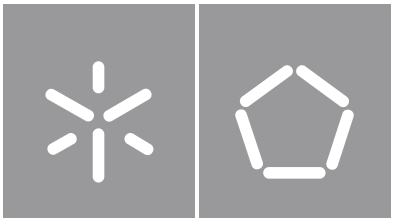
Universidade do Minho
Escola de Engenharia

José Carlos Rodrigues Carneiro

Artificial Intelligence for Tattoo Removal Treatment Planning

janeiro de 2023

Artificial Intelligence for Tattoo Removal Treatment Planning
José Carneiro
UMinho | 2023



Universidade do Minho
Escola de Engenharia

José Carlos Rodrigues Carneiro

Artificial Intelligence for Tattoo Removal Treatment Planning

Dissertação de Mestrado
Mestrado em Engenharia Eletrónica
Industrial e Computadores
Controlo, Automação e Robótica

Trabalho efetuado sob a orientação de
Professor Doutor Jaime Fonseca
Professor Doutor João Vilaça

janeiro de 2023

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Acknowledgements

This master's dissertation, which marks the completion of these five years at the University of Minho that have a significant meaning for me, is the product of many hours of effort, study, research, and development over the past few months. I matured academically and socially. I take with me knowledge but also memories. That's why I say goodbye with great emotion to this journey that is very significant in my life. But these paths are not taken alone, and for that, I have the following gratitude to express:

I want to start by thanking Professor Doctor Jaime Fonseca for his guidance, care, and availability. Thank you for always having the greatest advice to provide and always showing trust in me.

I would also like to thank Professor Doctor João Vilaça for welcoming me to the 2Ai (Applied Artificial Intelligence) laboratory and for all the support. To Bruno Oliveira, a special thanks for guiding me this year, and for his ideas and suggestions that made me progress throughout this project. Also, I want to thank Augusto Righetti for reviewing an article and for sharing his knowledge with me. I also met a number of colleagues at 2Ai, especially Nuno Costa, Bruno Duarte, Margarida Ferreira, Raúl Ferrete, Pedro Lobo, and Sérgio Pereira, with whom I shared ideas, thoughts, and pleasant moments. All of you were crucial to the completion of this dissertation.

To my friends, Rui Costa, Alexandre Mano, Rui Esteves, João Relvas, Filipa Lage, Pedro Duarte, Mariana Duarte, Luíz Alexandre, and Duarte Alves, these five years would not have been the same without you. Thank you for everything.

Sara Pereira, who was more than a partner not only on this journey but on many others as well, deserves special mention. We shared many moments and I will be forever grateful. Thank you so much for always being there!

Last but not least, I want to show my gratitude to those who have supported me for as long as I can remember. They are the people I love the most and I strive to make them proud. Many thanks to my mother Carla Carneiro, my father João Carneiro, and my sister Carla João Carneiro.

To all those mentioned and to many others, I will always be grateful.

The SmartHealth project was funded by the project “NORTE-01-0145-FEDER-000045”, supported by Northern Portugal Regional Operational Programme (Norte2020), under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Resumo

Lesões e marcas na pele são cada vez mais frequentes. Uma das formas mais antigas de marcar o corpo humano de forma permanente é realizando uma tatuagem. Esta consiste numa pigmentação subcutânea feita através da inserção de tinta.

Cada vez mais pessoas em todas as faixas etárias, sexos e graus de escolaridade têm tatuagens e um número surpreendentemente alto de pessoas está arrependido da sua decisão. Estima-se que 40% dos adultos entre 26 e 40 anos tenham pelo menos uma tatuagem. Além disso, 19% das pessoas tatuadas consideram a sua remoção.

Existem técnicas rudimentares para remover tatuagens mas o tratamento atual e mais popular de remoção de tatuagens baseia-se em terapia laser, a qual com um feixe de laser quebra os pigmentos de tinta. Contudo este método de remoção depende da experiência do médico, isto é, pode não ser garantida uma remoção segura porque alguns parâmetros como a distância e o ângulo do laser com a pele, assim como o correto caminho para remover a tatuagem, podem não ser executados com precisão.

Portanto, é cada vez maior a necessidade de otimizar os processos clínicos. Uma das melhores maneiras de aprimorar sistemas médicos tais como cirurgias ou análise de imagens médicas, é utilizando soluções baseadas em Inteligência Artificial.

Os avanços da Inteligência Artificial e do Deep Learning têm ajudado inúmeros processos clínicos, apresentando excelentes performances na segmentação e registo de imagens, como também na análise de diversas imagens médicas e no tratamento de inúmeros problemas, incluindo lesões e marcas na pele.

Esta dissertação visa contribuir para uma solução para a remoção de tatuagens auxiliando no planeamento do tratamento da mesma através da construção de um dataset para a segmentação de tatuagens, através da criação de um método de inteligência artificial para uma segmentação robusta das mesmas e analisando a influência de imagens de próximas do infravermelho para uma melhor segmentação. Ao todo, para obter segmentação de tatuagens em tempo real para auxiliar no planeamento de trajetória de um sistema baseado em robótica. Tal sistema pode facilitar o tratamento de remoção de tatuagem por laser, tornando a terapia mais precisa e consistente.

keywords: *Inteligência artificial, deep learning, segmentação de imagem, lesões na pele, marcas na pele, tatuagens, dataset.*

Abstract

Lesions and marks on the skin are increasingly common. One of the oldest ways to permanently mark the human body is by getting a tattoo. This consists of a subcutaneous pigmentation made through the insertion of ink.

More and more people across all age groups, genders, and educational levels have tattoos and a surprisingly high number of people are regretting their decision. It is estimated that 40% of adults between 26 and 40 years old have at least one tattoo. Furthermore, 19% of tattooed people consider their removal.

There are rudimentary techniques to remove tattoos, but the current and most popular treatment for tattoo removal is based on laser therapy, which with a laser beam breaks down ink pigments. However, this removal method depends on the physician's experience, that is, safe removal cannot be guaranteed because some parameters such as the distance and angle of the laser to the skin, as well as the correct way to remove the tattoo, may not be performed with accuracy.

Therefore, the need to optimize clinical processes is increasing. One of the best ways to improve medical systems such as surgery or medical image analysis is by using Artificial Intelligence based solutions.

Advances in Artificial Intelligence and Deep Learning have helped numerous clinical processes, by presenting excellent performances in image segmentation and registration, as well as in the analysis of various medical images and the treatment of numerous problems, including lesions and skin tags.

This dissertation's major goal is to contribute to find a solution for tattoo removal by helping in its planning and treatment by creating a dataset for tattoo segmentation, developing an artificial intelligence method for robust tattoo segmentation, and analyzing the influence of Near-Infrared (NIR) for improved segmentation. Altogether to get real-time tattoo segmentation to aid the trajectory planning of a robotic-based system. Such a system may ease the laser tattoo removal treatment while making the therapy more precise and consistent.

keywords: *Artificial intelligence, deep learning, image segmentation, skin lesions, skin marks, tattoos, dataset.*

Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligence.
CNNs	Convolutional Neural Networks.
DL	Deep Learning.
GND	Ground.
HSV	Hue, Saturation, Value.
IR	Infrared.
JI	Jaccard Index.
NIR	Near-Infrared.
RGB	Red, Green, Blue.
TCA	TriChloroacetic Acid.
UV	Ultraviolet.
YCbCr	Luminance; Chroma: Blue; Chroma: Red.

Chapter 1: Introduction

Tattoos are becoming more and more popular. There are an increasing number of people who want tattoos but also wish to get them removed. It is estimated that almost 23% of the 30% of Americans who have tattoos have tattoo regret. According to the 2020 United States Census, the population was 329.5 million people. This means there were up to that year around 22,735,500 Americans with tattoo remorse [1]. Also, searches for "tattoo removal" on Google in the United States have climbed by 40% since 2020 [2].

A permanent tattoo is made by injecting a pigment just below the skin's dermal-epidermal junction with a needle or other similar tool, as can be seen in Figure 1.1 [3].

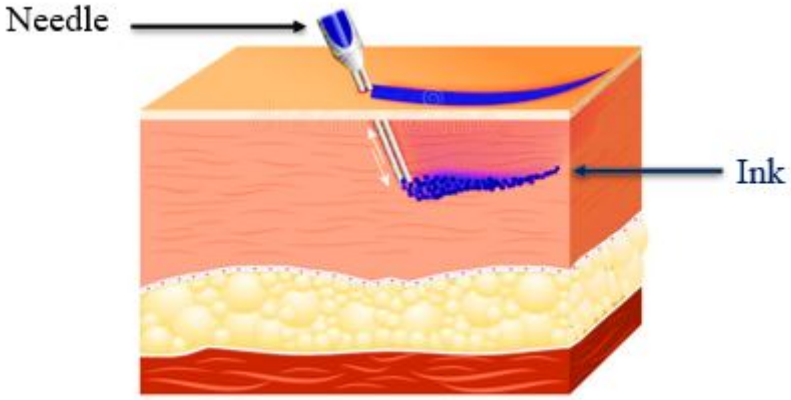


Figure 1.1: Tattooing process [4]

The immune system attacks the pigment, which then settles in macrophages and fibroblast cells, resulting in permanent skin pigmentation. Tattooing has been a feature of human civilization from ancient times, according to historical sources, with its popularity increasing and decreasing based on current cultural norms, [3].

The following anthropological constructs about tattooing as a manifestation of a basic human attitude are based on a review of prehistorical stories and historical literature. Body marking has been a part of man's repertoire since the dawn of time, and it has served a variety of purposes, [5]. Tattooing can be traced back to at least 8,000 years ago, according to evidence. Crude bone needles and bowls that stored

pigment have been discovered in caves and rock strata in France, Portugal, Romania, and Scandinavia. [6]. Tattoos are becoming more common as society's attitude toward this type of body art moves from stigma to acceptance. Around 33% of 18–25-year-olds and 40% of 26–40-year-olds in the United States have tattoos, [7]. Tattoos, unlike other body modifications like body piercing, are permanent, [3].

There are a variety of reasons why people get tattoos. Fashion, proving masculinity and toughness, or showing independence are among the most prevalent motivations for tattooing, but independently the motivation, some type of non-verbal expression appears to be important to tattooing.

The nature of communication varies and is influenced by the individual's inner world. The skin of a person interacts with the surroundings. It serves as a protective barrier, as well as a conduit for good and negative sensations to reach the neurological system. Skin plays a crucial part in people's personality, [8]. As a result, tattoos may symbolize a person's attempt at self-expression and originality, [3].

Tattoos have been linked to psychological characteristics of a person's personality, such as sexual immaturity, rebellion, and identity issues [9]. Low self-esteem has also been reported leading some to believe that tattoos are a form of emotional defense [3]. Certain studies have proposed that tattoos are a type of self-defence [10], denoting the bearer's hostility and strength. Tattoos relating to the individual's aggressive violent background and gang involvement are frequently seen in prison populations, and they may have been acquired to ward off unwanted attention or assaults in prison.

People sometimes get tattoos with the mistaken idea that they will make their bodies more appealing, which is relevant considering the rise in popularity of tattoos as a fashion statement in recent decades. In research on women's tattoo motivations [11], it was discovered that they got tattoos to improve their feminine image or as a form of public defiance, [3].

Tattoos have been acquired for religious, magical, social, and community causes over history [12]. Until now, several body marks have profound religious and superstitious significance in some regions of the world. Getting tattoos at a young age or impulsively is one of the primary drivers.

As tattoos continue to rise in popularity, so has the demand for an effective method of tattoo removal, [7].

1.1 Tattoo Industry

The tattoo industry has grown every year. All over the world, more and more people are getting tattoos for various reasons, such as expressing themselves, artistic issues, religious reasons, being addicted, or even feeling sexier, [13]. There are countless reasons people get tattoos.

The "ink business" is experiencing a boom. This sector is expected to grow even more in the next few years. In the US, the market size of tattoo artists has grown 9.2% per year, [14]. Figure 1.2 illustrates the market size of tattoo artists in the United States.

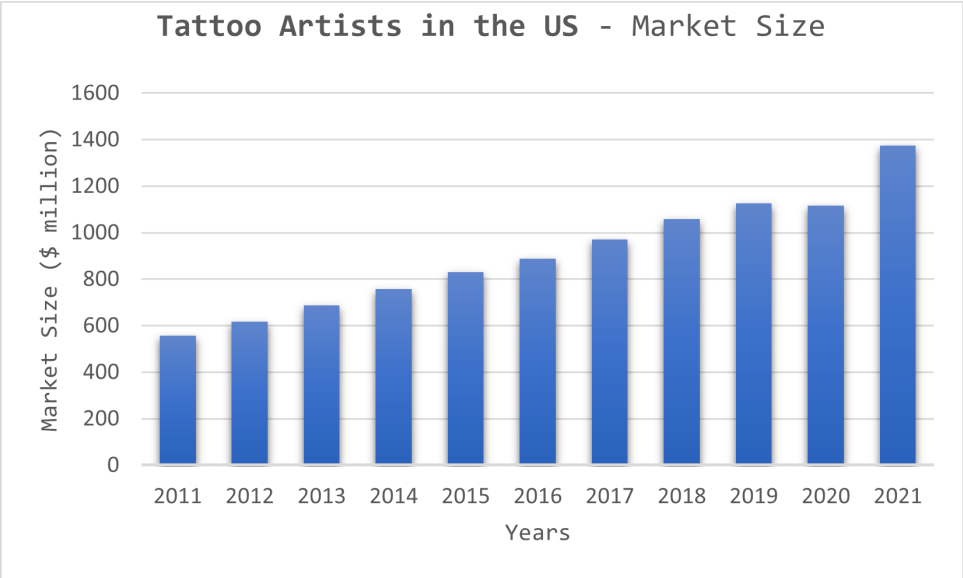


Figure 1.2: Market size of tattoo artists in the US [14]

The trend of getting tattooed has opened a massive opportunity for tattoo removal services. Many of the people who get tattoos later want to have them removed.

