UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Roxanne Silva Julia

Histological image segmentation of the bone vascular network through deep learning

Uberlândia, Brasil

2021

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Roxanne Silva Julia

Histological image segmentation of the bone vascular network through deep learning

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção do título de Bacharel em Ciência da Computação.

Supervisor: Dr. Bruno Augusto Nassif Travençolo

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Ciência da Computação

Uberlândia, Brasil

2021

Acknowledgements

Firstly, I would like to thank Doctor Bruno Augusto Nassif Travençolo for guiding me through this machine learning and image processing journey. It was all new territory for me and, not only he provided me with all the necessary equipment, but also did his best to support and offer me wisdom and knowledge whenever I needed it.

Secondly, I would like to thank the PHD student Dalí Freire Dias dos Santos for always, so promptly, answering my e-mails, which usually contained mountains of doubts. Your advice was much appreciated and really helped me a lot.

I would also like to thank Dr. Paula Dechichi, from the Federal University of Uberlândia, for providing the images that made this research possible.

Finally, I would like to thank my mum, my dad and my brother for being exactly who they are, which made them the most supportive, patient and loving people I could wish to have around throughout this adventure.

Abstract

Convolutional neural networks (CNN) are tools which have been intensively explored lately in the image processing field, showing very positive and promising results in classification and segmentation tasks. This project aims to study its application in the segmentation process of histological images of the bone vascular network. A PyTorch implementation of a U-Net was adapted to suit the purpose of this work. Training was performed with a provided set of bone vascular tissues images of rats and a corresponding manually marked set, where the canals, which are the objects of interest of the segmentation, were very roughly identified, presenting several flaws and wrong markings. This lack of precision in the ground truths had a huge negative impact on the model results, which ended up only achieving a Dice Coefficient of 0.20 due to overfitting.

Keywords: CNN, Segmentation, Histological images, Bone Vascular Network, U-Net

Resumo

Redes neurais convolucionais (CNN) são ferramentas que vêm sendo intensamente exploradas recentemente na área de processamento de imagem e demonstram resultados muito positivos e promissores em tarefas de classificação e segmentação. Este projeto têm o objetivo de estudar a sua aplicação ao processo de segmentação de imagens histológicas da rede vascular óssea. Uma implementação em PyTorch de uma U-Net foi adaptada para cumprir o propósito deste trabalho. O treinamento foi realizado com um conjunto fornecido de imagens da rede vascular óssea de ratos e um conjunto correspondente, manualmente marcado, em que os canais, os quais são as estruturas de interesse da segmentação, foram rusticamente identificados, apresentando assim diversas falhas e marcações erradas. Essa falta de precisão nas *ground truths* teve um enorme impacto negativo nos resultados do modelo, o qual foi capaz de atingir um *Dice Coefficient* de apenas 0.20 devido a *overfitting*.

Keywords: CNN, Segmentação, Imagens Histológicas, Rede Vascular Óssea, U-Net

List of Figures

List of abbreviations and acronyms

CNN	*Convolutional Neural Network*
DNN	*Deep Neural Network*
FCN	*Fully Convolutional Network*
ANN	*Artificial Neural Network*

Contents

1 Introduction

Primary data from various fields are acquired through images with multiple resolutions and dimensions (e.g. 2D, 3D) and there are several methods of capturing these images. For instance, in medical imagiology, images can be obtained through x-rays, one of the firsts and mostly used image capture techniques to assist medical diagnosis, manly when it comes to the bone structure (CHEN; SONG; CHEN XIAOYUAN ANDYANG, 2019). In biology, it is common to use histological images in the study of cells, tissues and organs. Basically, these images are obtained through the preparation of serial cuts which go through a colouring process that uses dye, since most tissues are colourless. All these processes are performed in slides, that are then taken to a microscope to capture the images. Histology is the study of tissues and how they compose organs. Tissues are classified into four groups: epithelial, conjunctive, muscular and nervous (JUNQUEIRA; CARNEIRO, 2013). Bone tissue, which is the main component of the skeleton, is part of the conjunctive tissue. It supports soft tissue and protects vital organs. This research will focus on the segmentation of the bone vascular network using deep learning techniques.

Deep learning is an artificial intelligence and machine learning field which is characterized by computational models composed of multiple processing layers. As the layers get deeper, the models can discoverer complex structures in big datasets using the back-propagation algorithm to designate how a machine should modify its internal parameters, which are used to compute the representation in each layer, based on the representation of the previous layer. These methods have improved the state-of-the-art in speech recognition, visual object recognition, object detection and many others domains. (LECUN; BENGIO; HINTON, 2015).

The complexity of the extracted image data, as well as the data volume obtained with the evolution of new and automatized microscopy technology, has transformed computers into one of the most viable resources for information extraction. Computational techniques have arisen with the goal of assisting humans in tasks that were previously done manually, which makes the whole process quicker and more efficient (SCHINDELIN et al., 2012).

Computerized methods for microscopical image analysis can significantly enhance the efficiency and objectivity of diagnosis and have gained a lot of focus in contemporary literature. Machine learning techniques in particular, like Convolutional Neural Networks (CNN), have been successfully applied in biological and medical researches (XING et al., 2018).

Seeing that convolutional neural networks have outperformed the state-of-the-

art in many visual recognition tasks (KRIZHEVSKY; SUTSKEVER; HINTON, 2012a; LONG; SHELHAMER; DARRELL, 2015a), new investigations about the application of CNN in digital images processing problems are very important. In this paper this has been done through histological images segmentation of the bone vascular network of rats femurs.

1.1 Objectives and Challenges of this research

The author's main research goal is to develop a computational methodology to segment serial cuts from histological images of the bone vascular network using deep learning. This main goal was divided into the following specifics goals:

- Analysing deep learning techniques to segment histological images of bone tissue;
- Proposing a convolutional neural network model for segmenting canals in histological images from the bone vascular network of rats femurs;
- Observing the behavior of a CNN when trained with poorly marked ground truths and how the results are affected by the wrong markings.

1.2 Justification

Studying the bone vascular network is very important because it can help preventing and treating possible bone tissue disorders. The histological analysis can detect both characteristics inherent in the individual who is being analysed and anomalies located in the tissue which were caused by therapy, medicament injections or incisions. These anomalies can cause vascularization reduction, bone metabolism deficiency and even cell loss or death (GONDIM et al., 2021).

Most histological images analysis of the bone vascular network is performed manually by a specialist. In this task, the specialist has to acquire the images, isolate the regions of interest, measure those regions and analyse the results. This is usually a long and subjective undertaking that can also be expensive. In addition, it can be a very tiring process for the specialist, turning the analysis susceptible to mistakes (GONDIM et al., 2021).

Recently, encouraged by promising results, deep learning based techniques are becoming increasingly popular in segmentation tasks (CUI et al., 2018). Figure 1 exhibits the amount of published papers that propose deep learning applications in histological images considering the network architecture and most common tasks.

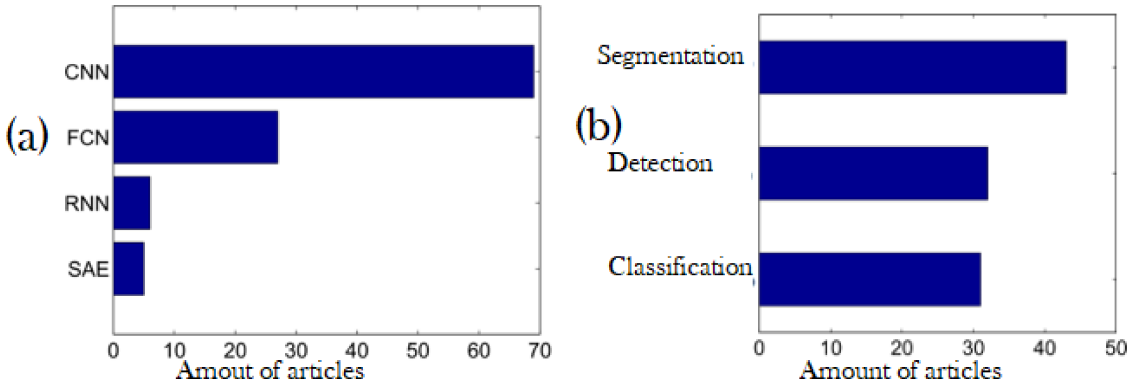


Figure 1 – Applications of deep learning in microscopical images analysis considering a) Network architecture. b) Executed tasks. CNN, FCN, RNN e SAE represent Convolutional Neural Network, Fully Convolutional Network, Recurrent Neural Network and Stacked AutoEncoder, respectively.
Source: adapted from Xing et al. (2018).

