some misclassified candidates. Analyzing these false positives, the researchers found that especially the edge of the cerebral parenchyma, high-signal regions near the ventricles, a part of the cerebral ventricle and perivascular spaces were misclassified as lacunes.

Uchiyama et al. [44] adopted the method of Yokoyama et al. [58]. However, in order to reduce the number of false positives, they replaced the approach in the second step with another false positive reduction approach. In this new step, false positives are eliminated based on their location, signal intensity difference and shape of the structure. The location is used, because lacunes occur within cerebral vessel regions and thus candidates on the periphery of the cerebral region are more likely to be false positives. Next, signal intensity difference is evaluated, as again lacunes show an intensity difference with their surrounding area and false positives may not. Finally, false positives are eliminated based on the shape of the structure, as lacunes are more likely to be of nodular shape and some false positives such as the cerebral sulcus will have a more linear shape. After misclassified candidates are eliminated using these features with cut-off thresholds, the remaining candidates are subdivided using a hyperplane into lacunes and false positives. The method shows improved results with detecting 96,8% of the lacunes and an average of 0.76 false positive per slice. However, still not all false positives were removed with the new approach. The types of remaining false positives included a part of the cerebral sulcus, a part of the cerebral ventricle and perivascular spaces.

To reduce specifically these false positives, Uchiyama et al. changed their method in a later paper [43]. That is, for the detection of candidates the approach of Yokoyama et al. [58] is used and for the first false positive elimination round the false positives are again removed by looking at the features location, signal intensity difference and shape using cut-off thresholds. The second round of elimination

is executed by applying neural networks instead of a separating hyperplane. The first neural network consists of 3-layers with having the features of above as input and the likelihood of a lacune as output. Thereafter, three parallel neural networks are applied to attack specific classes of false positives: classifier A distinguishes lacunes from parts of the cerebral sulcus, classifier B distinguishes lacunes from parts of the cerebral ventricle and classifier C distinguishes lacunes from perivascular spaces. This method identified 96.8% of the lacunes with an average of 0.30 false positive per scan.

Because Uchiyama et al. experienced that especially the differentiation between lacunes and perivascular spaces is challenging, they focused more on addressing this aspect in the methods that followed. In [45] all lesions are first enhanced and segmented. Then, features of location, size, signal intensity difference and degree of irregularity (a measure of how much the lesion deviates from a perfect circle) are used as input for a 3-layer neural network to classify the lesions. With this method they showed that size greatly contributes in distinguishing between lacunes and perivascular spaces and that location features are useful for differentiation between lacunes located in the periphery of the lateral ventricle and perivascular spaces. Furthermore, 93.3% of the lacunes were detected and 94.5% of the predictions were correct. In [46], the exact same method was applied, but with an extra step to construct a fused image. That is, on a T2-weighted image it can be difficult to distinguish lacunes from perivascular spaces. Therefore, since perivascular spaces have an elongated shape which is often only visible in 3D, a 3D vessel region is added to the T2-weighted image. As a result, from this fused image it is better visible whether a round or oval shaped lesion on a T2-weighted image is a perivascular space or a lacune.

Seven years after the first method in [44] of Uchiyama et al., they extended this original method with an extra false positive reduction step, which focused primarily on the differentiation between lacunes and perivascular spaces [47]. This means that the selection of candidates and the false positive elimination approach is similar to their method in [44]. However, after the two phased elimination step consisting of reduction using location, signal intensity difference and shape of the structure and grouping the false positives and lacunes using a hyperplane, another reduction phase was added. In order to delete extra information while maintaining important information that is useful for the differentiation between lacunes and perivascular spaces, a small area around a candidate is used to create an eigenspace. New unseen test data is then projected to this created eigenspace and evaluated for being a lacune or a false positive. As a result of this addition, 96.8% of the lacunes were detected while showing an average of 0.47 false positive per image.

Wang et al. developed a method to segment three lesions at once which also included lacunes [50]. Their procedure starts with delineating brain tissues based on T1 weighted images after which hyperintense regions are segmented using FLAIR and T1 weighted images. These hyperintense regions are then used to segment the lacunes with T1 weighted images, T2 weighted images and FLAIR images. To detect lacunes near white matter hyperintensities, the white matter hyperintensities are first dilated. After this dilation, voxels can be identified as a lacune by comparing the voxel intensity with the average intensity within the white matter hyperintensity region. For segmentation of lacunes in subcortical structures, the intensity of the voxels is compared with the averaged intensity of the specific subcortical structure to determine whether a voxel belongs to a lacune or not. With this method 80.6% of the lacunes were segmented while having an average of 0.06 false positive per scan.

The most recent method for the detection of lacunes is the two-stage method of Ghafoorian et al. [16]. In the first stage candidates are detected using a 7-layer CNN architecture. For this, small sub-images are used to capture a local neigborhood around each candidate. The second stage was created to eliminate the false positives with a 9-layer CNN. As location can be a discriminative factor for the differentiation between lacunes and perivascular spaces, explicit location features are added to this second CNN. In addition to this, the network is fed with three different sized images of a same candidate, such that the biggest sized image contains more neighborhood information and as a consequence also more information about the location than the smallest more detailed sized image. 97.4% of the lacunes were detected with this method with an average of 0.13 false positive per slice.