But other factors allow an increase in these removal services, such as advances in laser technology that permit different approaches in the area of aesthetics; the change in medical procedures that became less and less invasive, making post-surgery care easier [15].

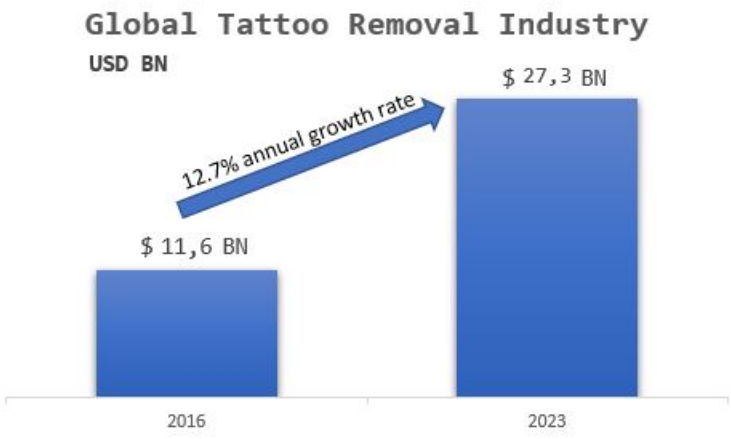


Figure 1.3: Global tattoo removal industry [16]

Figure 1.3 proves there is an enormous growth in tattoo removal procedures, and the importance of improving these services.

Tattoo removal may be desired for a variety of reasons, including remorse after receiving a tattoo, to improve the self-esteem, to get a more professional appearance, or to increase credibility [17]. Elimination procedures are costly and can result in scarification and incomplete removal.

1.2 Surgical Treatments

There are several methods to remove tattoos, namely salabrasion, dermabrasion, surgical excision, chemical removal, or laser therapy. This section provides a brief overview of the different types of removal technique.

Salabrasion

The salabrasion method consists of using an abrasive item, typically a stone covered in gauze, to massage salt and water into the tattoo for a period of about 30 minutes. The goal is to remove the top layers of skin by rubbing and peeling them off with the salty mixture [18].

After using the salabrasion combination to treat the tattoo, the area is treated with an antibiotic cream and covered with sterile gauze for a few days. This process is cyclic until the tattoo is removed [18] . In the Figure 1.4, it is possible to see the tattoo vanishing because of the repeated process of salabrasion.

Figure 1.4: Salabrasion treatment results [19]

The upper layer of the skin starts to die and is separated from the skin below, taking the ink pigments with it. However, this technique is outdated, extremely painful, and can cause several complications such as hyper and hypo-pigmentation, risk of infection, and permanent scarring, [18].

Dermabrasion

Another bygone method is dermabrasion, in which the specialist scrapes the top layers of the skin with a diamond-particle-filled burr or rough wire brush linked to a rotating tool, as can be seen in Figure 1.5.

Figure 1.5: Example of a rotating device to scrape the skin [20]

In the following days, the patient must take care of the irritated skin with gauzes and moisturizers [21].

As this method is cheaper and easier to do, similar to salabrasion, many people still use this technique even though it is outdated. Figure 1.6, it is illustrated a young man who successfully removed his tattoo.

Figure 1.6: Dermabrasion treatment results [21]

In the first image is the tattoo in its original state, then the tattoo removal during the procedure is presented, and finally the scar and final result after a few months of healing.

Nonetheless, the tattoo is not always removed successfully, and often the skin becomes lighter or darker, scars are more noticeable, and other side effects such as redness, bleeding, and infections can happen. Also, this treatment is occasionally combined with surgical excision.

Surgical Excision

Surgical excision is performed in extremely specific circumstances to remove minor tattoos caused by problems including ink that was introduced into the skin through an accident (crash with a car, fireworks), or in cases of allergy, tumors, and other adverse skin reactions with the tattoo.

A deep excision is performed, cutting the skin, fat and muscle around the intended area. By removing this piece that was cut, the tattoo is also removed. It is usually necessary to suture the wound and use a

bandage and an absorbent until it heals. After a few weeks, it is healed [22]. Figure 1.7, shows the before and after of this type of procedure.

Figure 1.7: Surgical excision treatment results [23]

This treatment is only recommended for small tattoos. Large tattoos can lead to different excision sessions.

Scarring is the most common side-effect of this procedure.

Chemical Removal

Another way to remove tattoos is through the use of acids. The three most common procedures are:

- Glycolic Acid Peel - The glycolic acid mixture, when applied at home or by a dermatologist, can inflame the skin, causing it to expel tattoo ink pigments. This procedure is painful, ineffective and results in a scar.
- TCA Peel - TriChloroacetic Acid (TCA) burns the top layers of the skin. The longer the acid stays on the skin the more layers are burned. However, it is inefficient and makes the skin lighter.
- Acid injections - This method is more invasive than the previous ones, as acid is inserted into the skin with a needle. The acid tries to bring the ink to the surface in the form of a crust, and when it comes off, the tattoo comes off with it [24].

Figure 1.8 illustrates this technique.

Figure 1.8: Chemical treatment [24]

However, this method is painful, and not recommended. Also, these treatments can lead to skin infection, hypo-pigmentation and scarring.

Laser Removal

Salabrasion, surgical excision, dermabrasion, and chemical destruction can all induce tissue injury [25], and recent advancements have led to the use of laser technologies, which have fewer side effects, [26]. Laser tattoo removal has surpassed all previous methods and is now the most effective. Some factors, such as advancements in laser technology and the use of various strategies and protocols, have made this method the most widely used worldwide.

The laser technology is generally used in the disciplines of medicine and aesthetics and it is based on the theory of photothermolysis, which states that by selecting the proper wavelength for a chromophore target in a time shorter than or equal to the thermal relaxation time, the target is destroyed. For tattoo removal it is necessary to choose the proper laser wavelength for the intended ink (target) to break down the tattoo ink pigments into smaller particles for lymphatic system cleaning [27], as it can be seen in Figure 1.9.

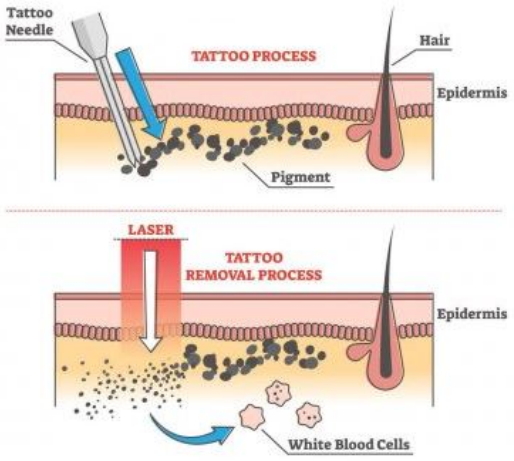


Figure 1.9: Tattoo removal by laser [28]

Different wavelengths correspond to different colors, so the laser wavelength must be selected based on the tattoo color. Table 1.1 displays some wavelengths according to tattoo color, [7].

Table 1.1: Laser wavelenghts and correspondent color

	Black	Blue	Green	Purple	Red	Yellow	Orange
Alexandrite 755 nm	Destroy	Destroy	Destroy				
Nd:YAG 1065 nm	Destroy	Destroy	Destroy				
Nd:YAG 532 nm				Destroy	Destroy	Destroy	Destroy
Pulse dye laser 585-595 nm					Destroy	Destroy	Destroy
Ruby 694 nm	Destroy	Destroy	Destroy				

Since tattoo pigments absorb light at specified wavelengths, the type of laser utilized must be able to emit enough energy within the pigment's wavelength range to be effective. Because black is the darkest hue of color people can get, black tattoos are typically the simplest to get rid of. The deeper color pigments absorb all laser wavelengths and draw the lightest energy from the laser, which helps break up the ink particles. Greater contrast between the tattoo's pigment and your skin's color will make laser tattoo removal simpler. Tattoo removal might be a little more difficult for colorful tattoos than for black ink. The more closely the laser matches the color of the tattoo, the more effective it will be removed [29]. Therefore, optimal laser selection is necessary.

In terms of removal protocol, the patient is normally anesthetized, each individual in the room must wear protection glasses, and if the tattoo is near the eyes, the patient must also wear intraocular protection.

The least amount of fluence necessary to provide the intended therapeutic outcome of instantaneous skin whitening upon laser impact should be used during laser therapy. Treatment sessions at different times should allow time for the last session to heal properly. Finally, the patient is entitled to post-treatment information. Also, it is necessary to use sunscreen and protective bandages after the procedure [30]. The Figure 1.10 shows the common laser tattoo removal.

Figure 1.10: Tattoo removal by laser [31]

In short, the tattoo removal procedure starts with the patient telling the doctor how the tattoo was acquired, so that he can analyze the best removal strategy. The skin must be numbed with a local anesthetic before the laser therapy. Every tattoo is unique. The tattoo's size, color, and ink type vary. Each individual also has a vastly different skin type. Therefore, it is important to study the best way to remove each tattoo. Then, the tattoo is subjected to a strong energy pulse that causes the ink to heat up and shatter. The ink pigments will collapse to a much more manageable size. This makes it possible for the white blood cells to remove the smaller ink particles so that the body can discard them. If the tattoo has many colors, it could be required to use lasers with different wavelengths. Most of the time, it is needed several treatments to entirely get rid of the tattoo, therefore it can take some time. As the removal session end, more and more ink particles will be dispersed. This implies that the tattoo will lose more ink as a result. After each removal session, it is necessary to take care of the wound. The basic post-treatment includes daily soap and water cleaning and gauze dressing replacement. The removal of tattoos has been revolutionized by laser technology since it is safer, more advanced, and the method with fewer side effects [32]. However, this removal is not always successful since the tattoo may remain visible. This might be due to the complexity of the tattoo or simply because the proper distance and angle between the laser and the skin were not achieved, or the optimal path to remove the tattoo was not chosen, because humans are prone to fatigue, lack of concentration, and even inexperience, which is why it is important to support the physician in developing the removal plan.

Figure 1.11 shows an example of an incomplete tattoo removal by a laser treatment.

Figure 1.11: Tattoo removal by laser [33]

1.3 Artificial Intelligence in Medicine

AI in medicine is the use of machine learning algorithms to explore medical data and reveal insights to assist improve health outcomes and patient experiences. AI is increasingly becoming a vital aspect of modern healthcare as a result of recent developments in the field of computer science and informatics. AI algorithms and other AI-powered apps are being utilized to aid medical practitioners in clinical settings and continuing research [34]. In the future years, AI will alter the traditional function of doctors. It is assisting in the resolution of many healthcare issues. As a result, Ai will be advantageous in diagnosing, treating diseases, detecting or identifying characteristics of a specific health problem, eliminating human errors, and digitally interacting with patients [35].

Various AI algorithms for segmenting medical images or situations containing objects of interest have been studied during the last few years [36], [37]. The implementations range from fundamental image processing approaches such as active contour, thresholding, and boundaries detection to contemporary DL algorithms, the latter of which has shown exponential growth in the fields of health informatics and medical imaging [38]. DL designs include convolutional neural networks (CNNs), which are comparable to primary visual networks [39]. The CNN architecture, in particular, can extract complicated information at the same level as humans, giving it a more efficient generalization capability than other machine-learning approaches.

DL in image segmentation may be used to identify lesions, discover anomalies, improve the visualization of specific diseases, aid in operations by accurately locating abnormalities, and assist with tracking in handling operations.

Consequently, image segmentation based on AI approaches will be able to deliver information to the physician, aid him, and provide him with benefits ranging from planning to occasions during the process, of various types of treatments. These benefits may also be advantageous to the tattoo removal problem. AI can improve the performance of the tattoo removal physician by ensuring the correct parameters in the removal, such as distance and perpendicularity with the skin, indicating the most efficient path, and eliminating the possibility of a flawed removal. This is only feasible if the tattoo has first been segmented/characterized.

1.4 Motivation

The tattoo industry is expanding globally as can be observed in the section 1.1. Therefore, there is a growing demand for tattoo removal services. It is estimated that 40% of adults between 26 and 40 years old have at least one tattoo. Furthermore, 19% of tattooed people consider their removal, [40]. Having an unwanted tattoo can lead to other problems, such as depression or low self-esteem. Therefore, removing undesired tattoos is crucial. The primary justifications for getting tattoos removed are [30]: dislike the way the tattoo looks; it no longer represents the same thing; the professional environment is negatively impacted by tattooing, making it more difficult to get employment. People just became bored of having tattoos on their bodies since they did not fit in with their current lifestyle.

Although more and more people intend to remove their tattoos and new advances in laser technology, the present method still depends exclusively on the physician's expertise. It is the physician who guarantees the angle and distance of the laser from the skin, and chooses the path to remove the tattoo. Even if a physician is skilled, performing several treatments back-to-back in one day can be exhausting, which affects concentration and precision. Lack of experience, focus, or precision during the removal procedure can result in skin complications like hypo-pigmentation, skin infections, and scarring. Therefore, it's essential to improve these removal methods. This improvement is made possible with the use of AI, which can segment tattoos and therefore define the region over which the laser must pass to remove them. This project seeks to demonstrate segmentation methods, research various neural networks that can segment tattoos, compare segmentation approaches, build databases that support neural networks, and investigate how infrared images affect segmentation. In this way, it is possible to start creating systems that in the future will assist doctors in the removal processes and thus improve the performance of the treatment. As a result, using systems that aid the physician in the tattoo removal process, it may be possible to have faster and more efficient sessions where the physician does not need to be extremely skilled or focused. This might reduce the number of sessions necessary and post-treatment complications, making this sort of technique more advantageous and attractive to anybody looking to remove tattoos.

1.5 Contribution

This project was developed within the scope of the project SmartHealth at 2Ai laboratory (2Ai, Applied Artificial Intelligence Laboratory, EST, IPCA, Barcelos, Portugal) and focuses on developing an intelligent solution that allows medical procedures to be faster and easier, by planning the treatment of skin lesions

by laser systems, specifically tattoo removal, with the use of artificial intelligence, to a later alliance with collaborative robotics.