In this light, the adaptation of a convolutional neural network model, capable of segmenting canals in histologial images of the bone vascular network, will be an important contribution to researchers of the medical field.

2 Background and Related Work

This chapter presents basic biological and computational concepts. In the biology topics, histology and bone tissue and its cellular formation are discussed. The computational concepts discussed are related to digital images, image segmentation, artificial neural networks and convolutional neural networks. In the last section, some related works about computational methods used for cellular structure detection in images are presented.

2.1 Histology

Histology is a biology field that studies not only about the microscopic structure of all components in an organism, but also the functions which formulate these components. In order to analyse the structure of a body, in a histological perspective, it is necessary, in addition to using microscopes, to have knowledge on these instruments components and on the most used practices of specimens preparation for observation (ABRAHAMSOHN, 2016). Histology is a field which studies body tissues microscopically, as well as how these tissues are organized to build organs. Tissues are groupings of cells and extracellular matrices that, operating in an integrated way, conduct specific functions. In animals, tissues are defined in four fundamental groups. Each group contains very unique organization and functions and they can be defined as: epithelial tissue, conjunctive tissue, muscular tissue and nervous tissue. (JUNQUEIRA; CARNEIRO, 2013).

Since cells and the other matrices structures are very small, the histological analysis procedure depends on microscopes. Hence, the samples to be studied by histological techniques need to be previously prepared. The most used procedure for studying tissues with microscopes consists in the preparation of histological cuts (JUNQUEIRA; CARNEIRO, 2013).

A growing tendency in the histological tissue analysis field is the whole slide tissue scanning after a staining process. In this way, pathologists can make diagnoses based solely in a digital image. Furthermore, storing digital images is a more desirable alternative than storing glass slides, since slides require a lot more space and can be damaged, lost or even disappear over time (McCann et al., 2015). Consequently, once the digital images is obtained, it is possible to apply image processing techniques to improve the qualities of the images and also to use diagnostic algorithms, which can help/complement specialists evaluations (Gurcan et al., 2009).

2.2 Bone tissue and its cellular formation

The main component of the skeleton, known as bone tissue, which is a conjunctive tissue, supports soft tissues and protects vital organs. It is formed by cells and calcified extracellular material, known as bone matrix. These tissue cells belong to two different groups: osteoclasts and osteoblasts. Osteoblastic cells are constituted by osteoblasts and osteocytes. As for the osteoblastic cells, they are composed by osteoclasts (JUNQUEIRA; CARNEIRO, 2013).

Osteocytes are flattened cells that are located in the extracellular matrix, where they occupy spaces which are called gaps. Osteoblasts are located in the bone surface in an arrangement that resembles a simple epithelium and which synthesize the organic part of the bone matrix. The osteoclasts, in comparison to other cells, are big, multinucleated, reabsorb bone tissue and take part in bone remodeling processes (JUNQUEIRA; CARNEIRO, 2013).

2.3 Digital images

An image can be defined as a bi-dimensional function $f(x, y)$, where x and y are spatial coordinates and the amplitude of f in any point (x, y) is called color intensity or image color in that point. When the amplitude values of f and of any point (x, y) are all finite (discretized), this image is called a digital image (BHABATOSH et al., 2011). In other words, a digital image can be interpreted as a matrix in which, the row and column index identifies an element of the image and the value of the element corresponds to the intensity or color at that point. Each element of this matrix is called a pixel.

In colored digital images, each pixel is described through a set of properties like hue, saturation and brightness. In general, the color of each pixel is represented by a point in space defined by a color model (BHABATOSH et al., 2011). A color model, also known as color space or color system, is a set of 3D coordinates where each color can be represented by a point in this three-dimensional space (BHABATOSH et al., 2011).

2.4 Segmentation

Image segmentation is an important step applied to many tasks in the image, video and computer vision field. This process consists of dividing a certain image into various regions, accordingly to some criteria, in order to identify objects of interest. Image segmentation can also be used to identify regions of interest (foreground) and differentiate them from the other regions in the back of the image (background) (SHAPIRO; GEORGE, 2002).

In a segmentation process, image pixels are grouped based on similarity criteria, such as color or texture. The identified regions must be uniform and homogeneous accordingly to the criteria. Therefore, two pixels from distinct regions will have significant differences based on the chosen criterion (SHAPIRO; GEORGE, 2002).

There are many techniques used for image segmentation tasks, such as clustering algorithms (CAI; CHEN; ZHANG, 2007), edge detection algorithms (GONZALES; WOODS, 2002), watershed transform (BEUCHER; MEYER, 1993), limiarization (OTSU, 1979) and deep learning (KRIZHEVSKY; SUTSKEVER; HINTON, 2012b).

2.5 Artificial neural networks

According to Aggarwal (2018), artificial neural networks are popular machine learning techniques which simulate the learning mechanism of biological organisms. The human nervous system contains cells, which are called neurons. The neurons are connected to each other by axons and dendrites and the connection regions between these structures are called synapses. The strength of these synaptic connections frequently changes in response to external stimuli. These changes are how living organisms learn.

This biological mechanism is simulated in artificial neural networks, which contain computational units known as neurons. These computation units are connected to each other through weights, which have the same role as the strength of the synaptic connections in biological organisms. Each neuron input possesses a weight that affects the function that is calculated in that unit. This architecture is illustrated in Figure 2. An artificial neural network computes the inputs function through the propagation of the calculated values from the input neurons to the output neurons and uses the weights as intermediary parameters. Learning occurs upon the update of the weights that connect the neurons. Just like external stimuli are necessary for the learning process in biological individuals, they are also required in artificial neural networks and are provided through training data, which contains input-output examples of the function to be learned.

2.6 Convolutional neural networks

Convolutional neural networks (CNN) have historically been the most successful among all types of artificial neural networks. They are usually used in image recognition tasks, object detection and even text processing (GOODFELLOW; BENGIO; COURVILLE, 2016).

CNNs are feedforward networks in which information flow takes place in a single direction: from their inputs to their outputs. Just as artificial neural networks (ANN) are biologically inspired, so are CNNs. The visual cortex in the brain, which consists of

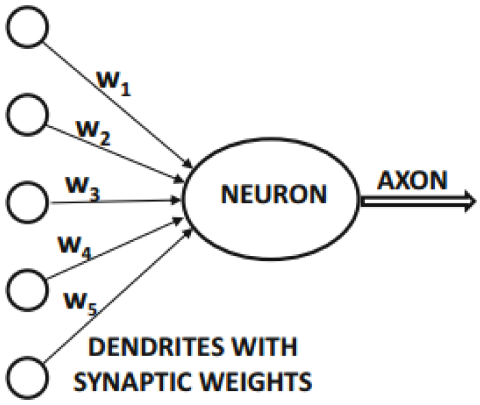


Figure 2 – Representation of an artificial neuron from an Artificial Neural Network.
Source: adapted from Aggarwal (2018)

alternating layers of simple and complex cells (HUBEL; WIESEL, 1962), motivates their architecture. CNN architectures, in general, consist of convolutional and pooling layers, which are grouped into modules. Either one or more fully connected layers, as in a standard feedforward neural network, follow these modules. Modules are frequently stacked on top of each other to form a deep model. Figure 3 illustrates typical CNN architecture for a vehicle image classification task: an image is passed as input to the network and this is followed by several stages of convolution and pooling. Subsequently, representations from these operations feed one or more fully connected layers. Finally, the last fully connected layer outputs the class label. The typical CNN architecture can be roughly divided in: convolutional, pooling and fully connected layers (RAWAT; WANG, 2017).