7

Segmentation challenges

7.1. Class imbalance problem

Compared to the entire brain, lacunes are very small lesions. This means that on a slice of a brain MRI image, only a few pixels will be occupied by the lacune. As a result, with the smallest size of the lesion, 3 mm, we have a lacune:non-lacune pixel ratio of 36:900 in our MRI images (see figure 7.1). This is called class imbalance and may cause a problem when we want to train a network. If there are many more samples of one class, which is called the majority class, than there are of the other class, the minority class, the learning process often gets stuck in a local minimum. When a network needs to learn from class imbalanced data, it will over-classify the majority class, the non-lacune pixels, because of its increased prior probability. As a consequence, the minority class, the lacune pixels, will often be misclassified [23]. So, what can we do to prevent this problem?
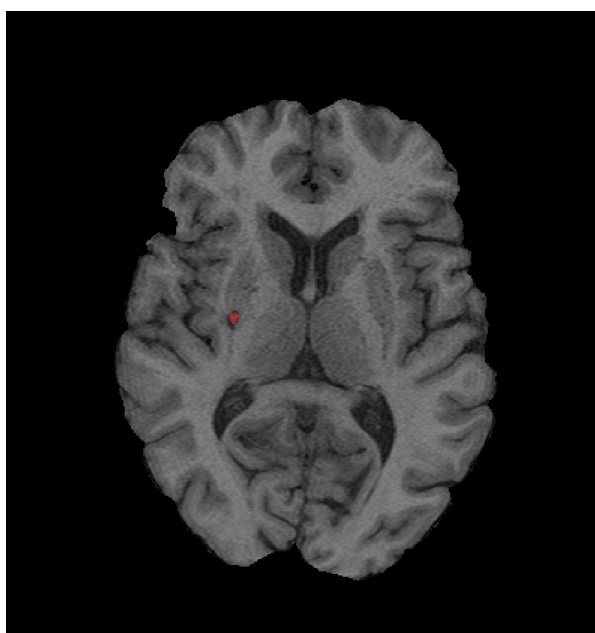

Figure 7.1: Lacunes occupy only a small amount of the pixels on an MRI image.

This problem might, however, be prevented. In recent years several loss functions were designed to handle this class imbalance problem [41]. Among them are the weighted cross-entropy (WBCE) loss and the Dice loss (DL). The weighted cross-entropy loss was first used in [38] where a weight, w, is applied to the minority class to give more importance to those pixels during training. With $y_l^{(i)}$ being the lth pixel value of the ground truth image $Y^{(i)}$ and $\hat{p}_l^{(i)}$ being the lth pixel value of the predicted

probabilistic map $\hat{P}^{(i)}$, the weighted loss of the binary cross-entropy loss given by equation 4.2 is as follows,

$$WBCE = -\frac{1}{L}\sum_{l=1}^{L} wy_l^{(i)} \log\left(\hat{p}_l^{(i)}\right) + \left(1 - y_l^{(i)}\right)\log\left(1 - \hat{p}_l^{(i)}\right), \tag{7.1}$$

where in [41] for the weight w, $\frac{L - \sum_l \hat{p}_l^{(i)}}{\sum_l \hat{p}_l^{(i)}}$ was used.

The Dice loss is based on the Dice coefficient which measures the relative overlap between the ground truth and the prediction. It was first used by Milletari et al. as Dice loss in [32] and reads as

$$DL = 1 - \frac{\sum_{l=1}^{L} \hat{p}_l^{(i)} y_l^{(i)} + \epsilon}{\sum_{l=1}^{L} \hat{p}_l^{(i)} + y_l^{(i)} + \epsilon} - \frac{\sum_{l=1}^{L}\left(1 - \hat{p}_l^{(i)}\right)\left(1 - y_l^{(i)}\right) + \epsilon}{\sum_{l=1}^{L} 2 - \hat{p}_l^{(i)} - y_l^{(i)} + \epsilon}, \tag{7.2}$$

where ϵ is a small number to ensure stability by avoiding divisions by 0 [41]. Since the Dice loss is focusing on the overlap rather than evaluating individual pixels, this may help tackling the class imbalance as well.

A final, also promising, approach is instead of using the entire image, splitting the image up in patches. This is also done by Ghafoorian et al. [16]. By using patches, we can control the amount of positive and negative samples the network sees, which are the lacunes and non-lacunes, respectively. This results in a more balanced dataset.

7.2. Differentiation with perivascular spaces

Another challenge that we might face during lacune segmentation is the differentiation with perivascular spaces. As we have seen in section 2.2, lacunes and perivascular spaces can look very similar. They have the same intensity on both T1-weighted and T2-weighted images and they can have a similar round or ovoid shape when imaged perpendicular to the course of the vessel. In chapter 6 we saw that this indeed can be a challenge for a network. In the papers from Yokoyama et al. and Uchiyama et al. [44, 58] where detection methods were developed, among the false positives also perivascular spaces were found. As a consequence, in the methods that followed [16, 45–47] more attention was given to reduce the misclassification of these perivascular spaces. Therefore, when designing a network for our lacune segmentation problem this differentiation challenge should also be taken into account. There are several options to address this challenge.