The main objective can be divided into three distinct phases:

- **Phase 1** - Construction of a dataset for tattoo segmentation: For segmentation tasks, there are no tattoo datasets available yet with ground truth, therefore pre-processing will be done on the datasets that already exist to obtain skin-tattoo images and the corresponding ground truth with aim of construct a new dataset that allow the implementation of Deep Learning methods, in the following phase.

- **Phase 2** - Creation of an artificial intelligence method for a robust tattoo segmentation : This phase will involve the implementation of an algorithm that receives the dataset, processes it, and segments tattoos using DL. The purpose is to set up a method to study various models and evaluate how well they perform in tattoo segmentation tasks. In addition, a traditional tattoo segmentation method was developed, to compare with the method based on AI, and therefore understand its relevance. In the end of this phase, the evaluation of segmentation models is based on defined metrics.

- **Phase 3** - Study the influence of the Near-Infrared (NIR) dataset in tattoo segmentation: A dataset of near-infrared images will be created in this phase, and the effects of NIR images on segmentation techniques will be studied.

Figure 1.12 illustrates the sequence of the main topics of this project.

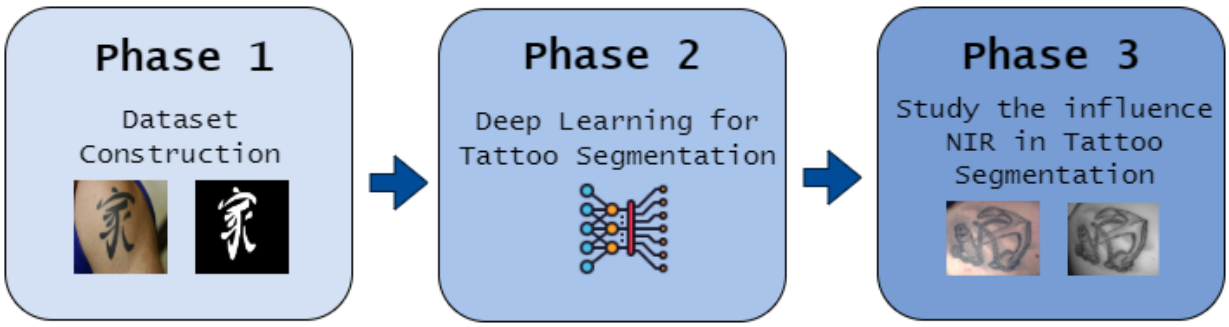


Figure 1.12: Sequence of overview of the project

1.6 Document Structure

There are four chapters in this dissertation. The first chapter covers the subject of tattoos. It outlines several tattoo removal methods used nowadays, as well as the exponential growth of these treatments in

the global market and which procedures need to be optimized. Furthermore, it relates to the project's contributions as well as its structure.

Then, in the second chapter, it is demonstrated how the dataset that serves as the basis for training the artificial intelligence models and testing for all models was built.

In the third chapter, a comparison of different deep learning models for tattoo segmentation is done, which will aid in tattoo removal treatment planning. First, the theme is introduced, and all principles are described, referring to the state-of-the-art of classic tattoo segmentation methods as well as advanced artificial intelligence-based methods. Following that, it is discussed how the DL algorithm was created. Various models and their outcomes are compared. Finally, a debate is held about all of the approaches used.

The NIR dataset is then introduced, and all bibliographic documents on the subject are addressed. The third chapter will look at how near-infrared images can aid in tattoo segmentation. It is made use of NIR in the system and recognizes their value and influence. This chapter ends with a discussion and conclusion about the work done.

Ultimately, an overall conclusion is made about the dissertation, and future work is presented.

Chapter 2: Dataset Construction

In this chapter, the construction of a tattoo dataset is presented. It will be explained the significance of a dataset, along with existing databases for projects involving tattoos, and how the pre-processing was carried out before reaching the final dataset.

2.1 Introduction

A dataset is a collection of different kinds of data that has been digitally stored. Any project using machine learning needs data as its primary input. To address different AI difficulties, such as image or video classification, object identification, face recognition, emotion, classification, voice analytics, and stock market forecasts - datasets are often composed of images, texts, sounds, videos, and numerical data points. It is not conceivable to implement an AI system without a good dataset. DL models are data-hungry and need a lot of data to build the best model or a strong system. The data must be good not only in quantity but also in quality. Even with excellent machine learning algorithms, the system fails without a proper dataset. Data preparation and interpretation is one of the most crucial and time-consuming activities of the Machine Learning project lifecycle, according to The State of Data Science 2020 study [41].

In machine learning, the term "ground truth" refers to the reality that will be used to train the machine learning model. When using a labeled dataset to train or validate a model, the target is sometimes referred to as "ground truth." When inferring, a segmentation model predicts a segmentation output that, if the ground truth label is available, can be compared with it [42]. In other words, the ground truth makes the machine learning model learn, as it tells it what it is supposed to learn because the ground truth indicates what is correct/on target and what is not. The model learns successively until it manages to predict the ground truth, generating an output. To evaluate the model, the "ground truth" generated by the model can be compared with the one generated manually when building the dataset. Figure 2.1 illustrates that process.

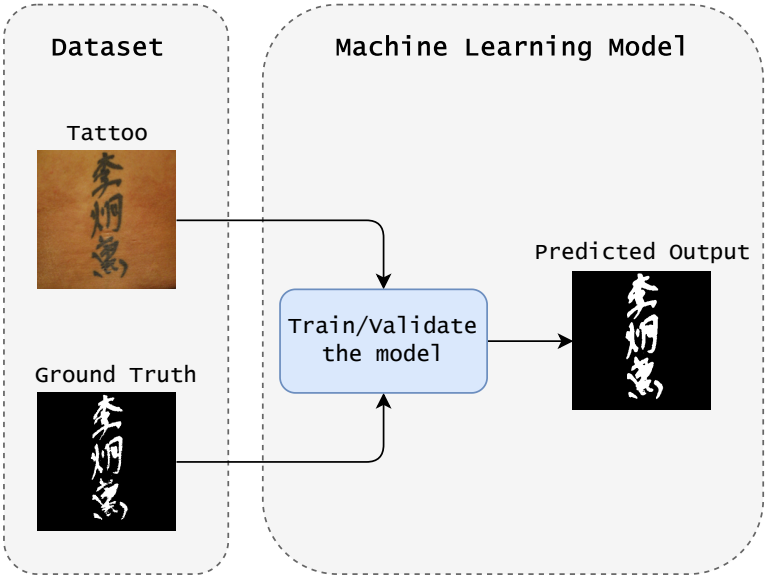


Figure 2.1: Tattoo and corrresponding ground truth mask to train the model

Therefore, it is necessary to build a dataset that is properly divided into training, validation, and test, with the respective ground truth masks for the machine learning models to learn. As a result, this chapter discusses the datasets that have been identified in the literature as well as every step that must be taken in order to produce an adequate dataset.

2.2 Related Work

To implement a DL algorithm, it is necessary to prepare the data, therefore it is essential to have a dataset that is adequate for the purpose intended to achieve and be of high quality. As a result, the dataset must be free of incoherent data, as accurate as possible, and have all of the necessary qualities for an algorithm to accomplish its goal [43]. So, to implement an artificial intelligence algorithm, it is necessary to obtain a reliable dataset. The objective is to gather the maximum number of tattoo images for further segmentation.

There are a lot of tattoo images on the internet, but only a few databases are public, as some images contain more intimate parts, and others may have gang-associated tattoos. Most works on tattoos found when reviewing the literature aim to help in forensic science, in the identification of suspects or criminals.

There are several tattoo datasets accessible, including Tatt-C, Tatt-E, WebTattoo, DeMSI, Flickr, and WVU-MediaTatt, however, some are not publicly available [44] [45] [46] [47] [48] [49], as shown in Table 2.1. The Flickr photographs were acquired from various points of view, poses, and complex backgrounds. Images range in raw size from 72 by 95 pixels to 500 by 500 pixels. The entire dataset contains color im-

ages. It features around 10000 photos, however, only 5740 of them have tattoos. To acquire this dataset it was necessary to sign an agreement file with the Nanyang Technological University. The database of the DeMSI tattoo dataset can be found in the Faculty of Electrical Engineering and Computing - University of Zagreb repository. This dataset is open to the public and can be downloaded. Tattoo images come in a variety of shapes, places, and situations. Most of them are in color. The image sizes are random and vary from image to image, ranging from 500x159 to 600x450. The WebTattoo dataset also contains random image dimensions since this dataset consists of a collection of different images present on the internet, however, it is notorious that in some cases the dimensions are smaller than those of the other datasets, for example, dimensions of 300 by 262 pixels and even smaller. However, it also presents other images with larger dimensions. The dataset consists of around 5000 images, mostly in color. To acquire this dataset, it was necessary to sign a license agreement. Annotations for tattoo detection are included in the WebTattoo and the DeMSI. However, these annotations are neither well ordered nor reliable for segmentation.

Table 2.1: Datasets available to segment tattoo

Dataset Name	**Number of images**	**Availability**	**Source**
DeMSI	1452	Available - public	FER - University of Zagreb
Flickr	10000	Available - with approval	NTU - Nanyang Technological University
WebTattoo	5000	Available - with approval	Website
Tatt-C	——	Not available	NIST
Tatt-E	——	Not available	NIST
WVU-MediaTatt	——	Not available	Social Media Website

It is concluded that the available datasets are DeMSI, Flickr, and Webtattoo.
Once the images from the datasets are organized, it is necessary to start their segmentation, specifically, the manual segmentation of the tattoo to create the ground-truth. In order to achieve that, it is imperative to prepare the data.

2.3 Data Preparation

In this section, it is explained all the steps implemented to get the ground-truth/masks to create the final dataset and therefore the neural networks can learn. First, the image selection procedure is explained, which focuses on the actual goal of tattoo removal in a medical environment, therefore the photos must not be too random. Then, from Photoshop to Matlab, it is described how the ground truth of each image was made. Finally, the images that compose the final dataset are divided into training, validation, and test set.

2.3.1 Images Selection

The preprocessed dataset is made by combining existing datasets and changing them to achieve the desired outcome. Considering the context of the tattoo removal problem, not all photographs of people with tattoos are relevant, because, in the actual treatment, the surrounding environment will be controlled, with a nearly non-existent background and with the focus on the tattoo (the tattoo must be centered and the greater part of the image).

That is why, from all of the photographs, only the relevant ones were selected, as some have a confusing background, others have the focus not on the tattoo, and others have too much information in the photograph.

For the purpose of distinguishing between images with and without tattoos, the Flickr dataset was developed. Thus, there are 4,260 non-tattoo images and 5,740 tattoo images in the database [50]. The University of Zagreb's (Faculty of Electrical Engineering and Computing) website hosts the DeMSI dataset, which contains 1452 tattoo images stored in its repository [51]. Around 5000 images make up the Webtattoo dataset, which was developed for tattoo detection purposes [45] [52]. The majority of the images, however, contain noise, or poor quality, too many artifacts, and complex backgrounds, therefore they don't all match the specifications. Only images that clearly show the tattoo were picked in order to remove noise. Additionally, the images needed to be manually segmented, producing a total of 770 tattoo images.

The first step is to select appropriate images to build an adequate dataset for segmentation, which is a collection of images with centered tattoos that are free of noise and have some variety. It is critical to select just images that will aid the system in segmentation. Low-quality images or images that do not focus on the tattoo are undesirable. Furthermore, images with accessories such as bracelets, earrings, and watches can be misinterpreted as tattoos. Also, backgrounds with too much information are likewise detrimental to the dataset. Figure 2.2 illustrates the process of image selection.

Images like the ones in the red box of figure 2.2 introduce a high variance in the dataset, making it imbalanced. Deep learning algorithms rarely learn with these images because they must focus on multiple image attributes, resulting in learning the noise or irrelevant information within the dataset, which causes overfitting. As a result, only images of high quality with the tattoo as the dominant part of the image were selected.

Figure 2.2: Comparison of some images that do not correspond to what was intended (red box) and some of the selected images (green box)

2.3.2 Ground Truth Creation

The images must be manually segmented after being chosen. Since trustworthy data is essential for DL models to be able to learn, the ground truth was created manually.

In image processing, the "ground truth" refers to what is considered to be correct and "true" within the data being examined. The ground truth in this particular situation indicates where the tattoo is. An ideal ground truth mask correctly identifies the original goal while omitting everything else, which in this case would identify the specific location of the tattoo.

To create the ground truth masks, a manual segmentation using the Paint.net program was first carried out. The process is straightforward and consists of choosing the desired areas, that also is, erasing everything that wasn't a tattoo, and then saving the results.

Figure 2.3 illustrates the initial stages of creating the ground truth masks for two dataset images using the Paint.net application.

Figure 2.3: Segmentation of the original images using the Paint.net software

The second step is to binary transform these images into black and white, where black represents the undesirable part and white represents the desired part (tattoo).

To transform the RGB images into binary images, MATLAB R2020b was used. In MATLAB, a folder was created for the output images and the folder for the segmented RGB images was loaded. The images within this folder were converted to grayscale so that they only had one channel. Then a threshold is implemented which transforms the 2D image into a binary image, turning all values above the threshold value to 1s and the rest to 0s. Thus, it is visually easier to select the area of interest and ignore the remaining areas. Finally, these images are saved in the initially created output folder. The process of binarize the images is shown in Figure 2.4.