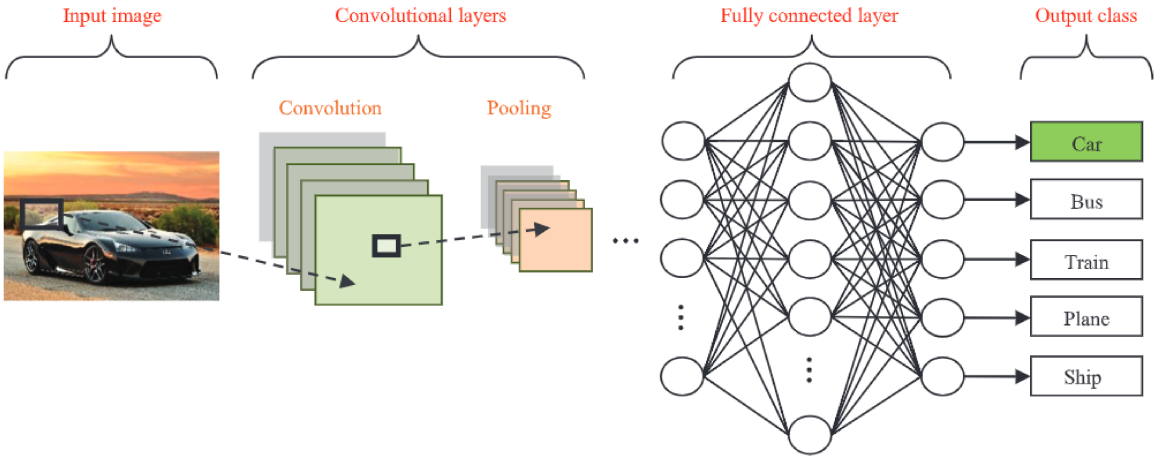


Figure 3 – CNN image classification pipeline.
Source: adapted from Rawat e Wang (2017).

The convolutional layers act as feature extractors, which learn the feature representations of their input images. The neurons in the convolutional layers are arranged into feature maps. Each neuron in a feature map has a receptive field, which is connected

to a neighborhood of neurons, located in the previous layer, through a set of trainable weights. Inputs are convolved with the learned weights in order to compute a new feature map, and the convolved results are sent through a nonlinear activation function (LECUN; BENGIO; HINTON, 2015). To summarize, the convolution process consists of the application of multiple kernels to an image with the aim of highlighting relevant features.

The purpose of the pooling layers is to reduce the spatial resolution of the feature maps and thus achieve spatial invariance to input distortions and translations (LECUN et al., 1990; LECUN et al., 1998; LECUN; BENGIO; HINTON, 2015; RANZATO et al., 2007; RAWAT; WANG, 2017). That means that the pooling layers shrink the images dimension in a way that the important information loss is minimal. The pooling process basically reduces the amount of parameters in the network.

Multiple convolutional and pooling layers are usually stacked on top of each other to extract more abstract feature representations in the process of moving through the network. The fully connected layers, which follow these convolutional and pooling layers, interpret these feature representations and perform the function of high-level reasoning (HINTON et al., 2012; SIMONYAN; ZISSERMAN, 2014; ZEILER; FERGUS, 2014). In other words, the fully connected layer performs the classification, which generates the networks prediction, also known as the network output. For classification problems, it is standard to use the softmax operator (RAWAT; WANG, 2017).

CNNs are usually used in classification tasks, where an image output is a single class label. Nevertheless, in many visual tasks, especially when dealing with biomedical images, a class label is supposed to be assigned to each pixel, since localization is important in the desired output. These visual tasks are known as segmentation tasks (RONNEBERGER; P.FISCHER; BROX, 2015).

2.7 Related work

The first studies approaching cellular structures detection and analysis in histological images worked with clustering algorithms such as K-means (FILIPCZUK; KOWAL; OBUCHOWICZ, 2011) and threshold techniques (FORD et al., 2018). Data about nuclear size were used as a prognosis factor for the first time in 1982 by Diamond et al. (1982). The authors developed a computerized method which, using histological H&E (haematoxylin and eosin stain) images, evaluate the shape of cellular nuclei through a measurement that was proposed by them, known as nuclear roundness.

Since then, a lot of progress has been achieved in histological images segmentation. Doyle et al. (2008) presented a model capable of distinguishing malignant lesions from non-malignant lesions (cancer vs. non-cancer) through studies of breast cancer cases. The model presented by Doyle et al. carries out the processing of histological images to extract

architectural features (space organization and histological structures distribution), as well as texture features (gray levels). After the extraction, the features were processed by a classifier, which made the distinction between cancer and non-cancer. The authors also presented a paper addressing which architectural and texture features were more relevant, not only for the cancer or non-cancer classification, but also for determining the stage of the disease.

CNNs have recently and frequently been applied to many image classification and segmantic segmentation tasks (Lecun et al., 1998), (CIREŞEAN et al., 2012), (KRIZHEVSKY; SUTSKEVER; HINTON, 2012a), (LONG; SHELHAMER; DARRELL, 2015b), (RONNEBERGER; P.FISCHER; BROX, 2015), (SIMONYAN; ZISSERMAN, 2015), (SZEGEDY et al., 2015), (MILLETARI; NAVAB; AHMADI, 2016), (KAMNITSAS et al., 2017), (HAVAEI et al., 2017). Medical images may come from a variety of imaging technologies, like ultrasound, X-ray, computed tomography (CT) and magnetic resonance imaging (MRI). They are often used for identifying different anatomical structures in the human body, such as blood vessels, bones, vertebrae and major organs. Since medical images portrait not only healthy human anatomy, but also different types of unhealthy structures (tumors, injuries, lesions, etc.), segmentation usually has two goals: delineating different anatomical structures (such as bone canals) and detecting unhealthy tissue (such as brain lesions) (KAYALIBAY; JENSEN; SMAGT, 2017).

Cireşean et al. (2012) used a Deep Neural Network (DNN), which is a structure inspired by CNN, to segment neuronal membranes through pixel classification. The network calculates the probability of a pixel belonging to the membrane, an operation that is applied to each pixel of the input image.

In 2015, the concept of a Fully Convolutional Network (FCN) was proposed for segmentation purposes by Long, Shelhamer e Darrell (2015b). FCNs are adapted CNNs, where a deconvolution process is done after the regular convolution. That allows the network to receive inputs of arbitrary size and, then, to produce segmentated outputs of the same size.

Basing themselves on the concept of FCNs, Ronneberger, P.Fischer e Brox (2015) presented a CNN model known as U-Net that is considered fast and precise for biomedical image segmentation and requires few training images. According to the authors, the proposed model is capable of segmentating 512×512 pixels images in less than one second. The model has won the Cell Tracking Challenge 2015 cellular image segmentation contest (RONNEBERGER; P.FISCHER; BROX, 2015). A further explanation of this model will be presented in Chapter 3.

The use of convolutional neural networks for medical images segmentation is becoming more common every day, however, a study about CNN applied to histological images of the bone vascular network hasn't been suggested yet, which is the proposition

of this work.

There are many approaches proposed in the literature to evaluate histological bone images. In Liu et al. (1999), they were able to measure how old the samples were through micro features analysis of bone tissue. In addition, the authors had to segment the images, which was performed using a neighborhood adaptive smoothing process that had the aim of lowering the amount of noise in the images. Then, the k-means algorithm was applied to the output of the previous step and, finally, each generated group was labeled as an image object. In Oliveira et al. (2006), morphological features of chickens were evaluated with the aim of analysing changes in the bone structure. To reach that goal, the authors segmented bone canals in images using manual limiarization, where pixels with a lower intensity than a predefined value (determined by the authors) were considered background, whereas pixel with equal or larger intensity values were considered the objects of interest. Rabelo, Beletti e Dechichi (2010) 's research aimed to study the alterations caused in rats bones by radiotherapy, through the analysis of histological images. First, the authors had to segment the images using a manual limiarization method and, then, area features were extracted, as well as their standard deviation, perimeter and lacunarity.