Because perivascular spaces are generally smaller in size (< 3 mm) than lacunes ($3 - 15$ mm), this can be a relevant determinant for the discrimination between these two lesions. It might be advantageous to apply a comparable approach as was used in [3], where constraints are put in the loss function on the size of the candidate that should be segmented. Particularly, using the loss function we can constrain the size of the candidates to be between 3 and 15 mm to differentiate lacunes from perivascular spaces.

In [16, 45] it was found that location is an important factor for the discrimination between lacunes and perivascular spaces. Using the idea proposed in [61] of adjusting the loss function to make sure that a model is able to use the location of a salient object, we could do something similar for our lacune segmentation task. That is, in order to reduce the false positives beforehand, locations where lacunes can never occur could be punished via the loss function. This location awareness might then also help to differentiate between lacunes and perivascular spaces.

If patches would be used as input image to overcome the data imbalance problem, it might be useful to use different sizes of patches as was also done by Ghafoorian et al. [16] and by Xu et al. [57] as it makes the network more location-aware. In these papers three sizes are used for the input patch, with the idea that the bigger patches are able to capture more information about the location while the smaller patches contain more detailed information to accurately catch the lesion boundary. Taking the information of the different sized patches together, would then result in a more accurate segmentation.

8
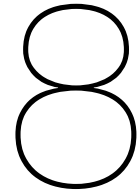
Conclusion

To assist in achieving the aim of developing an automated lacune segmentation method, the literature study provided a background on lacunes, previous detection an segmentation methods and segmentation challenges. In addition, it gave an introduction on convolutional neural networks and highlighted a promising architecture, U-Net.

Lacunes are lesions that are a result of the cerebral small vessel disease. cSVD is a disease that can have an influence on both physical and psychological abilities. Since little is known about the cause of the disease, detecting the lacunes might help to give more insight about the cause of cSVD.

We have seen that the U-Net architecture was successfully used in several other lesion segmentation methods. Therefore, it might also be promising to use this architecture as a basis for our segmentation task. However, other choices for the hyperparameters (e.g. activation function, optimizer) than the ones that are used in the original U-Net might work better for the segmentation of lacunes. Hence, to optimize our network, other options for the hyperparameters should be considered as well.

Furthermore, when developing a lacune segmentation method, we might encounter problems regarding the class imbalance and the differentiation with perivascular spaces. Several suggestions have been discussed to tackle these problems. For the class imbalance, a weighted cross-entropy loss or Dice loss might help to overcome the imbalance. Additionally, the network might benefit from using patches rather than full-sized images, as it helps with controlling the number of positive and negative samples it sees. For the differentiation with perivascular spaces, the loss function can be used to put a constraint on the size of the candidate or to punish locations where lacunes can never occur. Furthermore, if patches are used, the architecture can be given different sized patches to make the network more location-aware.

In conclusion, the U-Net can be a promising architecture to use as a basis for our problem, but it needs to be fine-tuned for it to also be applicable to lacunes.

9
Research Proposal

From the literature study it became clear that an automated lacune segmentation method can be beneficial. Therefore the main research question we would like to answer with this thesis is:

How can we develop an automated method that is able to segment lacunes of presumed vascular origin in brain MRI scans?

To answer this question, the U-Net architecture from [38] will be used as a starting point, as it has been proven to obtain good segmentation results for other lesions [11, 12, 14, 24, 49, 57]. Almost the complete structure of the architecture will be replicated. That is, the contracting path will also contain blocks of two $3x3$ convolutions, each followed by a ReLU activation function, and one $2x2$ max pool with stride 2. The expanding path will consist of blocks with two $3x3$ convolutions, each followed by ReLU, and an upsampling operation together with a $2x2$ up-convolution. Furthermore, copy and crop connections are used between the two paths to transfer information and at the end of the network another two $3x3$ convolutions are applied after which a final $1x1$ convolution follows. However, where the original U-Net architecture uses four max pooling layers, this might cause problems for the lacune segmentation task.
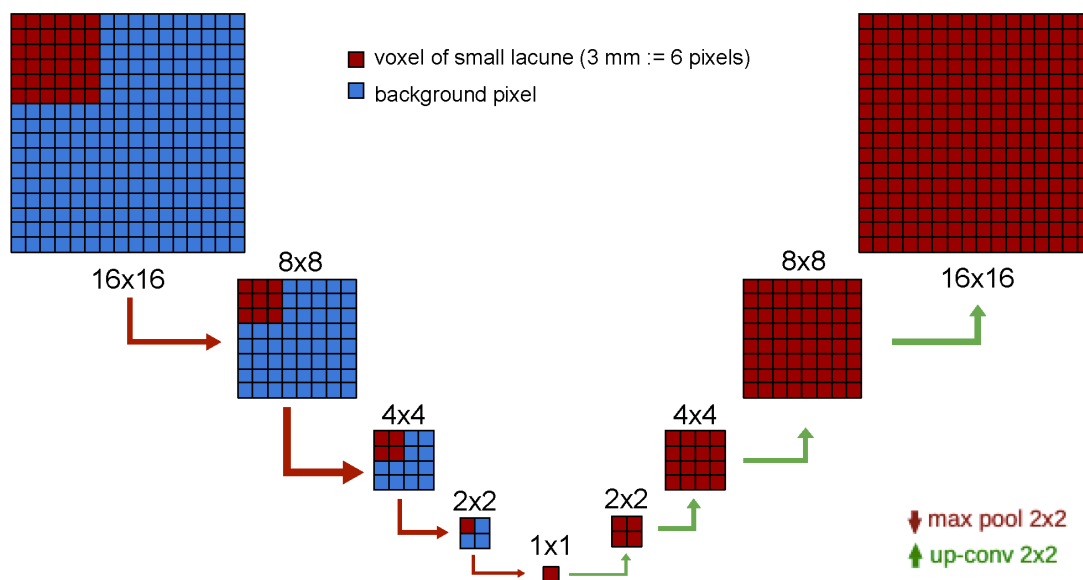