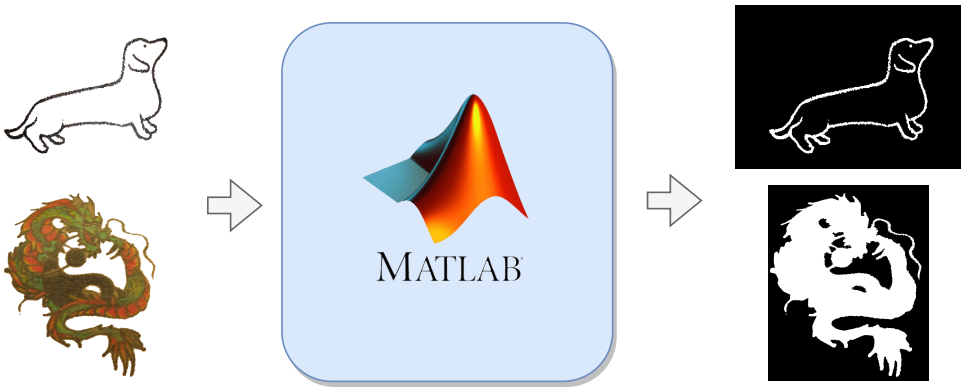


Figure 2.4: Normalization of images to 0 and 1 with MATLAB

In this way, the images have their respective ground truth, however most of the images don't have the same size. Most CNN architectures use an image size that is a multiple of 2. This way you can perform downsampling without problems. Also, using images of different sizes can cause errors, mismatching size exceptions, or even a random size output image. Thus, a resize was done so that the images would have a uniform size of 384x384. In the Figure 2.5, it is possible to see the images and the corresponding ground

truth resized.

Figure 2.5: Examples of images and the corresponding ground truth mask, already resized

2.3.3 Dataset Split

Once the tattoo images are organized with their ground truth mask, and all of the same sizes, the dataset is divided into three subsets: train set, validation set, and test set.

The separation into training, validation, and testing allows a better generalization of the DL models. The training set helps both in creating and training the model. In other words, the model can learn from this data. Additionally, the validation set is utilized to evaluate the model and adjust the hyperparameters during training. The test set is essential because it enables an evaluation of the model's global performance, which allows later model comparisons.

The dataset consisted of 770 manually segmented images and 638 tattoo images were for the training set, 76 for validation and 56 images for testing, as shown in Figure 2.6.

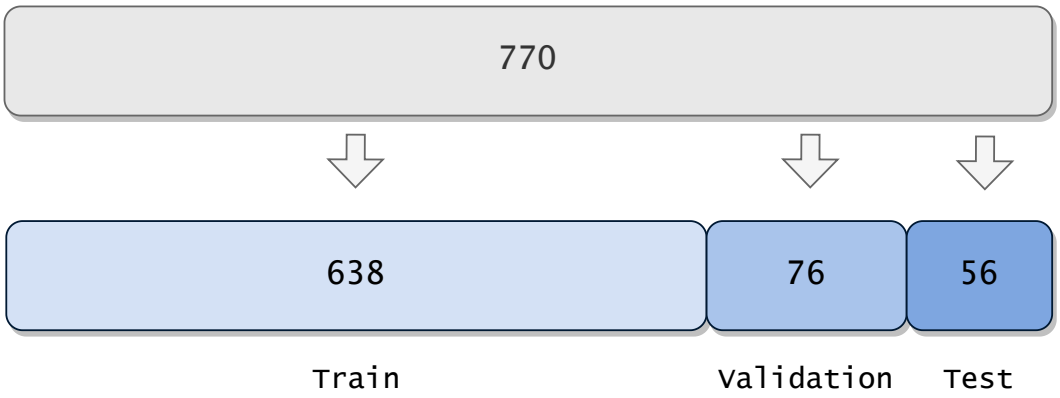


Figure 2.6: Dataset split into train, validation and test

Chapter 3: Different Techniques for Tattoo Segmentation

This chapter presents the implementation of two segmentation algorithms, one based on primitive techniques and the other on current approaches, in this instance DL. Finally, the work done on several defined metrics is discussed.

3.1 Introduction

As shown in the previous chapter, the need to improve clinical processes is increasing. As more individuals have tattoos and as many people regret obtaining tattoos, it is imperative to improve the tattoo removal procedure. Although this number is rising, the success of laser tattoo removal still entirely relies on the expertise and performance of the physician. AI-based segmentation successfully segments the tattoo. This segmentation/characterization may be used in the assistance of the physician in tattoo removal in the future. In addition to tattoo removal, tattoo segmentation can aid in criminal identification in the field of forensic science, as well as the development of platforms that enable the detection or removal of other skin problems.

This chapter aims to present the preparation of the tattoo dataset and six DL-based networks for the segmentation of tattoos, namely U-Net which served as the basis for development, and then the remaining networks that are derived from it, such as DynUNet, UNETR, RegUNet, SegResNetVAE, and SegResNet. In addition to these techniques, a method based on thresholds is presented, which serves as a term of comparison between advanced and traditional methods.

To evaluate the performance of the methods, six metrics were chosen, specifically, Dice Coefficient, Jaccard Index, Accuracy, Recall, Specificity, and Precision.

3.2 Related Work

This section presents the analysis of bibliographic documents that are of interest to the development of the project.

First, documents related to the traditional method are analyzed, without resorting to AI techniques.

Next, it is presented how the introduction of AI more specifically DL in a clinical context can improve these processes. Finally, a literature review is carried out on how a DL model is prepared and which metrics are commonly used to evaluate the model's performance.

3.2.1 Traditional Methods for Tattoo Segmentation

Segmenting tattoos is not an easy task due to the immensity of shapes, colors that the tattoo can have; the similarity it sometimes has to the skin; it can also be confused with accessories (bracelets, watches, glasses, along with others).
However, several methods of segmenting tattoos have been studied.

- **Tattoo skin detection and segmentation using image negative method** - Three steps make up the procedure. The first is skin detection, which employs a variety of skin patches to perform human skin color segmentation using the HSV model, with an emphasis on Asian skin color. The image negative approach is then employed in the second stage to detect the clear graphic picture of the tattoo segment. Finally, the tattoo segment is excised from the skin area of the negative image, generating the tattoo's negative image [53], as can be seen in Figure 3.1.

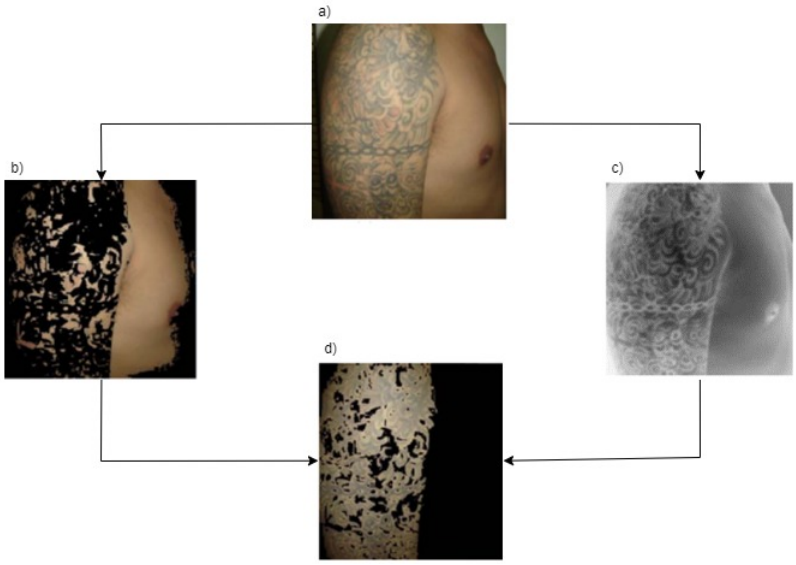


Figure 3.1: Tattoo segmentation based on image negative method [53]

a) Original image; b) Skin detection; c) Image negative; d) Tattoo segmentation

- **Using skin color space and K-mean clustering to segment tattoos** - To detect the skin region, K-mean clustering and human skin color segmentation in LAB color space are applied. The mathematical morphological processing is then used to smooth the clear graphic image of the

tattoo portion. Finally, the color tattoos segmentation is recovered and recognized based on the connection of tattoo region [54]. Figure 3.2 shows the sequence of this technique.

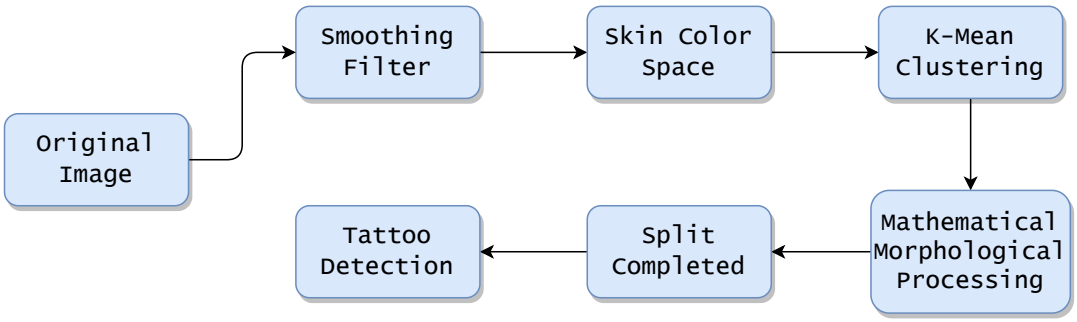


Figure 3.2: K-mean clustering algorithm to segment tattoos [54]

- **Graph-cut tattoo segmentation** - The tattoo segmentation problem is characterized as locating skin pixels surrounding a tattoo with the assumption that the tattoo is surrounded by skin. As a result, all pixels in the image do not need to be segmented. Only the regions along the image edges are considered for segmentation. A probabilistic skin color model based on a gaussian model is used to detect skin pixels in locations near image edges. The skin color model fails to detect the skin around the tattoo if there are sections in the backdrop that have a color comparable to skin. A visual saliency map was also employed to focus on the skin around tattoo areas. Following segmentation, it is determined which set of skin pixels are joined to make a closed contour with a tattoo. Tattoo refers to the areas that are surrounded by closed shapes [55], as shown in Figure 3.3.

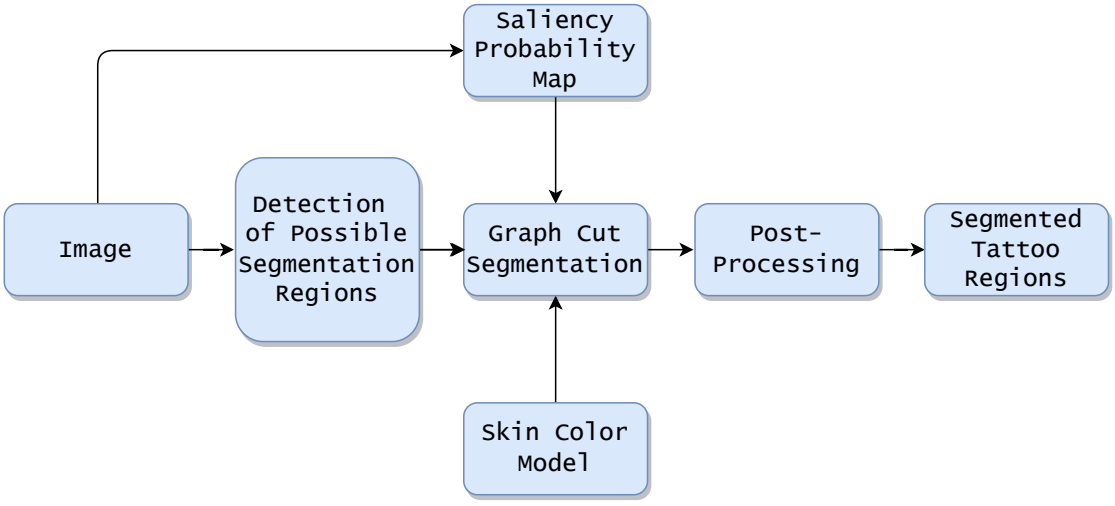


Figure 3.3: Graph-cut segmentation algorithm [55]

- **Unsupervised tattoo segmentation using bottom-Up and top-Down cues** - This method involves splitting each tattoo image into clusters using a bottom-up approach, learning to merge the

clusters containing skin, and then distinguishing tattoos from other skin in the image itself using a top-down approach. Tattoo segmentation with an undetermined number of clusters is adapted as figure-ground segmentation [56], as can be seen Figure 3.4.

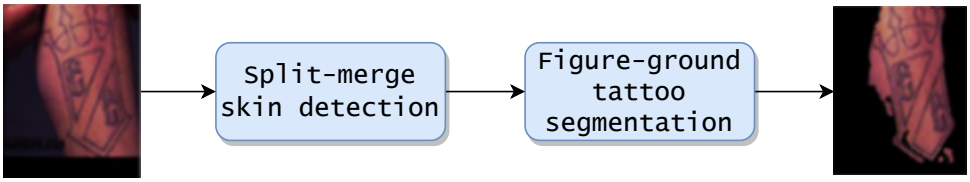


Figure 3.4: Bottom-up and top-down cues segmentation algorithm [56]

- **Manual segmentation with Paint.net** - Manual segmentation is slower but is more effective. Using paint.net software, it is possible to segment the tattoo in the selected way, erasing everything undesired.

  Paint.net is free and open-source computer software for manipulating and altering images and photographs [57].

These primitive/fundamental methods have some disadvantages mainly for varied datasets. Most of these methods can not deal with all types of tattoos because it would be very difficult to find a solution that covers all images. These methods are sensitive to noise, have difficulty processing unimodal histograms, and may confuse accessories/objects with tattoos. The same algorithm could have good results in some cases and bad in others. In this way, the most advanced and modern methods, such as AI, come to overcome these problems and present solutions for more varied images, obtaining better results than traditional methods.

3.2.2 Advanced Methods for Tattoo Segmentation

Not only were traditional tattoo segmentation methods studied, but also advanced methods for tattoo segmentation. The study carried out in this section and the literature review on these methods allow a greater understanding of more modern techniques such as Artificial Intelligence.