Gondim et al. (2021) proposed an automated bone canals segmentation method that can be divided in three steps. The first step is preprocessing, in which color deconvolution techniques were used (with the aim of highlighting the components that were tainted by hematoxylin and eosin), color space transformation (with the aim of highlighting the white parts of the image) and also contrast adjustment (to correct possible inadequate alteration in the tonalities that might have happens when the images were captured). The second step consists in segmentation, which was done through the k-means algorithm, to identify the areas of interest and image binarization. As for the third step, it is the post-processing step, where the results were refined through the withdrawal of objects that were not of interest.

The authors proposals also aims for a method for automated bone canals segmentation, however, as opposed to the paper presented by Gondim et al. (2021), the segmentation will be obtained using a convolutional neural network.

3 Methodology

The research methodology was divided in Materials, Methods and Dataset.

3.1 Materials

The images in this project are from bone vascular tissue of rats and were provided by Dr. Paula Dechichi, from the Federal University of Uberlândia. The intermediary region of the femur, known as the diaphysis, is the only bone region used in this study because it has a great concentration of vascular networks, which are responsible for blood irrigation. In this region there are two types of vascular canals: Haver's canals, which follow the long bone axis, and Volkman's canals, which connect the Haver's canals transversely (PAIVA, 2019). Figure 4 illustrates the detailing in the bone vascular network of a rat.

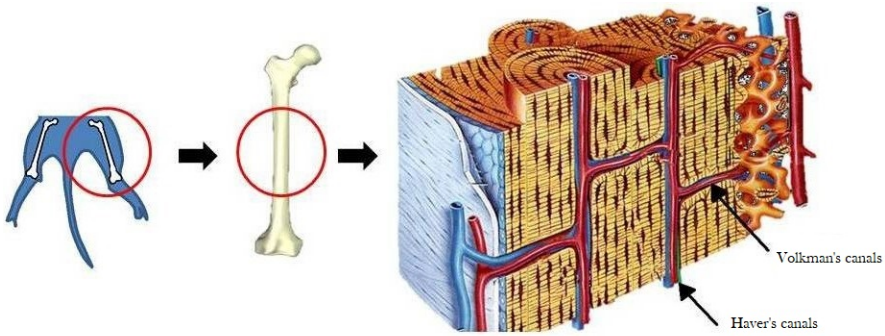


Figure 4 – Schematic illustration of the bone vascular network. (a) Localization of a rat femur. (b) Localization of the diaphysis in a femur. (c) Volkman's and Haver's canals. Volkman's canals are the vertical ones and Haver's are the transversal ones that connect Volkman's.
Source: Abreu (2016).

The images were acquired from a single rat that was monitored in a bioterium in the Federal University of Uberlândia. It weighted between 200g and 300g, was kept in a cage and was fed water and ration in an appropriate manner. A single dose of 30 Gy was applied to the rat's left femur. No medication was applied to the right femur so it could be used as a control mechanism. Sixty days after the application, the rat was scarified.

The histological images were captured with a high resolution ScanScope At TurbooR Scaner. Figure 5 shows an original histological image of a rat femur. In Figure 6, the canals, which are the structures of interest, are identified. Structures known as osteocytes are also identified in Figure 6, however they won't be segmented.

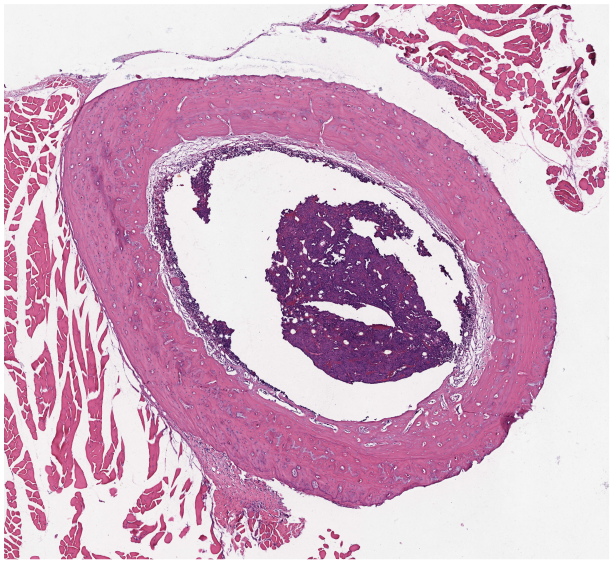


Figure 5 – Histological original image of the rat left femur

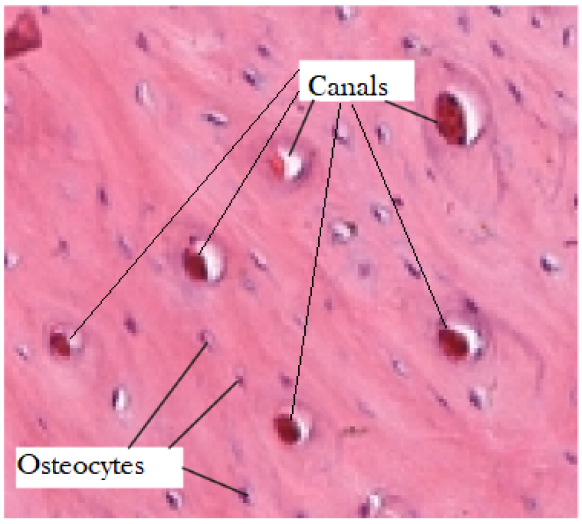


Figure 6 – Canals and osteocytes.

3.2 Methods

This research's Methods were divided in U-Net and Evaluation and loss metrics. U-Net was the chosen convolutional neural network architecture for this project. It is a CNN created for biomedical image segmentation, which has been showing very promising results and seemed suited for this work.

3.2.1 U-Net

The network architecture is illustrated in Figure 7. It consists of a contracting path (left side) and an expansive path (right side). The contracting path follows the typical architecture of a convolutional network. It consists of the repeated application of two 3×3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU)

and a 2×2 max pooling operation with stride 2 for downsampling. At each downsampling step the number of feature channels is doubled. Every step in the expansive path consists of an upsampling of the feature map followed by a 2×2 convolution ("up-convolution") that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3×3 convolutions, each followed by a ReLU. The cropping is necessary due to the loss of border pixels in every convolution. At the final layer a 1×1 convolution is used to map each 64-component feature vector to the desired number of classes. In total the network has 23 convolutional layers (RONNEBERGER; FISCHER; BROX, 2015).

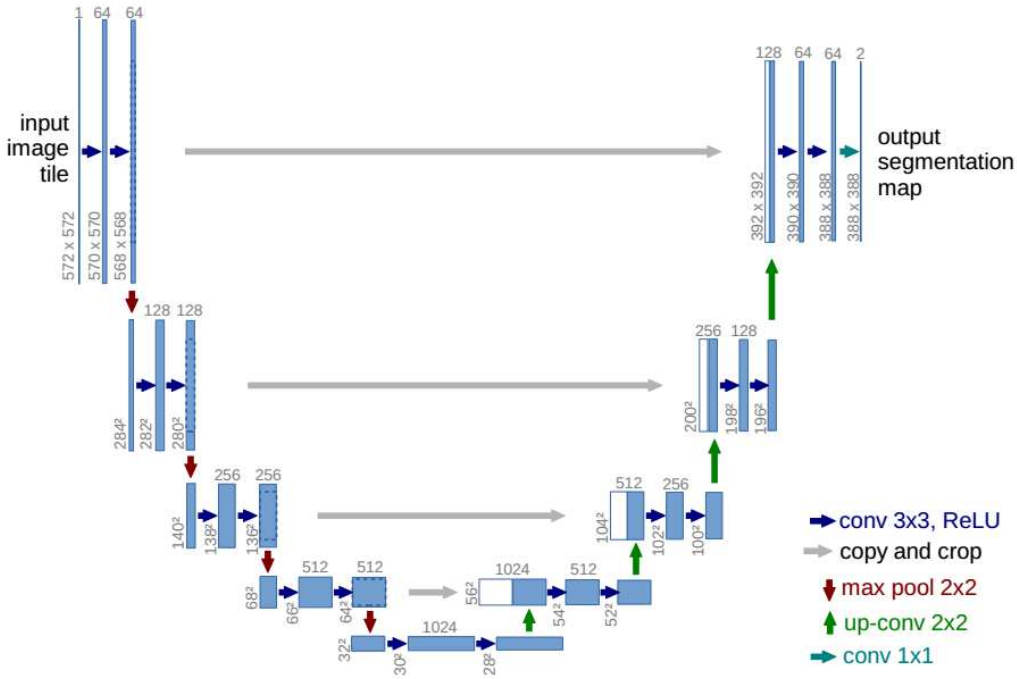


Figure 7 – U-Net architecture. Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.
Source: adapted from Ronneberger, Fischer e Brox (2015).