Figure 9.1: Graphical explanation of the effect of max pooling layers on lacunes. When a small lacune of 3mm is downsized four times, it will result in a representation that is more than two times bigger than the original. Here a pixel is 0.5 mm in height and width.

25

Since lacunes are very small in size, downsizing feature maps causes the lesion to almost vanish after only a few pooling layers. From figure 9.1, we can see the effect of applying four pooling layers to the smallest lacune of 3 mm. Here, one pixel has a size of 0.5 mm for both its height and width, which means that a small lacune will occupy $6x6$ pixels of the $16x16$ image. After three pooling layers there is only one pixel left with information about the lacune. When we apply another pooling layer, there will still be one pixel containing the lacune information. This is causing the problem, because if we use the upsampling operation four times, the original lacune is now represented by $16x16$ pixels. The lacune has increased by a factor of almost 3. Therefore it is decided to not use more than three pooling layers for the lacune segmentation task, starting with two max pool operations.

With the U-Net architecture as a basis for the lacune segmentation problem, the aim will be to adapt this architecture such that accurate segmentation results can be obtained. To achieve this, the method should be optimized by considering different choices of hyperparameters. Therefore, the following subquestion is formulated:

Which options of general hyperparameters (e.g. activation function, optimizer) should be chosen to obtain the most accurate results?

Where other options for the activation functions are leaky ReLU and ELU, which were discussed in section 3.2. A good choice for the optimizer would be Adam [25] or AdaDelta [59], as they have been used more often in combination with the U-Net [11, 12, 14, 24, 49].

Furthermore, from the literature study in chapter 7, we have seen that there are at least two challenges we may encounter when applying a segmentation method. In order to overcome these problems, the two subquestions stated below are proposed:

Which approach should be used to tackle the data imbalance problem?

How can we make sure that the model is able to differentiate between lacunes and perivascular spaces?

Where choices for tackling the data imbalance problem, as suggested in section 7.1, might be, choosing for a weighted cross-entropy loss or Dice loss, or making use of patches instead of entire images. In section 7.2 ideas are posed to overcome the differentiation issue. That is, we could try to use the loss function to put on the size of a candidate or to penalize locations where lacunes never occur. Moreover, when patches are used, it might be beneficial to use different sizes to make the network more location-aware.

Finally, it would also be interesting to investigate how the developed lacune segmentation method can be applied to other data. This way the approach can be made more generalizable. To assess this issue the next subquestion is defined:

Can we make the model applicable to another dataset as well?

Acronyms

BBB blood-brain barrier. 3

CNN convolutional neural network. 8

CSF cerebrospinal fluid. 4

cSVD cerebral small vessel disease. 3

ELU exponential linear unit. 10

FLAIR fluid-attenuated inversion recovery. 4

MRI magnetic resonance imaging. 3, 4

ReLU rectified linear unit. 10

RSS Rotterdam scan study. 5

SGD stochastic gradient descent. 14

Glossary

Amyloid

Aggregate of proteins. 3

Arteriole

Small blood vessel that branches out from an artery into the capillaries. 3

Biomarker

Measurable indicator of the presence or severity of a disease state. 3

Brain parenchyma

Functional brain tissue that consists of neurons and glial cells. 3

Capillary

The smallest blood vessel in the body. They supply blood to the surrounding tissues. 3

Cardiovascular system

System that consists of the heart and blood vessels. 3

Cerebellum

Area at the back and bottom of the brain, which is associated with movement and coordination. 6

Cerebral sulcus

Grooves on the surface of the brain. 19

Cerebrospinal fluid

Clear, colorless body fluid found in the brain and spinal cord that acts as a cushion or buffer. 4

False positive

A positive predicted outcome that should have been negative. 19

Fluid-attenuated inversion recovery sequence

A particular setting of radiofrequency pulses and field gradients that influences the appearance of an image. 4

Glial cells

Supportive cell in the nervous system. They surround and support neurons. 3

Ground truth

A measurement that approaches the exact outcome, but is assumed to be the exact outcome. 6

Haemorrhage

Blood escaping from the cardiovascular system caused by damaged blood vessels. 4

Hyperintense

Brighter appearance of an abnormality or structure than the structures it is compared to. 4

Hyperparameter

Chosen parameter whose value is used to control the learning process. 23

Hypointense

Darker appearance of an abnormality or structure than the structures it is compared to. 4