DL in Clinical Procedures

One strategy to develop Artificial Intelligence systems is Machine Learning. Machine Learning is the discipline of utilizing algorithms to analyze data, learn from it, and then determine or predict anything about the outside world. As a result, rather than manually writing software programs with a particular set of instructions to perform a task, the machine is "trained" using vast amounts of data and algorithms, giving it

the capacity to learn how to carry out the operation. Deep Learning is a Machine Learning implementation approach that has the potential to shorten therapy time and enhancing treatment effectiveness and will almost certainly assist healthcare by moving away from traditional medicine, where one medicine fits all, and toward targeted treatments and individualized therapies. AI in medicine makes doctoring easier and less error-prone while making therapy faster and safer for patients [35].

In the tattoo removal service, Deep Learning can help with better tattoo characterization, as well as defining the parameters to remove tattoos by laser therapy.

Deep Learning in Image Segmentation

Various DL techniques have recently appeared to solve numerous real life problems. With several applications, including scene interpretation, medical image analysis, augmented reality, video surveillance, and picture compression, robotic sensing, image segmentation is a fundamental topic in image processing and computer vision.

From the early methods, such as histogram-based bundling, watersheds, k-means clustering, and thresholding, multiple image segmentation algorithms have been developed in the literature. However, for the past several years, DL models have produced a new generation of image segmentation models with amazing performance increases, frequently obtaining the best accuracy rates on common benchmarks, leading to a paradigmatic change in the area [58].

Computer vision faces some difficulties, such as lighting, perspective variation, or deformation [59] and DL typically produces more accurate results than traditional methods.

Convolutional Neural Networks (CNNs)

CNNs are one of the most well-known deep learning techniques in computer vision. They are among the most important categories in image recognition and classification.

When dealing with large image resolutions, the computational cost and resulting memory occupation are implicit, however, the use of CNNs manages to overcome these disadvantages and produce good results at this level [59].

In CNN models, each image is processed through a series of convolution layers with filters (Kernels), Pooling, and the Softmax function to define an object with probabilistic values ranging from 0 to 1 [59]. Figure 3.5 shows the CNN sequence.

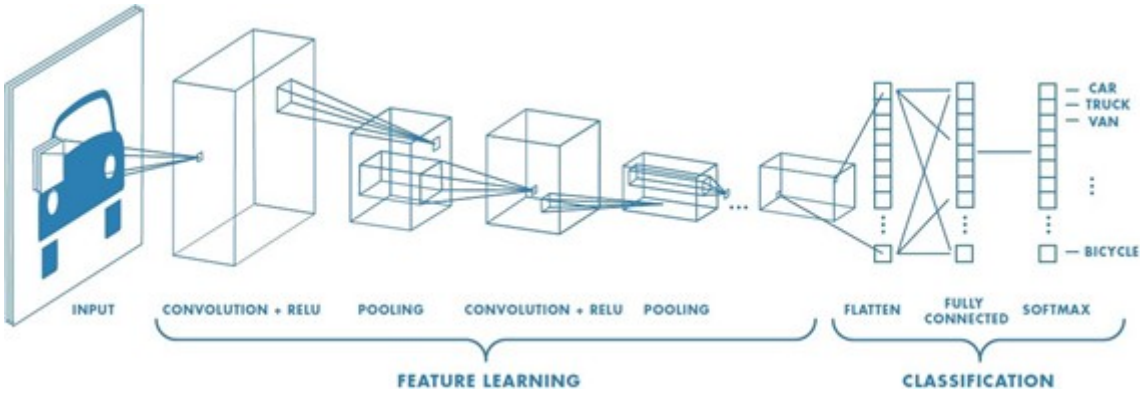


Figure 3.5: CNN sequence [60]

Convolution Layer - Convolution is the first layer in the process of extracting features from an input image. By learning features of an image with small squares of input data, convolution maintains the correlation between pixels. It is a mathematical operation with two inputs: kernel or an image matrix and a filter [60].

Non Linearity (ReLU) - ReLU stands for Rectified Linear Unit. ReLU is an activation function. The activation function in a neural network is in control of converting the node's summed weighted input into the node's activation or output for that input. If the input is positive, it will be output directly; otherwise, it will be output as zero [60]. In Figure 3.6 is represented the ReLU activation function.

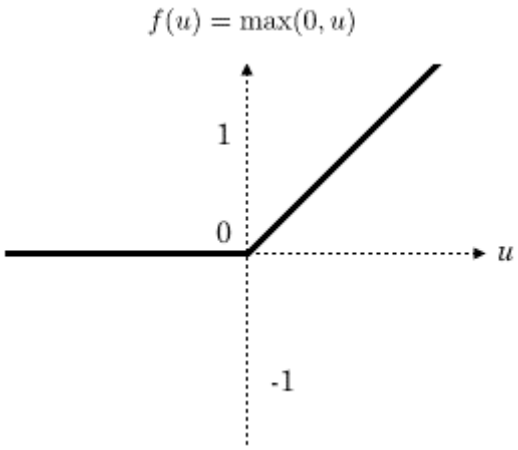


Figure 3.6: ReLU representation [61]

Other nonlinear functions, such as tanh or sigmoid, can be used in place of ReLU. ReLU is used by most data scientists because it outperforms the other two, in many systems.

Pooling Layer - When the images are too large, the pooling layer's part will indeed reduce the number of parameters. Spatial pooling, also known as subsampling or downsampling, reduces the dimensionality of each map while retaining critical information [60] [61]. There are several types of spatial pooling:

- Max Pooling
- Average Pooling
- Sum Pooling

Fully Connected Layer - The matrix is flattened to form a vector, which is then fed into a fully connected layer. The fully connected layers connect features to construct a model [60] [61]. Figure 3.7 represents a fully connected layer.

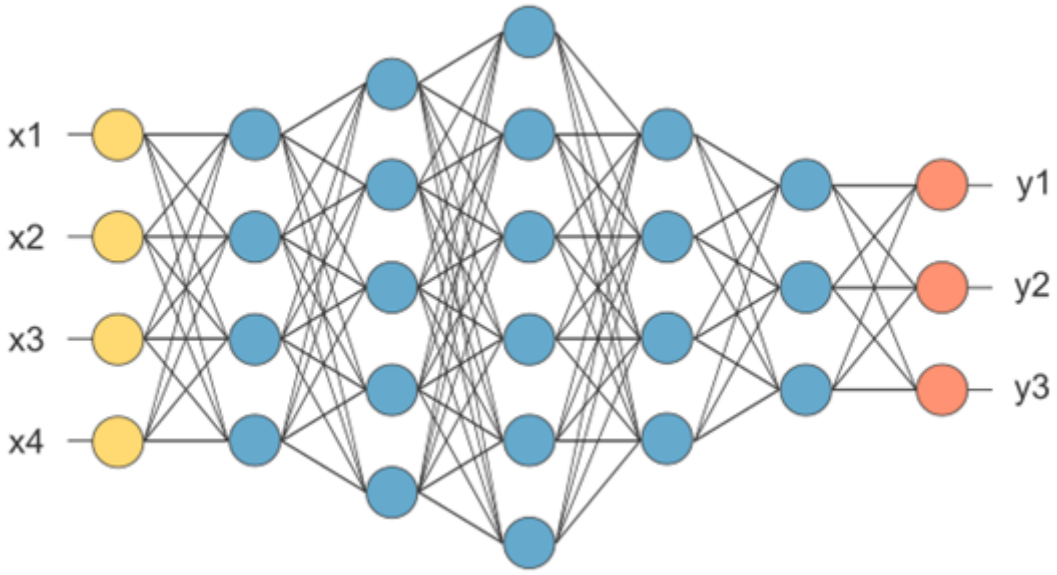


Figure 3.7: Fully connected layer representation [60]

Finally, to classify the outputs, an activation function such as softmax or sigmoid is used.

Softmax - The Softmax layer is the final layer of the convolutional neural network. This layer is in charge of predicting the likelihood of the input belonging to one of the two labels [61].

CNNs architectures

There are various CNNs architectures available, which have been critical in building algorithms that power and will power AI as a whole in the near future, such as LeNet, AlexNet, VGGNet, GoogLeNet, ResNet, ZFNet, Siamese and many others [62]. In [63], AlexNet and Siamese are presented as networks to detect tattoos.

AlexNet - It is made up of five convolutional layers, which are followed by three fully connected layers. One of the most notable differences between AlexNet and other object detection algorithms is the use of ReLU for the non-linear part of the algorithm rather than the commonly used Sigmoid or Tanh functions

in traditional neural networks. AlexNet uses ReLU's quickest training to make the algorithm faster as well [63] [64]. In Figure 3.8 can be seen the AlexNet network.

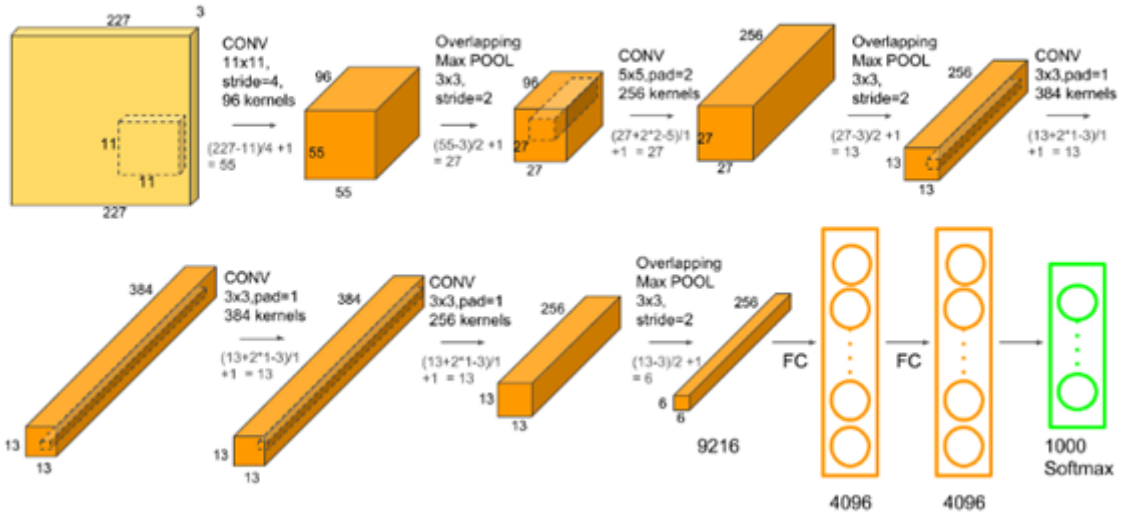


Figure 3.8: AlexNet architecture [64]

Siamese - As the name implies, a Siamese network is an architecture with two parallel layers as can be seen in Figure 3.9. Instead of learning to classify its inputs using classification loss functions, the model in this architecture learns to differentiate between two given inputs [65]. It compares two inputs using a similarity metric and determines whether they are the same or not. This network is made up of two identical neural networks with similar parameters, each with one input data point. It extracts similar types of features in the middle layer because the weights and biases are the same. The final layers of these networks are fed to a loss function layer, which computes the similarity of the two inputs [66].

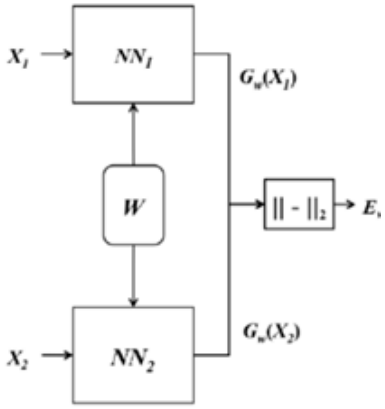


Figure 3.9: Siamese architecture [66]

U-Net - Semantic Segmentation

Semantic image segmentation aims to assign a class to each pixel in an image. This task is usually referred to as dense prediction because we're predicting for every pixel in the image.

The intended output in semantic segmentation is not merely labels and bounding box parameters, as it was in the prior challenges. The result is a high-resolution image (usually the same size as the input) with each pixel classified into a different class. As a result, the classification is at the pixel level [59].

Olaf Ronneberger created the U-Net for Biomedical Image Segmentation. There are two paths in the architecture. The compression path (also known as the encoder) is the first path, and it is used to capture the image's context. The encoder is simply a convolutional and maximum pooling layer stack. The symmetric expanding path (also known as the decoder) is the second path, and it is employed to achieve exact localisation via transposed convolutions [67].

As a result, it is an end-to-end fully convolutional network (FCN) [59]. In Figure 3.10 is represented the U-Net structure.

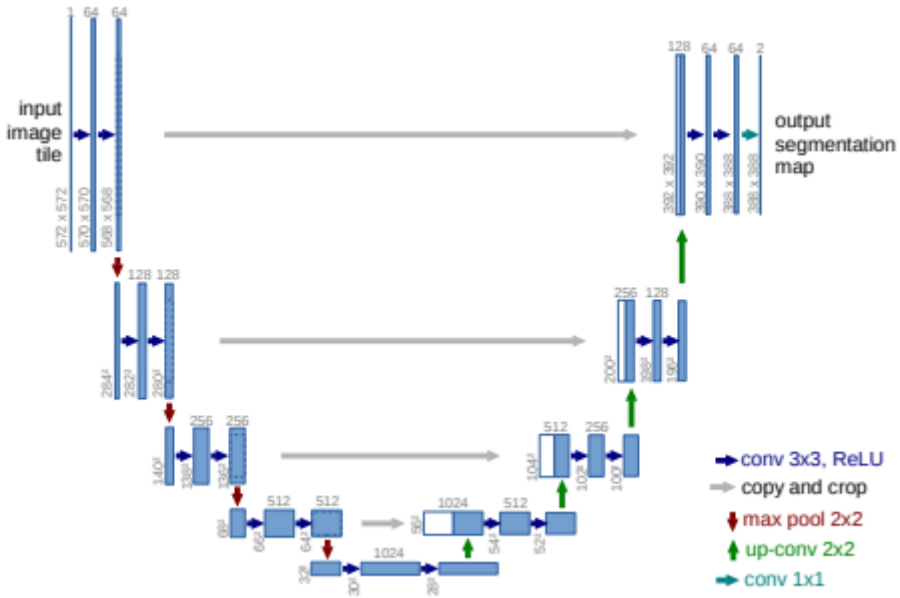


Figure 3.10: U-net architecture [67]

Algorithm Preparation

A deep learning algorithm must first be prepared, which entails carrying out a series of tasks like handling the dataset, performing data augmentation, defining the model, choosing the loss, and saving the results.