A Pytorch implementation of a U-Net, originally proposed for FLAIR abnormality segmentation in brain MRI, provided in Buda, Saha e Mazurowski (2019), was altered to fit the task proposed in this work. The main alterations were performed in the Dataset class, where the images are preprocessed and passed as input to the network.

3.2.2 Evaluation and loss metrics

The Evaluation and loss metrics used in this work were Dice coefficient, Precision and Recall.

3.2.2.1 Dice coefficient

In image segmentation tasks, severe data imbalance, a phenomenon where negative examples outnumber positive ones, is very common and can prove to be a problem in the training process. Data imbalance results in training-test discrepancy: if the labels aren't balanced, the learning process tends to converge to a point that strongly biases towards the class with the majority label (LI et al., 2020).

In this light, Dice Coefficient, or Sørensen–Dice coefficient, or F1 score, or Dice Loss has proven to be a very good loss function and evaluation method for semantic segmentation problems with highly imbalanced data, which is the case for the dataset used in this work.

As defined in (Wikipedia contributors, 2021), given two sets, X and Y, it is defined as:

$DSC = \frac{2|X \bigcap Y|}{|X|+|Y|}$ where |X| and |Y| are the cardinalities of the two sets (i.e. the number of elements in each set). The Sørensen index equals twice the number of elements common to both sets divided by the sum of the number of elements in each set.

When applied to Boolean data, using the definition of true positive (TP), false positive (FP), and false negative (FN), it can be written as:

$$DSC = \frac{2TP}{2TP+FP+FN}$$

In the segmentation task proposed by this work, in which the objects of interest are the canals, each pixel is classified by the network as canal or not canal. A TP consists of a pixel that was correctly classified as canal, whereas a FP consists of a pixel that was incorrectly classified as canal, in other words, a pixel that isn't part of a canal in the ground truth, but was classified as so by the network. Following the same logic, a true negative (TN) is a pixel that was correctly classified as not canal by the network, whereas a FN is a pixel that was incorrectly classified as not canal.

3.2.2.2 Precision

Precision is a good metric to evaluate how many true positives were predicted, out of all the positive predictions. It can be defined as:

$$Precision = \frac{TP}{TP+FP}$$

3.2.2.3 Recall

Recall is a metric that evaluates how many positives the model predicted correctly, out of all the actual positives. It can be defined as:

$$Recall = \frac{TP}{TP+FN}$$

3.3 Dataset

The Dataset is composed of seventy images. A set of corresponding manually marked images, where the canals were identified, were also provided by the Biology faculty. Figure 5 shows an example of an original image, whereas Figure 8 shows the corresponding manually marked image.

Figure 8 – Corresponding marked image of the rat left femur.

The original images had to be manually aligned with their corresponding marked ones, since the original provided pairs of image/corresponding marked image weren't originally aligned. Subsequently, a threshold was applied to the marked images and the background was painted black. Figure 9 shows the corresponding mask of Figure 8.

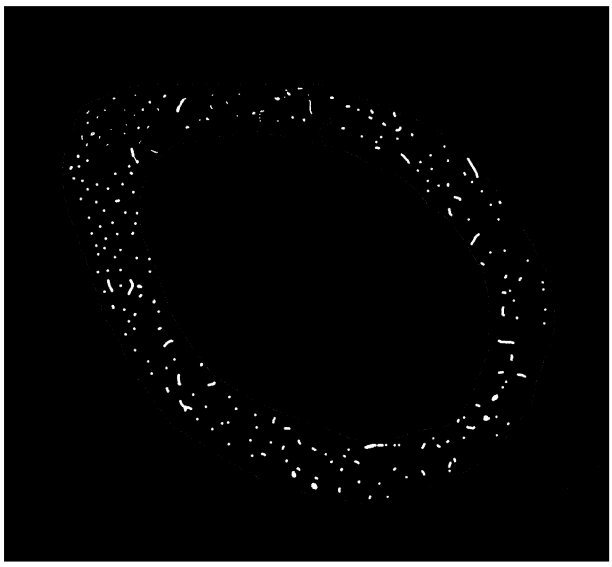
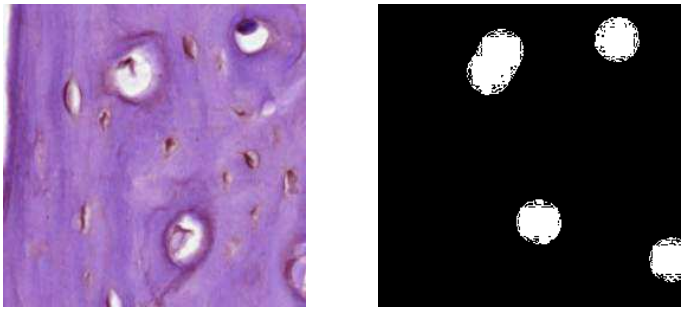


Figure 9 – Corresponding mask of the rat left femur.



(a) Original (b) Mask

Figure 10 – Examples of image patches. (a) Patch from the original image; (b) Corresponding mask patch

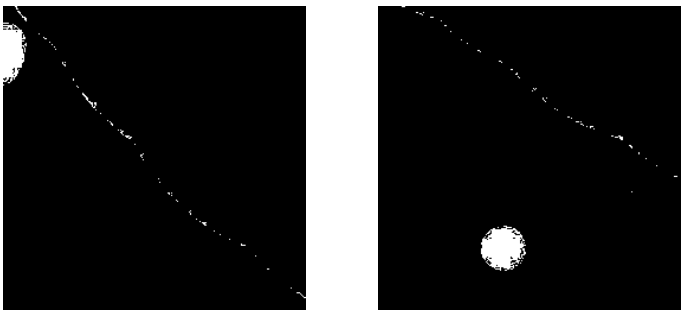


Figure 11 – Examples of noise in the dataset caused by the JPEG format.

Since the images are very large (around 7800×7800 pixels), they, and their corresponding masks, were sliced into 10266 256×256 pixels image patches. The patch size was chosen based on the standard input size of the U-Net implementation that was used. Finally, after performing a pixel-wise normalization, the patches were passed as input to the network. In Figure 10, an input example is presented.

It is important to highlight that the original images were provided in JPEG, which turned out to be a problem in the threshold step, generating noise in the dataset. This happens because, when compressing the image, the JPEG format changes the values of objects borders in the image. These modifications are not perceivable to the human eye, but they can affect image processing routines. Examples of this issue can be observed in Figure 11.

Moreover, the provided manually marked images contained various flaws and wrong markings. That occurred because the images were originally marked for the identification of the canals (and not their exact borders), for a better understanding of the influence of radiation in the architecture of the diaphysis. Hence, precision was neglected. Besides, the markings were made in the original images, which was also a key factor for inaccuracy in the ground truths, since the images are very large and the canals are small, making it very difficult for a human to mark them accurately without a close-up. Examples of such flaws and wrong marking can be observed in Figure 12.