Infarct

Resulting lesion from tissue death due to inadequate blood supply. 4

Ischaemic stroke

Blood supply to a part of the brain is decreased, leading to dysfunction of the brain tissue. 3

Kernel

Filter that is able to extract features. 8

Microbleed

Lesion that results from cerebral small vessel disease. 1

MRI sequence

A particular setting of radiofrequency pulses and field gradients that influences the appearance of an image. 4

Neuroimaging

The process of producing images of the structure or activity of the nervous system. 3

Neuron

Nerve cell that communicates with other cells with electrical impulses. It is the basic unit of the nervous system. 3

Pathogen

Any disease-producing agent such as bacteria, a virus or other microorganisms. 3

Perivascular space

Lesion that results from cerebral small vessel disease. 1

Segmentation

The division of a visual input into segments. 1

Semipermeable

Some molecules or ions can pass through, but others can not. 3

Stride

Stepsize of a sliding window. 10

Subcortical

Part of the brain that is located below the cerebral cortex, the outer part of the brain. 3

T1-weighted sequence

A particular setting of radiofrequency pulses and field gradients that influences the appearance of an image. 5

T2-weighted sequence

A particular setting of radiofrequency pulses and field gradients that influences the appearance of an image. 19

Ventricle

CSF-filled space in the brain. 19

Venule

Small blood vessel that transports the regained blood from the capillaries to the veins. 3

White matter hyperintensity

Lesion that results from cerebral small vessel disease. 1

Bibliography

[1] N.J. Abott, A.A.K. Patabendige, D.E.M. Dolman, S.R. Yusof, and D.J. Begley. Structure and function of the blood-brain barrier. *Neurobiology of Disease*, 37(1):13–25, 2010. doi: 10.1016/j.nbd. 2009.07.030.

[2] J. Bamford, P. Sandercock, L. Jones, and C. Warlow. The natural history of lacunar infarction: The oxfordshire community stroke project. *Stroke*, 18(3):545–551, 1987. doi: 10.1161/01.str. 18.3.545.

[3] M. Bateson, J. Dolz, H. Kervadec, H. Lombaert, and I.B. Ayed. Constrained domain adaptation for segmentation. In *Proceedings of the 22nd Medical Image Computing and Computer Assisted Intervention*, pages 326–334, Shenzhen, China, 2019. doi: 10.1007/978-3-030-32245-8_37.

[4] J.M. Biesbroek, H.J. Kuijf, Y. van der Graaf, K.L. Vincken, A. Postma, W.P.T.M. Mali, G.J. Biessels, and M.I. Geerlings. Association between subcortical vascular lesion location and cognition: A voxel-based and tract-based lesion-symptom mapping study. the smart-mr study. *PLOS ONE*, 8 (4), 2013. doi: 10.1371/journal.pone.0060541.

[5] A. Biffi and S.M. Greenberg. Cerebral amyloid angiopathy: A systematic review. *Journal of Clinical Neurology*, 7(1):1–9, 2011. doi: 10.3988/jcn.2011.7.1.1.

[6] A. Charidimou, Q. Gang, and D.J. Werring. Sporadic cerebral amyloid angiopathy revisited: recent insights into pathophysiology and clinical spectrum. *Journal of Neurosurg Psychiatry*, 83(2):124–137, 2012. doi: 10.1136/jnnp-2011-301308.

[7] F. Chollet. *Deep Learning with Python.* Manning Publications Co., Shelter Island, New York, 2018.

[8] D.A. Clevert, T. Unterthiner, and S. Hochreiter. Fast and accurate deep network learning by exponential linear units (ELUs), 2016. URL arXiv:1511.07289.

[9] E. Cuadrado-Godia, P. Dwivedi, S. Sharma, A. Ois Santiago, J. Roquer Gonzales, M. Balcells, J. Laird, M. Turk, H.S. Suri, A. Nicolaides, L. Saba, N.K. Khanna, and J.S. Suri. Cerebral small vessel disease: A review focusing on pathophysiology, biomarkers, and machine learning strategies. *Journal of Stroke*, 20(3):302–320, 2018. doi: 10.5853/jos.2017.02922.

[10] K.F. de Laat, A.M. Tuladhar, A.G. van Norden, D.G. Norris, M.P. Zwiers, and F.E. de Leeuw. Loss of white matter integrity is associated with gait disorders in cerebral small vessel disease. *Brain*, 134(1):73–83, 2011. doi: 10.1093/brain/awq343.

[11] J. Dolz, I.B. Ayed, and C. Desrosiers. Dense multi-path u-net for ischemic stroke lesion segmentation in multiple image modalities. In *Proceedings of the 4th International MICCAI Brainlesion Workshop*, pages 271–282, Granada, Spain, 2018. doi: 10.1007/978-3-030-11723-8_27.

[12] F. Dubost, G. Bortsova, H. Adams, M.A. Ikram, W.J. Niessen, M.W. Vernooij, and M. de Bruijne. Gp-unet: Lesion detection from weak labels with a 3d regression network. In *Proceedings of the 20th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 214–212, Québec City, Canada, 2017. doi: 10.1007/978-3-319-66179-7_25.

[13] V. Dumoulin and F. Visin. A guide to convolution arithmetic for deep learning, 2016. URL arXiv: 1603.07285.