- **Dataset Structure:** High-quality images are necessary for deep learning methods to train, validate, and test algorithms in the best possible way. The lack of suitably vast, curated, and representative training data is one of the main barriers to the development and clinical application of AI algorithms (eg, ground-truth).

  Finding public datasets that might have images in the desired domain is the first step. It is frequently necessary to have access keys, ethical approvals, or authorization licenses from specific institutions. The dataset participants' consent may be necessary for some circumstances. A local data storage or external data storage is used to transfer and store data. The chosen images are then selected after the dataset has been analyzed.The data must then be organized in standardized, machine-readable ways. Linking the images to ground truth data, which might be one or more labels, segmentations, or electronic phenotypes, is the last step [68].

- **Train, Validation and Test sets:** When the dataset is built, it has to be divided into training, validation, and testing.

  The training set is responsible for training the model. The model learns from these data. Validation is responsible for adjusting the hyperparameters and evaluating the training. On the other hand, the test set is only employed at the end of training and validation, with the purpose of assessing the model's overall performance [69].

- **Data Augmentation:** The training dataset is made larger artificially by a process known as "data augmentation." This is commonly done in medical imaging by applying modifications to both the ground truth and the images equally, resulting in distorted representations of the training data.

  Transformations including crops, rotations, or elastic deformations are commonly used in dataset augmentation techniques to create training images that closely mimic a given training example [70].

  Data augmentation can increase the training dataset and add new data. As a result, it is used to reduce overfitting.

- **Model Definition:** For any deep learning algorithm, it is necessary to choose a model. There are several factors to choose a model, among them is performance which is a key aspect to consider when selecting a model.

  Models that maximize the performance of the output are the best ones. Explainability is another significant factor. It's crucial to select an understandable model while making a choice. The results must be simple to comprehend, and the algorithm must be easy to interpret. Another important

factor is the size of the dataset. The model used must be appropriate for the size of the dataset. Certain models can process more data than others, while some are better suited to smaller datasets.

Finally, it is necessary to realize that the best performance is not always desired, time and cost are very important variables when choosing a model [71].

- **Loss Function:** The stochastic gradient descent optimization algorithm is used to train deep learning neural networks. The error for the model's current state must be regularly calculated as part of the optimization algorithm. This necessitates the selection of an error function, which is typically referred to as a loss function, that may be used to estimate the loss of the model and be used to update the weights to lower the loss on the subsequent assessment [72].

  It is essential to assess the problem before selecting the loss function. The typical selection of a loss function for Regression Problems is based on the Mean Squared Error. When using Binary or Multi-Class Classification, Cross-Entropy is typically used as a function. The most valuable loss functions for Segmentation tasks are those related to the Dice.

- **Saving Results:** At the end of a deep learning algorithm, it is necessary to save the results, as they will allow evaluating the performance of the model. The evaluation can be done through training and validation graphs, to understand possible cases of overfitting or underfitting, or for a more accurate evaluation, save the output and compare it with the ground truth. The results must be saved in the same way so that later it is possible to compare models or model components.

To evaluate the results there are several metrics available, such as Precision, Accuracy, Dice, and Jaccard Index, among others.

3.2.3 Metrics

In any machine learning or deep learning model, it is necessary to evaluate the model's quality in relation to the task's aim.
Statistical functions have been developed to ensure in the evaluation of error capacity and model reliability. Metrics are the name given to certain mathematical functions. [59]

There are different types of evaluation metrics [73]:

- Confusion matrix - Gives a matrix as output and describes the complete performance of the model. Table 3.1 represents the confusion matrix table.

Table 3.1: Confusion matrix

		True Value	
		Positive	Negative
Predicted Value (PV)	Positive	**True Positive (TP)**	**False Positive (FP)**
	Negative	**False Negative (FN)**	**True Negative (TN)**

True Positive (TP): values are those that fall into the positive category and were classed as such.

False Positive (FP): Values that should have been categorised as negative but were instead classified as positive.

False Negative (FN): Positive values that were incorrectly labeled as negative.

True Negative (TN): Values that are in the negative range and have been identified as such [59].

- Classification accuracy - The number of correct predictions divided by the total number of input samples is the classification accuracy [73]. The accuracy Equation is the 3.1.

$$Accuracy = \frac{TP + TN}{TOTAL} \tag{3.1}$$

- Precision - The number of correct true positive results divided by the number of positive results predicted by the classifier equals precision [73], as can be seen in Equation 3.2.

$$Precision = \frac{TP}{TP + FP} \tag{3.2}$$

- Recall - The number of correct positive results divided by the total number of all relevant samples is what is known as recall [73]. Recall Equation is presented in 3.3.

$$Recall = \frac{TP}{TP + FN} \tag{3.3}$$

- F1 Score or Dice - The Harmonic Mean of precision and recall is the F1 Score. F1 Score has a range of [0, 1]. It indicates how accurate the classifier is [59] [73]. The Dice Equation can be seen in 3.4.

$$Dice = \frac{2 * TP}{FN + (2 * TP) + FP} \tag{3.4}$$

- Jaccard Index (JI) - The JI is a technique for calculating the percentage overlap between the ground truth mask and the predicted output [74]. Equation 3.5 shows the Jaccard Index mathematical formula.

$$JI = \frac{TP}{TN + TP + FP} \tag{3.5}$$

- Specificity - Specificity assesses the capability to correctly identify true negative classes. Its Equation is calculated by dividing the True Negatives by the sum of True Negatives and False Positives, as shown in 3.6

$$Specificity = \frac{TN}{TN + FP} \tag{3.6}$$

- Mean Absolute Error- Mean Absolute Error is the average of the difference between the True Values and the Predicted Values [73].

- Mean Squared Error - The sole difference between Mean Squared Error and Mean Absolute Error is that MSE takes the average of the square of the difference between the True and Predicted Values [73].

3.3 Implemented Methods

A review of the literature was conducted in the previous section, and two methodologies are described in this section. One approach is based on conventional techniques and another is based on modern techniques, such as the case of Deep Learning. The goal of implementing the two methods is to compare which strategies are better and to learn different methods of segmentation.

3.3.1 Deep Learning Method

Once the dataset is already split, the DL algorithm just loads this data, making sure the images and ground truths match are correct.

Data Augmentation

Then, data augmentation was implemented. By transforming the original images, data augmentation intends to generate a larger and more varied dataset. This was accomplished using a Python library called **Albumentations**, which allows for fast and adaptable image augmentations and therefore boost the performance of deep convolutional neural networks. Albumentations provide a clear but effective image

augmentation interface for many computer vision tasks, such as object classification, image segmentation, and detection. Also, Albumentations are able to operate with different deep learning frameworks like PyTorch and Keras and have an impact on a variety of domains including satellite imagery, normal photos, medical images, or even industrial applications. Albumentations effectively implements a wide range of image transform operations that are optimized for better performance [75]. Some examples of transforms are presented in 3.11.

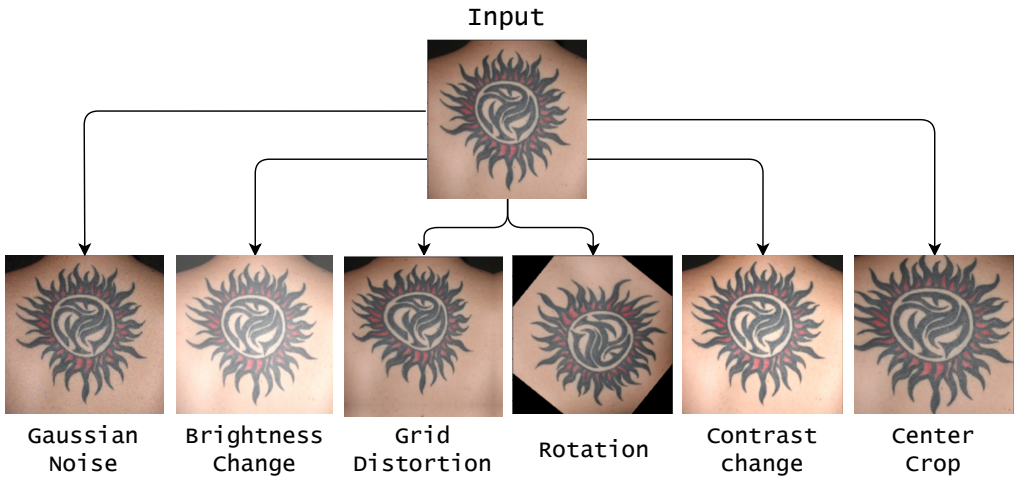


Figure 3.11: Example of transforms to implement data augmentation

Some of the transformations performed on the initial dataset from the Albumentations library were:

- **Rotate:** Rotate the input by a chosen random angle from a uniform distribution. The parameters were selected to allow the initial image to rotate between a limit angle of -180 and 180. Also, a probability of 20% was chosen to perform the transformation in the input images.
- **GaussNoise:** Apply gaussian noise to the input image with a variance range of 0 to 10. The probability of changing the input images is 10%.
- **RandomScale:** Randomly rescale the input image with a scaling factor range of -0.3 to 0.4. The probability of applying the transform chosen was 20%.
- **RandomBrightness:** Randomly change brightness of the input image, with a factor range for changing the brightness of -0.25 to 0.25. The probability of applying the transform chosen was 15%.
- **RandomContrast:** Randomly change the contrast of the input image, with a factor range for changing the contrast of -0.25 to 0.25. The probability of applying the transform chosen was 15%.

- **Downscale:** Decreases image quality by downscaling and upscaling back. The lower bound on the image scale range from 0.75 to 0.99.The probability of applying the transform chosen was 25%.
- **CenterCrop:** Crop the centre of an image. The probability of applying the transform chosen was 100%.

Neural Networks Architectures

After the data augmentation, it is necessary to define the neural networks. U-Net is known to be effective in segmentation tasks. However, more recent research has shown network variants that could be more effective for this purpose. In this regard, it was considered to use U-Net as a reference and five other neural networks in this segmentation task, namely UNETR, DynUNet, RegUNet, SegResNet, and SegResNetVAE. The networks consist of many layers, and when the model is trained, each convolutional layer employs filters to extract the needed features. In Figure 3.12 is illustrated a basic design of a deep neural network for segmentation.

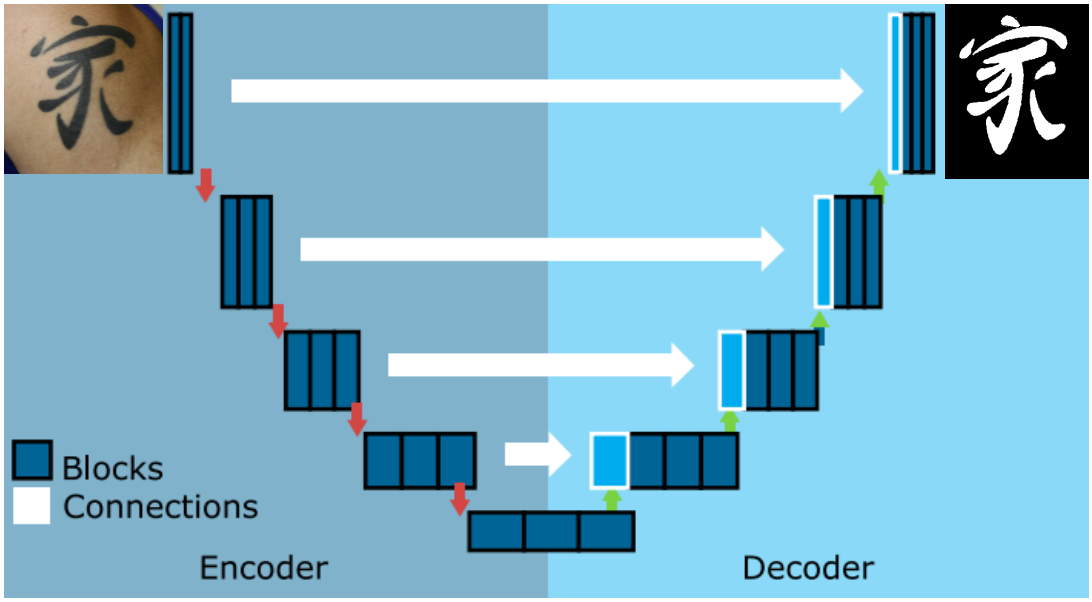


Figure 3.12: Deep neural network's basic schematic for segmentation (adapted from [76])

The U-Net architecture follows two paths. The first path is the compression path (also known as the encoder), which is used to capture the image's context. The encoder is a stack of convolutional and maximum pooling layers. The second path is the symmetric expanding path (also known as the decoder), which is used to achieve exact localization via transposed convolutions [67]. The U-net Transformer, also known as UNETR, is a Transformer-based architecture for medical image segmentation that makes use of a pure transformer as the encoder to learn sequence representations of the input volume, effectively

capturing the global multi-scale information. To compute the final semantic segmentation output, the transformer encoder is directly connected to a decoder via skip connections at various resolutions, much like a U-Net [77] . The DynUNet automatically adjusts its architectures to the provided image geometry and defines the additional steps around it by developing a set of rules that choose the best kernel sizes, strides, and network depth from the training set. It is a dynamic implementation of U-Net [78], in other words. The RegUNet implements an adapted U-Net model, it makes use of an encoder-decoder architecture to predict the general non-rigid transformation and it is best suited for image registration however can be used in segmentation [79]. The SegResNet architecture is an encoder-decoder-based convolutional neural network with hyperparameter optimization and tuning used to automatically select the hyperparameters [80]. The SegResNetVAE is identical to SegResNet, but with the addition of a variational autoencoder (VAE) branch to reconstruct the input images in combination with segmentation to regularize the encoder. In the Table 3.2 it is summarized all the neural networks implemented.