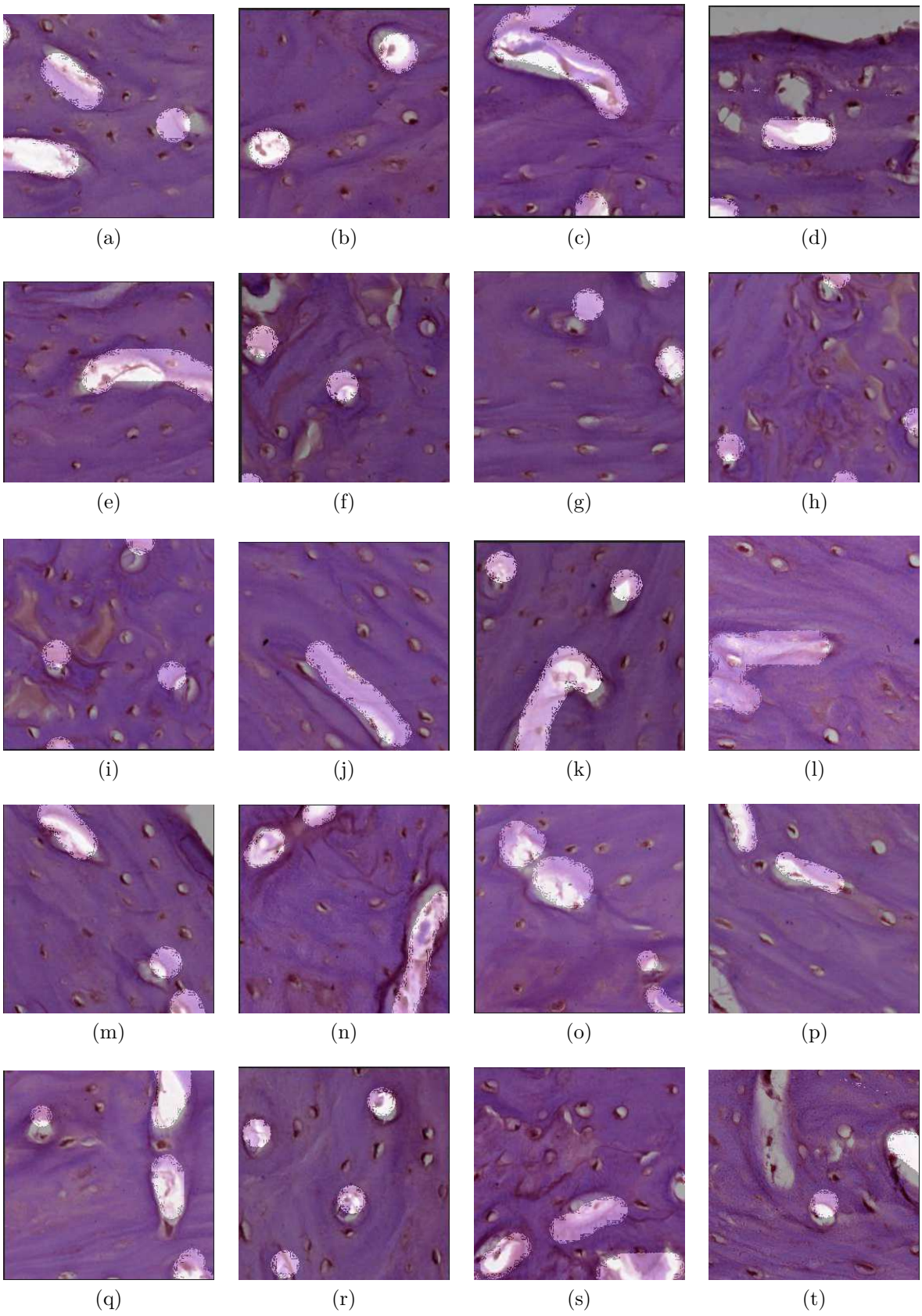


Figure 12 – Examples of originals patches overlapped with their corresponding masks that illustrate some of the dataset's flaws and wrong marking.

4 Results and Discussion

As mentioned before, training was done with a dataset of 10266 patches (70 images) and, out of those, two randomly chosen images were used for validation. In addition, three images (419 patches) that weren't part of the dataset were used for testing the trained model. It is also important to highlight that the model contained data augmentation (rotation and horizontal flip).

The best accuracy achieved by the model were a mean dice coefficient of 0.18 for the validation images and of 0.20 for the test images. These results were reached with 100 epochs, a batch size of 16 and a learning rate of 0.001. The poor performance of the model can be explained by overfitting, which is when the model is unable to generalize well. It learns the features of the training set, but is unable to classify accurately any data that slightly deviates from the exact data used during training. This usually happens when the dataset doesn't correctly represents the proposed problem.

During training, it was observed that, although both the training and validations dice losses kept diminishing throughout the epochs, the accuracy stopped improving very quickly, which indicates overfitting. Tests with different learning rates were also performed, but the results did not improve. As mentioned in Chapter 3, the provided marked images of the dataset contained very serious flaws and wrong marking, which was probably the main reason for such results.

The mean precision and recall for the test outputs were also calculated and the results were 0.16 and 0.32, respectively. Figure 13 shows examples of output images from the test set. The green outlines correspond to ground truth (masks markings) and red to model predictions. It is possible to observe that the model tends to predict several false positives. Not only it is segmenting larger areas than the canals, which are the areas of interest, but it is also getting confused by the osteocytes, which are objects very similar to the canals. The model also predicted a lot of false negatives, which can be observed in Figure 14.

It is also important to discuss the possible reasons for such discrepancy between the results of this work and the one proposed by Gondim et al. (2021), which were very good, considering that the same images and corresponding marked images were used in both works. Gondim et al. (2021) proposed an automated bone canals segmentation method that used traditional image segmentation techniques, which did not involve deep learning. In their approach, the marked images were only used after the segmentation process, for validation, which means that they do not have any impact in the actual segmentation. In this work, on the other hand, the flawed marked images were used for training, hence,