[14] M.T. Duong, J.D. Rudie, J. Wang, L. Xie, S. Mohan, J.C. Gee, and A.M. Rauschecker. Convolutional neural network for automated flair lesion segmentation on clinical brain mr imaging. *American Journal of Neuroradiology*, 40(8):1282–1290, 2019. doi: 10.3174/ajnr.A6138.

[15] C.M. Fisher. Lacunes: Small, deep cerebral infarcts. *Neurology*, 15(8):774–784, 1965. doi: 10.1212/wnl.15.8.774.

[16] M. Ghafoorian, N. Karssemeijer, T. Heskes, M. Bergkamp, J. Wissink, J. Obels, K. Keizer, F.e. de Leeuw, B. van Ginneken, E. Marchiori, and B. Platel. Deep multi-scale location-aware 3d convolutional neural networks for automated detection of lacunes of presumed vascular origin. *NeuroImage: Clinical*, 14:391–399, 2017. doi: 10.1016/j.nicl.2017.01.033.

[17] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, Cambridge, Massachusetts, 2016. http://www.deeplearningbook.org.

[18] L.L. Herrmann, M. Le Masurier, and K.P. Ebmeier. White matter hyperintensities in late life depression: a systematic review. *Journal of Neurology, Neurosurgery Psychiatry*, 79(6):619–624, 2008. doi: 10.1136/jnnp.2007.124651.

[19] F.C. Higham and Higham D.J. Deep learning: An introduction for applied mathematicians. *SIAM Review*, 61(4):860–891, 2019. doi: 10.1137/18M1165748.

[20] M.A. Ikram, A. van der Lugt, W.J. Niessen, P.J. Koudstaal, G.P. Krestin, A. Hofman, D. Bos, and M.W. Vernooij. The rotterdam scan study: design update 2016 and main findings. *European Journal of Epidemiology*, 30(12):1299–1315, 2015. doi: 10.1007/s10654-015-0105-7.

[21] M.A. Ikram, G. Bruselle, M. Ghanbari, A. Goedegebure, M.K. Ikram, M. Kavousi, B.C.T. Kieboom, C.C.W. Klaver, R.J. de Knegt, A.L. Luik, T.E.C. Nijsten, R.P. Peeters, F.J.A. van Rooij, B.H. Stricker, A.G. Uitterlinden, M.W. Vernooij, and T. Voortman. Objectives, design and main findings until 2020 from the rotterdam study. *European Journal of Epidemiology*, 2020. doi: 1oi.org/10.1007/s10654-020-00640-5.

[22] K. Jarrett, K. Kavukcuoglu, M.A. Ranzato, and Y. LeCun. What is the best multi-stage architecture for object recognition? In *IEEE 12th International Conference on Computer Vision*, pages 2146–2153, 2009. doi: 10.1109/ICCV.2009.5459469.

[23] J.M. Johnson and T.M. Khoshgoftaar. Survey on deep learning with class imbalance. *Journal of Big Data*, 6(27), 2019. doi: doi.org/10.1186/s40537-019-0192-5.

[24] R. Karthik, U. Gupta, A. Jha, R. Rajalakshmi, and R. Menaka. A deep supervised approach for ischemic lesion segmentation from multimodal MRI using fully convolutional network. *Applied Soft Computing Journal*, 84, 2019. doi: 0.1016/j.asoc.2019.105685.

[25] D.P. Kingma and L.J. Ba. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations*, San Diego, United States of America, 2015.

[26] P. Klarenbeek, R.J. van Oostenbrugge, R.P.W. Rouhl, I.L.H. Knottnerus, and J. Staals. Ambulatory blood pressure in patients with lacunar stroke: Association with total mri burden of cerebral small vessel disease. *Stroke*, 44(11):2995–2999, 2013. doi: 10.1161/STROKEAHA.113.002545.

[27] R.P. Kloppenborg, P.J. Nederkoorn, A.M. Grool, L.J.L. De Cocker, W.P.T.M. Mali, Y. van der Graaf, and M.I. Geerlings. Do lacunar infarcts have different aetiologies? risk factor profiles of lacunar infarcts in deep white matter and basal ganglia: The second manifestations of arterial disease-magnetic resonance study. *Cerebrovascular Diseases*, 43(3-4):161–168, 2017. doi: 10.1159/000454782.

[28] Y. Ling and H. Chabriat. Progression of mri markers in cerebral small vessel disease: Sample size considerations for clinical trials. *Journal of Cerebral Blood Flow Metabolism*, 36(1):228–240, 2016. doi: 10.1038/jcbfm.2015.113.

[29] Y. Ling and H. Chabriat. Incident cerebral lacunes: A review. *Journal of Cerebral Blood Flow Metabolism*, 40(5):909–921, 2020. doi: 10.1177/0271678X20908361.

[30] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.W.M. van der Laak, B. van Ginneken, and C.I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017. doi: 10.1016/j.media.2017.07.005.

[31] A.L. Maas, A.Y. Hannun, and A.Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, United States of America, 2013.

[32] F. Milletari, N. Navab, and S.A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *Proceedings of the 4th International Conference on 3D Vision*, pages 565–571, Stanford, United States of America, 2016. doi: 10.1109/3DV.2016.79.