Table 3.2: Summary of all implemented neural networks

Networks	Date of publication	Summary
U-Net [67]	2015	The U-Net is divided into two sections: a decoder/expansion path on the right and an encoder/contraction path on the left.
DynUNet [78]	2018	DynUNet is a dynamic implementation of U-Net and adapts its structures to the provided image size automatically.
UNETR [77]	2021	UNETR uses a transformer as the encoder to efficiently collect the global multi-scale data and learn sequence representations of the input volume.
RegUNet [79]	2020	The encoder-decoder design of RegUNet, which is adapted from U-Net, predicts the general non-rigid transformation.
SegResNetVAE [80]	2018	SegResNetVAE is based on autoenconder regularization but with a varitational autoencoder
SegResNet [80]	2018	The SegResNet architecture is a convolutional neural network that employs autoencoder regularization.

Loss Function

Image segmentation is the process of classifying pixels into several types of elements, often those connected to the background and the central focus (e.g., tattoo). The networks are often biased toward the bigger element due to the disparity in the regions that the elements span. Some loss functions,

nevertheless, can get around this issue. Therefore, it is proposed to utilize **Dice Loss**, one of the most used loss functions, to fit the segmentation models to the training data [76].

A statistic called the Dice coefficient is used to determine how similar two samples are. By assessing the similarity of pixels that are spatially matched, its application to images may be expanded [76].

The coefficient is also included in the loss function and is shown by the expression 3.7:

$$DL(l,p) = 1 - \frac{2 * l * p + 1}{l + p + 1} \tag{3.7}$$

where p is the predicted probability of the label and the l is the label. The dice loss function that was used was **DiceLoss** from **MONAI**, which computes the average dice loss between two tensors and supports tasks with many classes and labels.

Hyperparameters Definition

The six DL networks were trained with the Dice Loss function, for **1000 epochs** using the training dataset, and their parameters were tuned to the optimal results. The models were assessed using training and validation data at each epoch. Finally, predictions for the test data were made using the trained models, and the resulting scores were determined using the ground-truth masks and evaluation metrics.

The **optimizer** chosen was **Adam**, since it is described in the literature as fast, as it trains a neural network in a short time and efficiently [81]. A **learning rate** of **1e-5** was chosen to regulate how much of the model must be altered in accordance to the estimated error for each update of the model weights. It was decided to use a **weight decay** of **1e-5** to manage the trade-off between having a satisfactory model or a overfitting model. The **batch size** used was **6** to fully exploit the GPU according to the available computational resources to complete the training. Table 3.3 summarizes the hyperparameters selected.

Table 3.3: Hyperparameters definition summary

Number of epochs	1000
Batch Size	6
Optimizer	Adam
Learning Rate	1e-5
Weight Decay	1e-5

Evaluation Metrics

Once the predictions have been generated, it is possible to evaluate the performance of the various networks. It is feasible to determine the results of the metrics that evaluate the model using the predictions

from the training models and the ground-truth masks.

The metrics that were selected to evaluate the models were as follows:

- Jaccard Index;
- Dice or F-1 Score;
- Precision;
- Recall;
- Specificity;
- Accuracy.

A more detailed explanation of these metrics is present in section 3.2.3 .

Deep Learning Method Overview

In the Figure 3.13 it is represented the overview of the Deep Learning method.

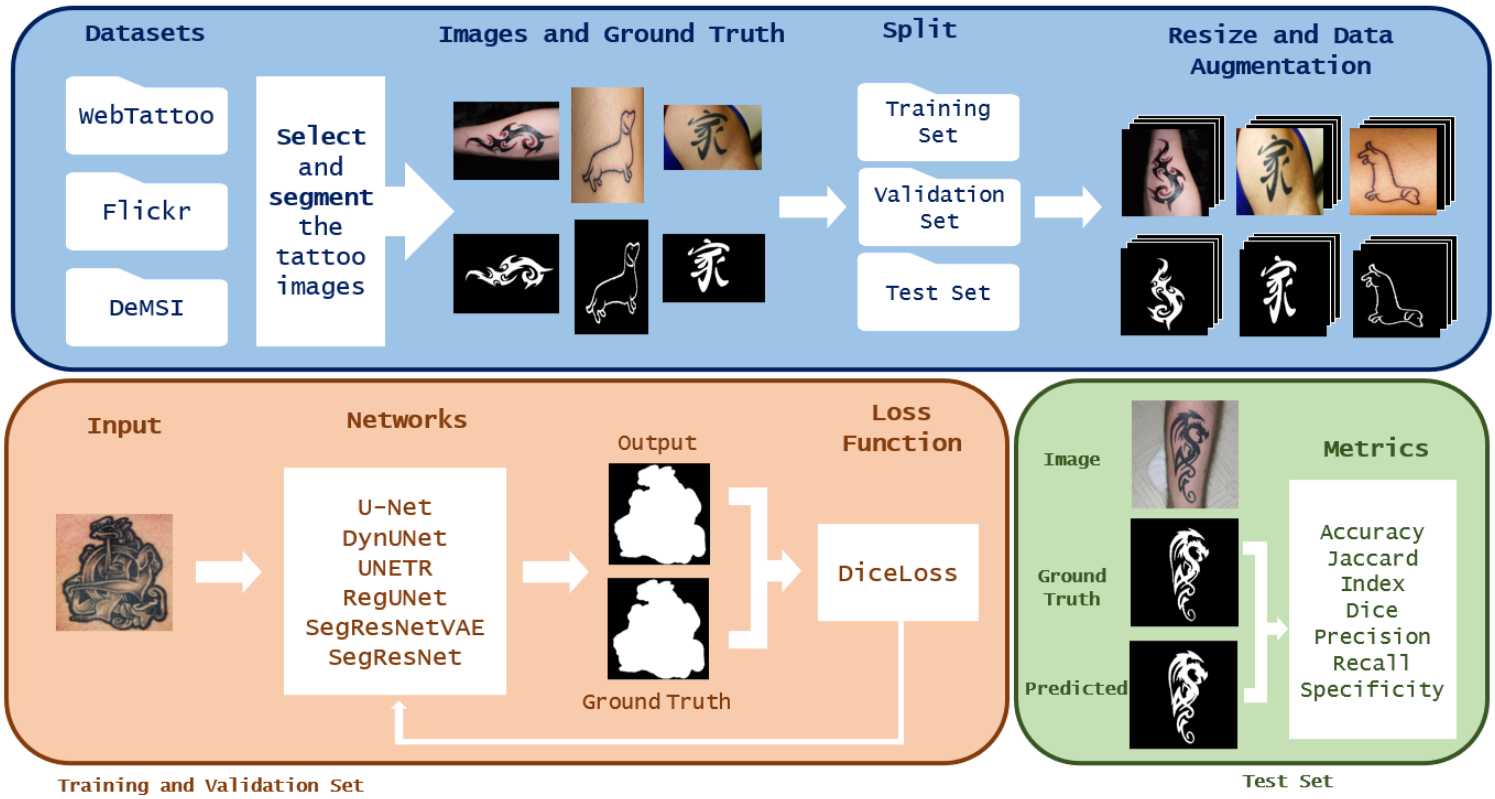


Figure 3.13: Sequential overview of the DL method

3.3.2 Traditional Method

For a comparative study, a segmentation method was carried out, using a traditional method. In order to implement this method it was necessary the use of different strategies, such as thresholding, and morphological operations.

As shown in section 3.2.1, traditional tattoo segmentation techniques frequently use skin segmentation as the first step, so it is possible to extract what is within the skin. Normally it is tattoos, skin tags/marks, or accessories.

Although several color spaces can segment the skin, combining different color spaces can increase the threshold's robustness. RGB is not recommended as a single colorspace because of the close relationship between the R, G, and B values and their reliance on outside factors like lighting. YCbCr can't segment by itself but makes it possible to separate luminance and chrominance.HSV is also valuable since its primary benefit is that it is invariant to sources of white light [82].

As a result, a threshold was determined for each color space, and their mixture produced better skin segmentation.

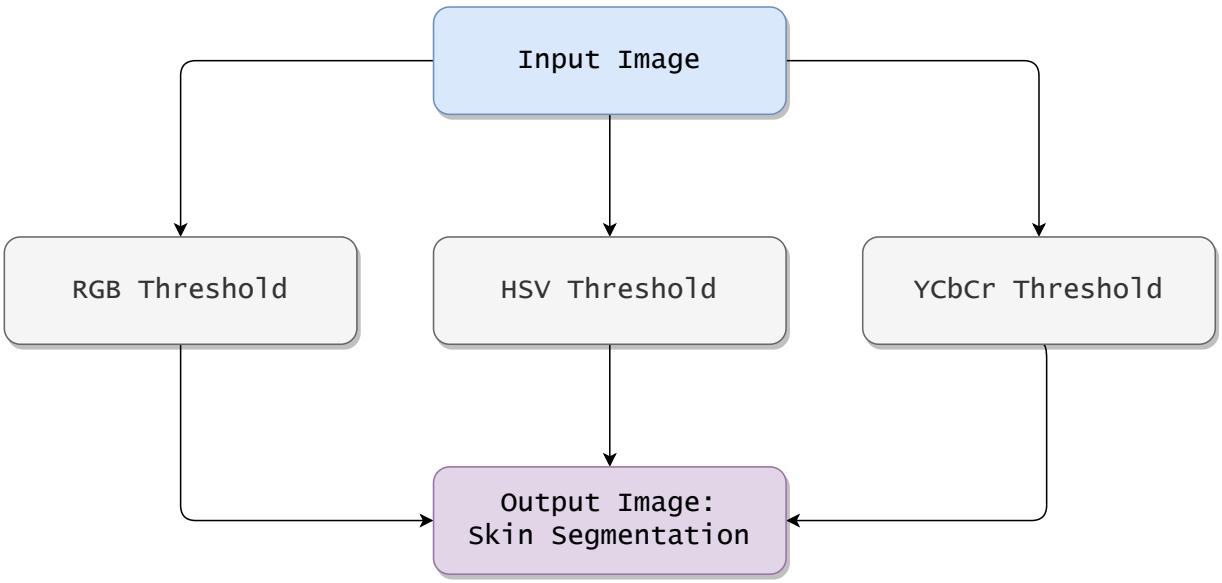


Figure 3.14: Multi-color space threshold

The implementation of this method was based on the article presented in [82]. For this, Equations 3.8, 3.9, and 3.10 were used to select the desired part of each color space.

From the HSV color space, it was taken advantage of the H that corresponds to the Hue:

$$HSV = 0.2 > H > 0.8 \tag{3.8}$$

Regarding the color space YCbCr the equation is as follows:

$$YCbCr = (Cr > 135) \wedge (Cb > 85) \wedge (Y > 80) \wedge Cr \leq (1.5862Cb + 20) \quad (3.9)$$

Finally, to segment the RGB color space with the aim of segmenting the skin, it was used the following equation:

$$RGB = (R > 95) \wedge (G > 40) \wedge (B > 20) \wedge (R > G) \wedge (R > B) \wedge (|R - G| > 15) \quad (3.10)$$

The next step is get the boundaries of the segmented skin. These boundaries will correspond to the region of interest, because tattoos are always on the skin. Figure 3.15 shows the detection of the skin area, delimited by a blue line.

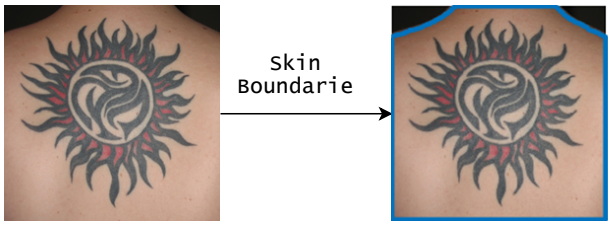


Figure 3.15: Skin Boundarie Detection

Therefore, the objective is to segment the skin to find the **region of interest of the tattoo**, which will be a binary mask of the skin. The region of interest will define the boundaries where the tattoo can be. To obtain a binary mask, the input image is **converted to grayscale**. Grayscale conversion simplifies visual interpretation by removing all color and leaving only shades of gray, which takes away the complexity that color introduces. Grayscale conversion can make it easier to differentiate between various image components in image segmentation and object detection, particularly in the medical industry, for this reason, ultrasound and tomography images are in this format. Furthermore, an image must be passed to just one channel to be normalized to [0,1] to turn into a binary black-and-white image. Thus, the color image to grayscale conversion has been performed. The grayscale image is then converted into a **binary image** (black and white image) using the Otsu method that lowers the intraclass intensity variance, and as a result, can automatically and adaptively generate a threshold that divides the pixels into foreground and background, or in other words, divides the pixels into 0s and 1s. Then, a morphological **closing** is implemented. A closing, in image processing, is a dilation followed by an erosion. Dilation adds pixels to the edges of the image/object. On the other hand, erosion removes pixels from the edges of images/objects.

The addition or removal of pixels depends on the definition of the structuring element.

Subsequently, a **fill** function was implemented. This function aims to fill and seal holes in the binary image. The purpose of this function is to make sure that the mask is filled in white in the intended area. After this part, the binary skin mask is ready. In Figure 3.16 it is possible to observe the skin mask produced.

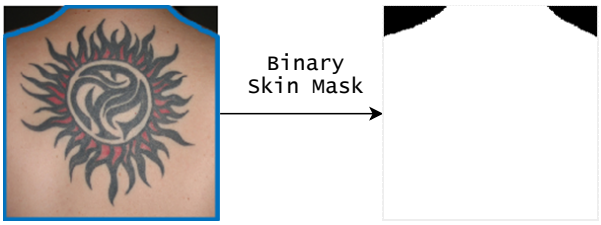


Figure 3.16: Binary Skin Mask

The original image in grayscale is then **concatenated** with the binary mask. This results in the grayscale image but only in the region of interest. Hence, it is removed some noise, and artifacts in the background,as it can be seen in Figure 3.17.