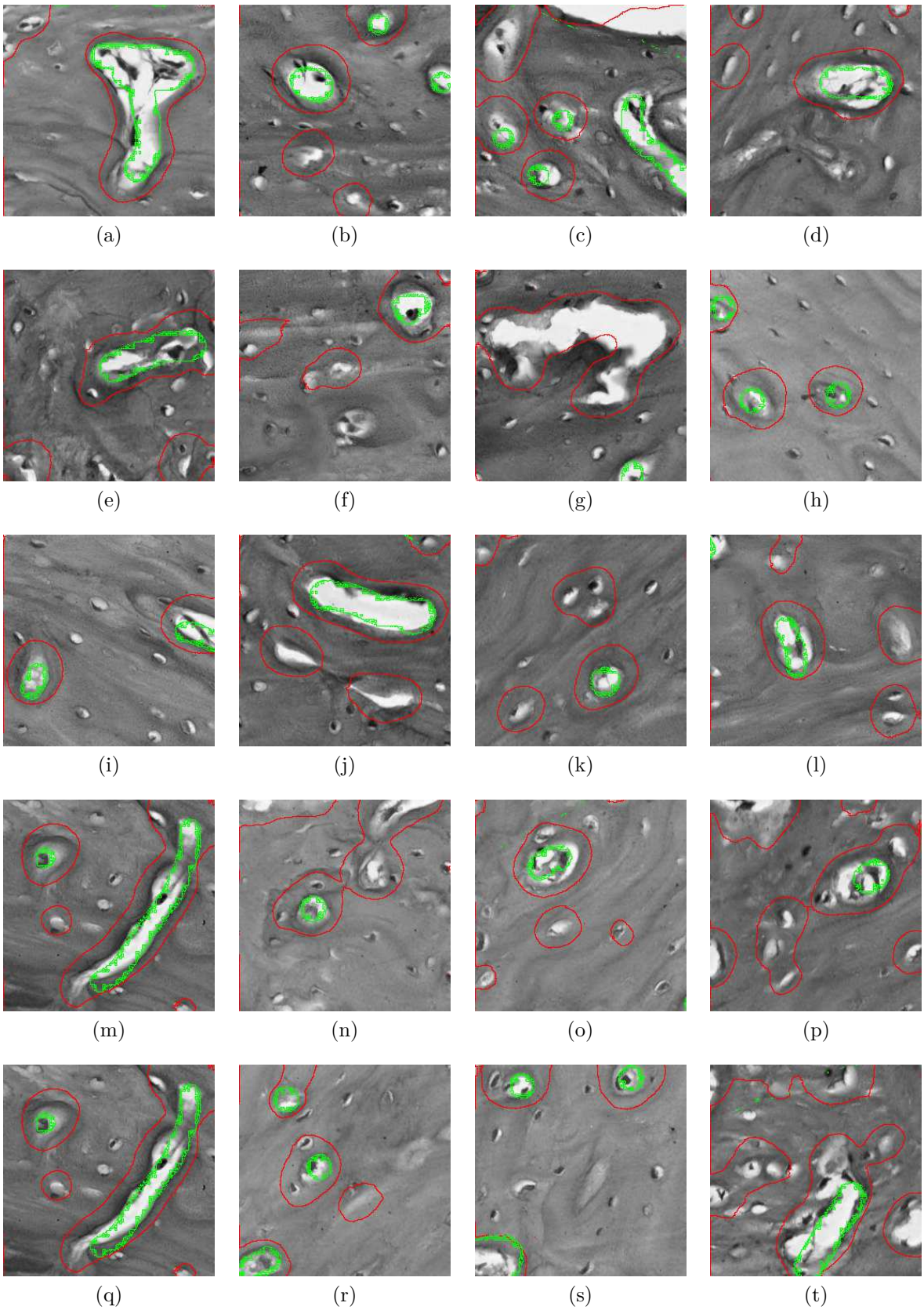
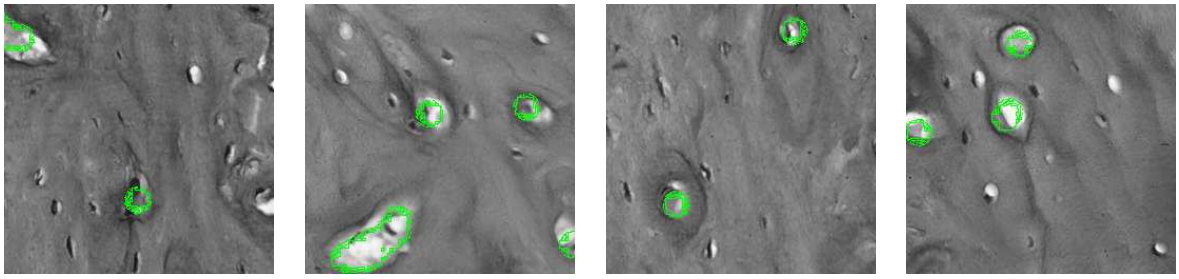


Figure 13 – Examples of model outputs for test images. The green outlines correspond to ground truth (masks markings) and red to model predictions.



(a) (b) (c) (d)

Figure 14 – False negative examples of model outputs for test images. The green outlines correspond to ground truth (masks markings). These regions were not predicted by the model (i.e, they are false negatives).

they had a direct impact in the segmented outputs, which was probably why the results of this work were much more affected by the poorly marked images than the ones presented by Gondim et al. (2021).

5 Conclusion

This work had as a main goal the development of a computational methodology to segment serial cuts from histological images of the bone vascular network using deep learning. Throughout the process of trying to achieve such a goal, a lot of research about machine learning theory and frameworks, CNNs architectures, as well as deep learning and image processing techniques had to be performed, which was a very enriching and rewarding experience for the authors, since this isn't a topic much explored yet by the Computer Science undergraduate course.

Reaching the achieved results required not only learning and getting familiarized with Pytorch, the machine learning framework, but also modifying an implementation of a U-Net, originally proposed for FLAIR abnormality segmentation in brain MRI, to fit the proposed task. Furthermore, the authors had to work with a very flawed and weak dataset, which was an extremely challenging process.

The main conclusion reached by this research is that the dataset's flaws resulted in the overfitting of the model. Not only it is filled with wrong and poorly made markings, a consequence of the method used for marking and of the initial purpose of the marked images, but also holds many residue border pixels, since the originals images were provided in JPEG, which created noise in the masks. Moreover, it was observed that the osteocytes proved to be a problem, since they are very similar to the canals and the model is frequently classifying them as objects of interest.

Although the results ended up being quite poor, a lot has been learned by the authors about machine learning and image processing, which are largely researched fields at the moment. The knowledge acquired in this work will create a very strong basis for future projects and we believe that, after the adjustments in the training and test dataset, the methodology will perform well.

5.1 Future works

In future works, the authors intend to correct the flaws found in the dataset by manually marking the masks again and removing the noise present in them, with the aim of obtaining better results. In addition, they plan to find a solution to diminish the impact the osteocytes are having on the model.

Bibliography

ABRAHAMSOHN, P. *Histologia*. 1a edição. ed. [S.l.]: Guanabara Koogan, 2016. ISBN 978-8527729819. Cited on page 11.

ABREU, A. F. *Estudo e desenvolvimento de algoritmos de esqueletização com aplicação em redes vasculares ósseas*. Dissertação (Mestrado) — Universidade Federal de Uberlândia, 2016. Cited on page 18.

AGGARWAL, C. C. *Neural Networks and Deep Learning*. [S.l.]: Springer-Verlag GmbH, 2018. ISBN 9783319944630. Cited 2 times on pages 13 and 14.

BEUCHER, S.; MEYER, F. The morphological approach to segmentation: the watershed transformation. *Mathematical morphology in image processing*, v. 34, 1993. Cited on page 13.

BHABATOSH, C. et al. *Digital image processing and analysis*. [S.l.]: PHI Learning Pvt. Ltd., 2011. Cited on page 12.

BUDA, M.; SAHA, A.; MAZUROWSKI, M. A. Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. *Computers in Biology and Medicine*, Elsevier, v. 109, 2019. Cited on page 20.

CAI, W.; CHEN, S.; ZHANG, D. Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. *Pattern recognition*, Elsevier, v. 40, n. 3, 2007. Cited on page 13.

CHEN, X.; SONG, J.; CHEN XIAOYUAN ANDYANG, H. X-ray-activated nanosystems for theranostic applications. *Chemical Society Reviews*, Royal Society of Chemistry, 2019. Cited on page 8.

CIREŞEAN, D. C.; GIUSTI, A.; GAMBARDELLA, L. M.; SCHMIDHUBER, J. Deep neural networks segment neuronal membranes in electron microscopy images. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2*. Red Hook, NY, USA: Curran Associates Inc., 2012. (NIPS'12), p. 2843–2851. Cited on page 16.

CUI, Y.; ZHANG, G.; LIU, Z.; XIONG, Z.; HU, J. *A Deep Learning Algorithm for One-step Contour Aware Nuclei Segmentation of Histopathological Images*. 2018. Cited on page 9.

DIAMOND, D. A.; BERRY, S. J.; JEWETT, H. J.; EGGLESTON, J. C.; COFFEY, D. S. A new method to assess metastatic potential of human prostate cancer: relative nuclear roundness. *The Journal of urology*, Wolters Kluwer Philadelphia, PA, v. 128, n. 4, 1982. Cited on page 15.

DOYLE, S.; AGNER, S.; MADABHUSHI, A.; FELDMAN, M.; TOMASZEWSKI, J. Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. In: IEEE. *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. [S.l.], 2008. Cited on page 15.

FILIPCZUK, P.; KOWAL, M.; OBUCHOWICZ, A. Automatic breast cancer diagnosis based on k-means clustering and adaptive thresholding hybrid segmentation. In: *Image processing and communications challenges 3*. [S.l.]: Springer, 2011. Cited on page 15.

FORD, N.; SHUBBER, Z.; JARVIS, J. N.; CHILLER, T.; GREENE, G.; MIGONE, C.; VITORIA, M.; DOHERTY, M.; MEINTJES, G. Cd4 cell count threshold for cryptococcal antigen screening of hiv-infected individuals: a systematic review and meta-analysis. *Clinical Infectious Diseases*, Oxford University Press US, v. 66, n. suppl_2, p. S152–S159, 2018. Cited on page 15.