[33] V. Nair and G.E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning*, pages 807–814, Haifa, Israel, 2010.

[34] L. Pantoni. Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. *The Lancet Neurology*, 9(7):689–701, 2010. doi: 10.1016/S1474-4422(10)70104-6.

[35] M. Pasi, G. Boulouis, P. Fotiadis, E. Auriel, A. Charidimou, K. Haley, A. Ayres, K.M. Schwab, J.N. Goldstein, J. Rosand, A. Viswanathan, L. Pantoni, S.M. Greenberg, and M.E. Gurol. Distribution of lacunes in cerebral amyloid angiopathy and hypertensive small vessel disease. *Neurology*, 88 (23):2162–2168, 2017. doi: 10.1212/WNL.0000000000004007.

[36] A. Payan and G. Montana. Predicting alzheimer's disease: a neuroimaging study with 3D conolutional neural networks. In *Proceedings of the 4th International Conference on Pattern Recognition Applications and Methods*, 2015.

[37] A. Pogessi, G. Pracucci, H. Chabriat, T. Erkinjuntti, F. Fazekas, A. Verdelho, M. Hennerici, P. Langhorne, J. O'Brien, P. Scheltens, M.C. Visser, M. Crisby, G. Waldemar, A. Wallin, D. Inzitari, and L. Pantoni. Urinary complaints in nondisabled elderly people with age-related white matter changes: the leukoaraiosis and disability (ladis) study. *Journal of the American Geriatrics Society*, 56(9):1638–1643, 2008. doi: 10.1111/j.1532-5415.2008.01832.x.

[38] O. Ronneberger, P. Fisher, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the 18th Medical Image Computing and Computer Assisted Intervention*, pages 234–241, Munich, Germany, 2015. doi: 10.1007/978-3-319-24574-4_28.

[39] E.E. Smith and S.M. Greenberg. Beta-amyloid, blood vessels and brain function. *Stroke*, 40(7): 2601–2606, 2009. doi: 10.1161/STROKEAHA.108.536839.

[40] E.E. Smith and J.M. Lee. Lacunes: Black holes in our understanding of cerebral amyloid angiopathy. *Neurology*, 88(23):2158–2159, 2017. doi: 10.1212/WNL.0000000000004017.

[41] C.H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M.J. Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Proceedings of the 3rd Deep Learning in Medical Image Analysis and 7th Multimodal Learning for Clinical Decision Support*, pages 240–248, Québec City, Canada, 2017. doi: 10.1007/978-3-319-67558-9_28.

[42] R. Topakian, T.R. Barrick, F.A. Howe, and H.S. Markus. Blood-brain barrier permeability is increased in normal-appearing white matter in patients with lacunar stroke and leucoaraiosis. *Journal of Neurology, Neurosurgery Psychiatry*, 81(2):192–197, 2010. doi: 10.1136/jnnp.2009.172072.

[43] Y. Uchiyama, R. Yokoyama, H. Ando, T. Asano, H. Kato, H. Yamakawa, H. Yamakawa, T. Hara, T. Iwama, H. Hoshi, and Fujita. Improvement of automated detection method of lacunar infarcts in brain MR images. In *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1599–1602, Lyon, France, 2007. doi: 10.1109/IEMBS.2007.4352611.

[44] Y. Uchiyama, R. Yokoyama, H. Ando, T. Asano, H. Kato, H. Yamakawa, H. Yamakawa, T. Hara, T. Iwama, H. Hoshi, and H. Fujita. Computer-aided diagnosis scheme for detection of lacunar infarcts on MR images. *Academic Radiology*, 14(12):1554–1561, 2007. doi: 10.1016/j.acra. 2007.09.012.

[45] Y. Uchiyama, T. Kunieda, T. Asano, H. Kato, T. Hara, K. Masayuki, T. Iwama, H. Hoshi, Y. Kinosada, and H. Fujita. Computer-aided diagnosis scheme for classification of lacunar infarcts and enlarged virchow-robin spaces in brain MR images. In *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3908–3911, Vancouver, Canada, 2008. doi: 10.1109/IEMBS.2008.4650064.

[46] Y. Uchiyama, T. Asano, T. Hara, H. Fujita, H. Hoshi, T. Iwama, and Y. Kinosada. Cad scheme for differential diagnosis of lacunar infarcts and normal virchow-robin spaces on brain MR images. In *IFMBE Proceedings of the World Congress on Medical Physics and Biomedical Engineering*, pages 126–128, Germany, Munich, 2009. doi: 10.1007/978-3-642-03904-1_36.

[47] Y. Uchiyama, A. Abe, C. Muramatsu, T. Hara, J. Shiraishi, and H. Fujita. Eigenspace template matching for detection of lacunar infarcts on MR images. *Journal of Digital Imaging*, 28(1):116–122, 2015. doi: 10.1007/s10278-014-9711-2.

[48] W. M. van der Flier, E.C.W. van Straaten, F. Barkhof, A. Verdelho, S. Madureira, L. Pantoni, D. Inzitari, T. Erkinjuntti, M. Crisby, G. Waldemar, R. Schmidt, F. Fazekas, and P. Scheltens. Small vessel disease and general cognitive function in nondisabled elderly. *Stroke*, 36(10):2116–2120, 2005. doi: 10.1161/01.STR.0000179092.59909.42.

[49] K.M.H. van Wijnen, F. Dubost, P. Yilmaz, M.A. Ikram, W.J. Niessen, H. Adams, M.W. Vernooij, and M. de Bruijne. Automated lesion detection by regressing intensity-based distance with a neural network. In *Proceedings of the 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–242, Shenzhen, China, 2019. doi: 10.1007/978-3-030-32251-9_26.

[50] U. Wang, J.A. Catindig, S. Hilal, H.W. Soon, E. Ting, T.Y. Wong, N. Venketasubramanian, C. Chen, and A. Qiu. Multi-stage segmentation of white matter hyperintensity, cortical and lacunar infarcts. *NeuroImage*, 60(4):2379–2388, 2012. doi: 10.1016/j.neuroimage.2012.02.034.

[51] J.M. Wardlaw, P.A.G. Sandercock, M.S. Dennis, and J. Starr. Is breakdown of the blood-brain barrier responsible for lacunar stroke, leukoaraiosis, and dementia? *Stroke*, 34(3):806–812, 2003. doi: 10.1161/01.STR.0000058480.77236.B3.

[52] J.M. Wardlaw, F. Doubal, P. Armitage, F. Chappell, T. Carpenter, S. Muñoz-Maniega, A. Farall, C. Sudlow, M.S. Dennis, and B. Dhillon. Lacunar stroke is associated with diffuse blood-brain barrier dysfunction. *Annals of Neurology*, 65(2):194–202, 2009. doi: 10.1002/ana.21549.

[53] J.M. Wardlaw, C. Smith, and M. Dichgans. Mechanisms underlying sporadic cerebral small vessel disease: insights from neuroimaging. *Lancet Neurology*, 12(5):483–497, 2013. doi: 10.1016/S1474-4422(13)70060-7.

[54] J.M. Wardlaw, E.E. Smith, G.J. Biessels, C. Cordonnier, F. Fazekas, R. Frayne, R.I. Lindley, J.T. O'Brien, F. Barkhof, O.R. Benavente, S.E. Black, C. Brayne, M. Breteler, H. Chabriat, C. DeCarli, F.E. de Leeuw, F. Doubal, M. Duering, N.C. Fox, S. Greenberg, V. Hachinski, I. Kilimann, V. Mok, R. van Oostenbrugge, L. Pantoni, O. Speck, B.C.M. Stephan, S. Teipel, A. Viswanathan, D. Werring, C. Chen, C. Smith, M. van Buchem, B. Norrving, P.B. Gorelick, and M. Dichgans. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *The Lancet Neurology*, 12(8):822–838, 2013. doi: 10.1016/S1474-4422(13)70124-8.

[55] J.M. Wardlaw, C. Smith, and M. Dichgans. Stroke subtype, vascular risk factors, and total mri brain small-vessel disease burden. *Neurology*, 83(14):1228–1234, 2014. doi: 110.1212/WNL. 0000000000000837.

[56] J.M. Wardlaw, S.J. Makin, M.C. Valdés Hernández, P.A. Armitage, A.K. Heye, F.M. Chappell, S. Muñoz-Maniega, E. Sakka, K. Shuler, M.S. Dennis, and M.J. Thrippleton. Blood-brain barrier failure as a core mechanism in cerebral small vessel disease and dementia: evidence form a cohort study. *Alzheimer's Dementia*, 13(6):634–643, 2017. doi: 10.1016/j.jalz.2016.09.006.

[57] B. Xu, Y. Chai, C. Galarza, C. Vu, B. Tamrazi, B. Gaonkar, L. Macyszyn, T. Coates, N. Lepore, and J. Wood. Orchestral fully convolutional networks for small lesion segmentation in brain MRI. In *Proceedings of the 15th IEEE International Symposium on Biomedical Imaging*, pages 889–892, Washington, United States of America, 2018. doi: 10.1109/ISBI.2018.8363714.

[58] R. Yokoyama, X. Zhang, Y. Uchiyama, H. Fujita, T. Hara, X. Zhou, M. Kanematsu, T. Asano, H. Kondo, S. Goshima, H. Hoshi, and T. Iwama. Development of an automated method for the detection of chronic lacunar infarct region in brain MR images. *IEICE Transactions on Information and Systems*, E90-D(6):943–954, 2007. doi: 10.1093/ietisy/e90-d.6.943.

[59] Zeiler. Adadelta: an adaptive learning rate method, 2012. URL `arXiv:1212.5701`.

[60] C.E. Zhang, S.M. Wong, H.J. van de Haar, J. Staals, J.F. Jansen, C.R. Jeukens, P.A. Hofman, R.J. van Oostenbrugge, and W.H. Backes. Blood-brain barrier leakage is more widespread in patients with cerebral small vessel disease. *Neurology*, 88(5):426–432, 2017. doi: 10.1212/WNL.0000000000003556.

[61] P. Zhang, W. Liu, H. Lu, and C. Shen. Salient object detection with lossless feature reflection and weighted structural loss. *IEEE Transactions on image processing*, 28(6):3048–3060, 2019. doi: 10.1109/TIP.2019.2893535.