GONDIM, P. H. C. C.; LIMIRIO, P. H. J. O.; ROCHA, F. S.; BATISTA, J. D.; DECHICHI, P.; TRAVENÇOLO, B. A. N.; BACKES, A. R. Automatic segmentation of bone canals in histological images. *Journal of Digital Imaging*, Springer, p. 1–13, 2021. Cited 4 times on pages 9, 17, 25, and 27.

GONZALES, R. C.; WOODS, R. E. *Digital image processing*. [S.l.]: Prentice hall New Jersey, 2002. Cited on page 13.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>. Cited on page 13.

Gurcan, M. N.; Boucheron, L. E.; Can, A.; Madabhushi, A.; Rajpoot, N. M.; Yener, B. Histopathological image analysis: A review. *IEEE Reviews in Biomedical Engineering*, v. 2, 2009. Cited on page 11.

HAVAEI, M.; DAVY, A.; WARDE-FARLEY, D.; BIARD, A.; COURVILLE, A.; BENGIO, Y.; PAL, C.; JODOIN, P.-M.; LAROCHELLE, H. Brain tumor segmentation with deep neural networks. *Medical Image Analysis*, Elsevier BV, v. 35, p. 18–31, Jan 2017. ISSN 1361-8415. Disponível em: <http://dx.doi.org/10.1016/j.media.2016.05.004>. Cited on page 16.

HINTON, G. E.; SRIVASTAVA, N.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. R. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012. Cited on page 15.

HUBEL, D. H.; WIESEL, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, Wiley Online Library, v. 160, n. 1, p. 106–154, 1962. Cited on page 14.

JUNQUEIRA, L. C.; CARNEIRO, J. *Histologia Básica*. 1. ed. [S.l.]: Editora Guanabara Koogan Ltda, 2013. ISBN 9788527723114. Cited 3 times on pages 8, 11, and 12.

KAMNITSAS, K.; LEDIG, C.; NEWCOMBE, V. F.; SIMPSON, J. P.; KANE, A. D.; MENON, D. K.; RUECKERT, D.; GLOCKER, B. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical Image Analysis*, Elsevier BV, v. 36, p. 61–78, Feb 2017. ISSN 1361-8415. Disponível em: <http://dx.doi.org/10.1016/j.media.2016.10.004>. Cited on page 16.

KAYALIBAY, B.; JENSEN, G.; SMAGT, P. van der. *CNN-based Segmentation of Medical Imaging Data*. 2017. Cited on page 16.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F.; BURGES, C. J. C.; BOTTOU, L.; WEINBERGER, K. Q. (Ed.). *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., 2012. p. 1097–1105. Disponível em: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>. Cited 2 times on pages 9 and 16.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. Cited on page 13.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, n. 7553, p. 436–444, 2015. ISSN 1476-4687. Disponível em: <https://doi.org/10.1038/nature14539>. Cited 2 times on pages 8 and 15.

LECUN, Y.; BOSER, B. E.; DENKER, J. S.; HENDERSON, D.; HOWARD, R. E.; HUBBARD, W. E.; JACKEL, L. D. Handwritten digit recognition with a back-propagation network. In: *Advances in neural information processing systems*. [S.l.: s.n.], 1990. p. 396–404. Cited on page 15.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Ieee, v. 86, n. 11, p. 2278–2324, 1998. Cited on page 15.

Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998. Cited on page 16.

LI, X.; SUN, X.; MENG, Y.; LIANG, J.; WU, F.; LI, J. *Dice Loss for Data-imbalanced NLP Tasks*. 2020. Cited on page 21.

LIU, Z.-Q.; LIEW, H. L.; CLEMENT, J. G.; THOMAS, C. D. L. Bone image segmentation. *IEEE Transactions on biomedical engineering*, IEEE, v. 46, n. 5, 1999. Cited on page 17.

LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015. Cited on page 9.

LONG, J.; SHELHAMER, E.; DARRELL, T. *Fully Convolutional Networks for Semantic Segmentation*. 2015. Cited on page 16.

McCann, M. T.; Ozolek, J. A.; Castro, C. A.; Parvin, B.; Kovacevic, J. Automated histology analysis: Opportunities for signal processing. *IEEE Signal Processing Magazine*, v. 32, n. 1, p. 78–87, 2015. Cited on page 11.

MILLETARI, F.; NAVAB, N.; AHMADI, S.-A. *V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation*. 2016. Cited on page 16.

OLIVEIRA, R. C. d.; COSTA, L. d. F.; FERNANDES, E. A.; MATIOLI, S. R.; BELETTI, M. E. et al. Bone histomorphometry of broilers submitted to different phosphorus sources in growing and finisher rations. *Pesquisa Agropecuária Brasileira*, SciELO Brasil, v. 41, n. 10, 2006. Cited on page 17.

OTSU, N. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, IEEE, v. 9, n. 1, 1979. Cited on page 13.

PAIVA, B. V. d. *Implicações da obesidade induzida pela dieta hiperlipídica na composição molecular e rugosidade superficial do osso cortical*. Dissertação (Mestrado) — Universidade Federal de Uberlândia, 2019. Cited on page 18.

RABELO, G. D.; BELETTI, M. E.; DECHICHI, P. Histological analysis of the alterations on cortical bone channels network after radiotherapy: a rabbit study. *Microscopy research and technique*, Wiley Online Library, v. 73, n. 11, 2010. Cited on page 17.

RANZATO, M.; HUANG, F. J.; BOUREAU, Y.-L.; LECUN, Y. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In: IEEE. *2007 IEEE conference on computer vision and pattern recognition*. [S.l.], 2007. p. 1–8. Cited on page 15.

RAWAT, W.; WANG, Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation*, v. 29, n. 9, p. 2352–2449, 09 2017. ISSN 0899-7667. Disponível em: <https://doi.org/10.1162/neco_a_00990>. Cited 2 times on pages 14 and 15.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: NAVAB, N.; HORNEGGER, J.; WELLS, W. M.; FRANGI, A. F. (Ed.). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing, 2015. p. 234–241. ISBN 978-3-319-24574-4. Cited on page 20.

RONNEBERGER, O.; P.FISCHER; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2015. (LNCS, v. 9351), p. 234–241. (available on arXiv:1505.04597 [cs.CV]). Disponível em: <http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a>. Cited 2 times on pages 15 and 16.

SCHINDELIN, J.; ARGANDA-CARRERAS, I.; FRISE, E.; KAYNIG, V.; LONGAIR, M.; PIETZSCH, T.; PREIBISCH, S.; RUEDEN, C.; SAALFELD, S.; SCHMID, B. et al. Fiji: an open-source platform for biological-image analysis. *Nature methods*, Nature Publishing Group, v. 9, n. 7, p. 676, 2012. Cited on page 8.

SHAPIRO, L.; GEORGE, C. Stockman g: computer vision. In: *Prentice Hall*. [S.l.: s.n.], 2002. Cited 2 times on pages 12 and 13.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Cited on page 15.

SIMONYAN, K.; ZISSERMAN, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. Cited on page 16.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCKE, V.; RABINOVICH, A. Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015. Cited on page 16.

Wikipedia contributors. *Sørensen–Dice coefficient — Wikipedia, The Free Encyclopedia*. 2021. [Online; accessed 28-April-2021]. Disponível em: <https://en.wikipedia.org/w/index.php?title=S%C3%B8rensen%E2%80%93Dice_coefficient&oldid=1007400543>. Cited on page 21.

XING, F.; XIE, Y.; SU, H.; LIU, F.; YANG, L. Deep learning in microscopy image analysis: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, v. 29, n. 10, p. 4550–4568, Oct 2018. ISSN 2162-2388. Cited 2 times on pages 8 and 10.

ZEILER, M. D.; FERGUS, R. Visualizing and understanding convolutional networks. In: SPRINGER. *European conference on computer vision*. [S.l.], 2014. p. 818–833. Cited on page 15.