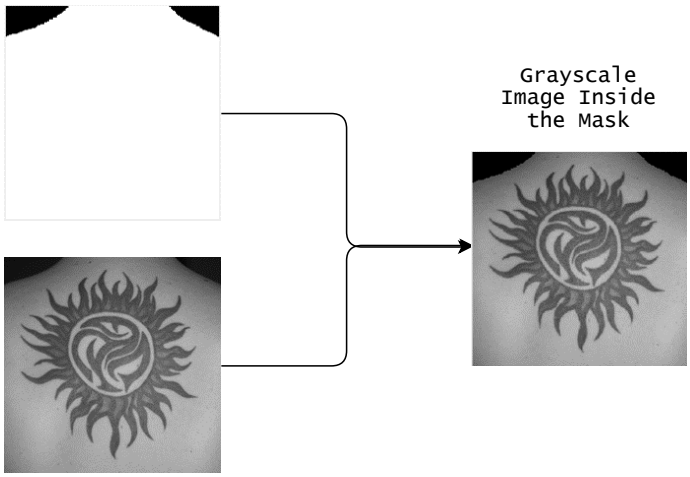


Figure 3.17: Original Grayscale Image Inside the Mask

Afterwards, it is important to balance the histogram of the images. To do that, low light areas were illuminated to **improve image brightness**. Certain images are darker or have more shadows, and improving the brightness in those images is critical to overall segmentation success.

Furthermore, with the purpose of improving the contrast, it was applied a contrast-limited adaptive **histogram equalization** to enhance the contrast of grayscale image. When the image has a limited range of intensity, enhancing contrast is necessary. It is possible to make the image components more distinguishable by using implementing this. As a result, the segmentation of the test set images is improved generally. Figure 3.18 illustrates these two steps.

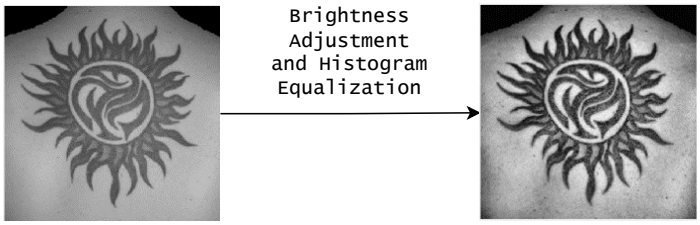


Figure 3.18: Brightness Adjustment and Histogram Equalization

Following that, a **high-sensitivity adaptive threshold** was applied so that the majority of pixels were considered to be in the foreground because the risk of noise was reduced as a result of the preceding steps. Figure 3.19 shows the adaptative threshold.

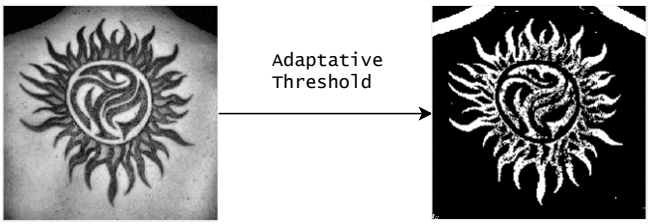


Figure 3.19: Adaptative Threshold

The resultant image is inverted and concatenated with the skin mask to reduce undesired noise. A few morphological operations were performed to further minimize the noise in the image. The morphological techniques enable the removal of several image imperfections, and they are often applied in segmentation techniques. Through the use of a structural element, these operations transform an image and produce an output. First, the small unwanted objects present in the binary image were removed by eliminating any connected components with fewer than 8 pixels. Then, a magnitude and direction gradient was implemented using the **Sobel** method. The Sobel operator accentuates areas of high spatial frequency that correlate to edges by performing a 2-D spatial gradient measurement on an image. It is usually used to determine the roughly absolute gradient magnitude at every location in an input image. Then, the filling function was applied again, because the gradient caused some holes. Lastly, it was employed a function to **suppress any image border-related light structures**. This function helps to clean the edges of the image and thus removes the garbage that remains in the image.

Figure 3.20 shows the final image after some morphological operations to reduce the noise.

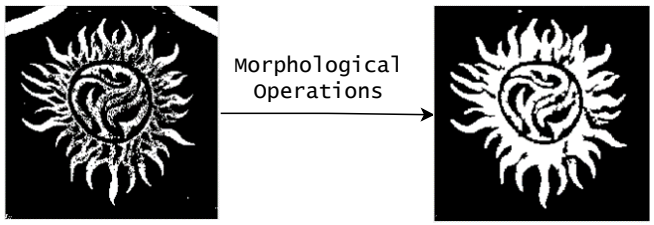


Figure 3.20: Final Image After Morphological Operations

The final segmentation is accomplished after these processes. The sequence of the complete procedure is depicted in the figure below.

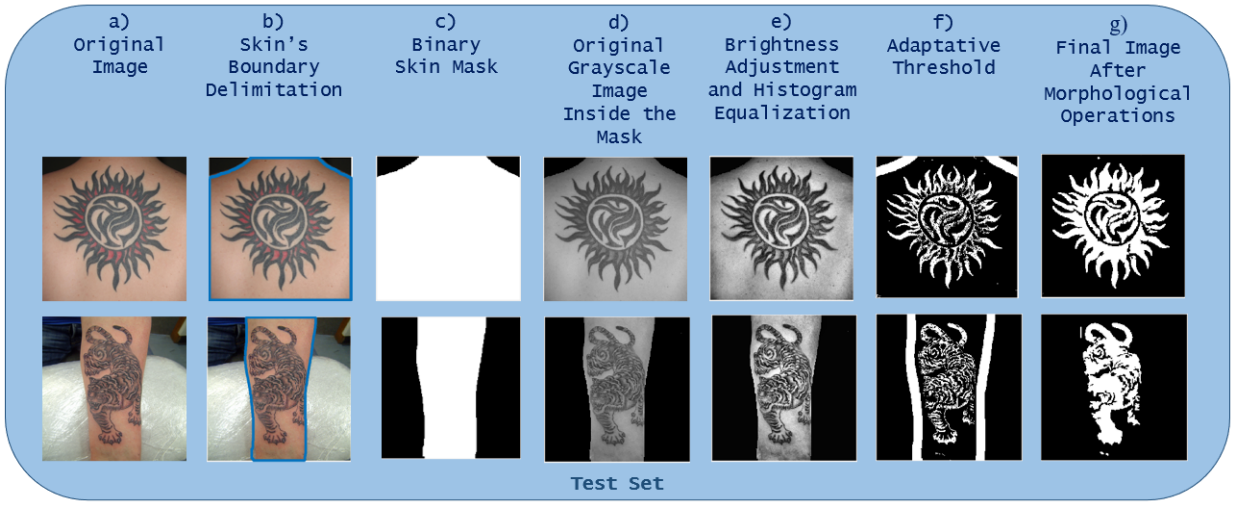


Figure 3.21: Sequence Overview of the Threshold Method

3.3.3 Software Tools

The programming and computing platform **MATLAB, version R2020b** [83], was used to implement conventional methods for segmenting tattoos. This platform was also used to support several other tasks, like file organization or obtaining binary image formats, among others.

Additionally, the software **Anaconda, version 4.11.0** [84], which is an open-source distribution of the Python language for data science that seeks to simplify package management and deployment, was installed to use the Deep Learning methodology. Also, NVIDIA's **CUDA (version 10.2.89**) [85], which is a parallel computing platform and programming paradigm, was also installed. **Python 3.9.7** [86] was the programming language used in conjunction with **Pytorch (version 1.9.0+cu111)** [87], an open source machine learning framework that optimizes the transition from research prototype to production deployment. **JupyterLab** was picked as an interface since it is quite versatile and provides better work organization in data science or machine learning projects. Consequently, it is very useful for coding or

simply serving as a notebook [88].

3.4 Results and Discussion

This section shows the results generated by the different networks, and also by the traditional method. In this way, the graphs of the training and validation related to the deep learning models are presented first, then the visual results of each technique including the traditional method, and finally the quantitative results obtained by each metric applied in all scenarios.

All networks were trained across 1000 epochs, as was already mentioned. Analyzing the training and validation loss graphs is crucial since doing so allows comprehension of how each network behaves. After performing the training, validation, and testing process, no neural network overfitted. The training and loss graphs for each network are displayed in the images below. To visualize metrics such as Loss, and Dice it was used the TensorBoard toolbox from Pytorch, which is used for machine learning exploration. The TensorBoard allows the tracking of the training and validation. Therefore it is possible to analyze how the model is behaving in real-time.

The first neural network implemented was the **U-Net**, and the Figure 3.22 shows the training and the validation loss.

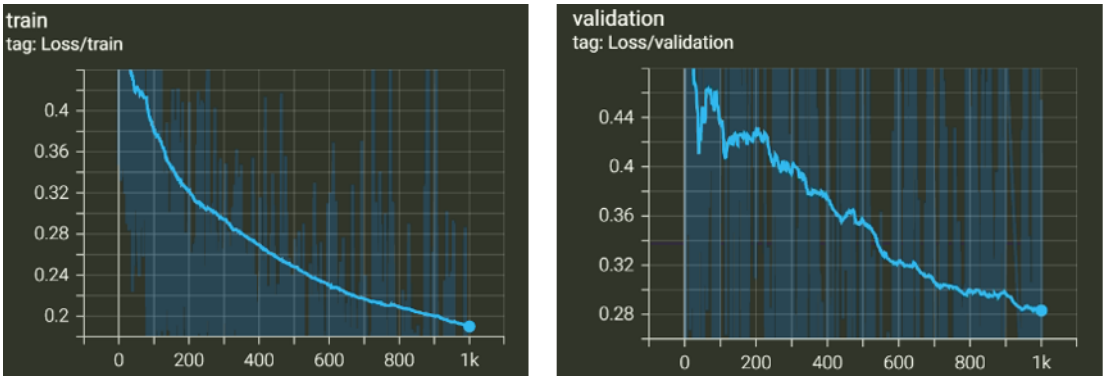


Figure 3.22: U-Net train and validation loss function over the epochs

Then, it was implemented the **DynUNet** as can be seen in the Figure 3.23, that presents the training and validation of this neural network.

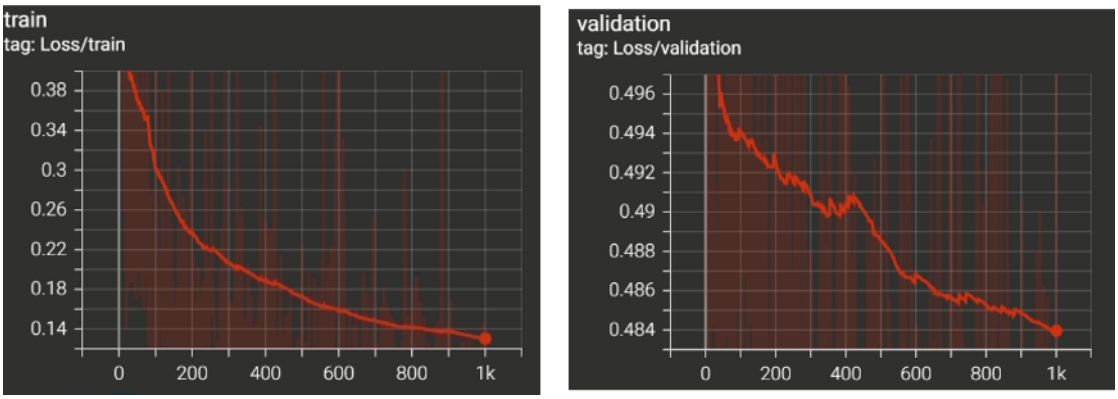


Figure 3.23: DynUNet train and validation loss function over the epochs

After the DynUNet, it was the **UNETR**. The training and validation losses over the epochs are illustrated in Figure 3.24.

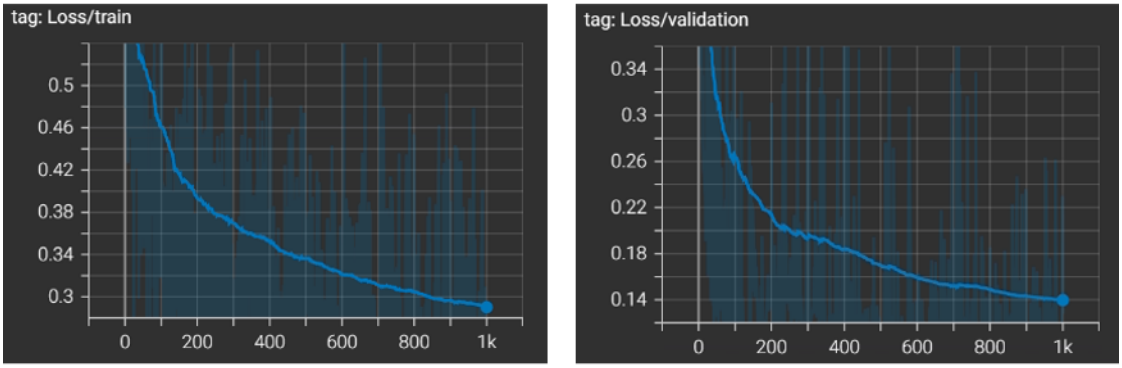


Figure 3.24: UNETR train and validation loss function over the epochs

In addition, the loss graphics of **RegUNet**'s training and validation are depicted in Figure 3.25.

Figure 3.25: RegUNet train and validation loss function over the epochs

The training and validation of the loss function of the neural network **SegResNetVAE**, are depicted in the Figure 3.26.

Figure 3.26: SegResNetVAE train and validation loss function over the epochs

Finally, the training and validation losses of the SegResNet neural network are also presented in Figure 3.27.

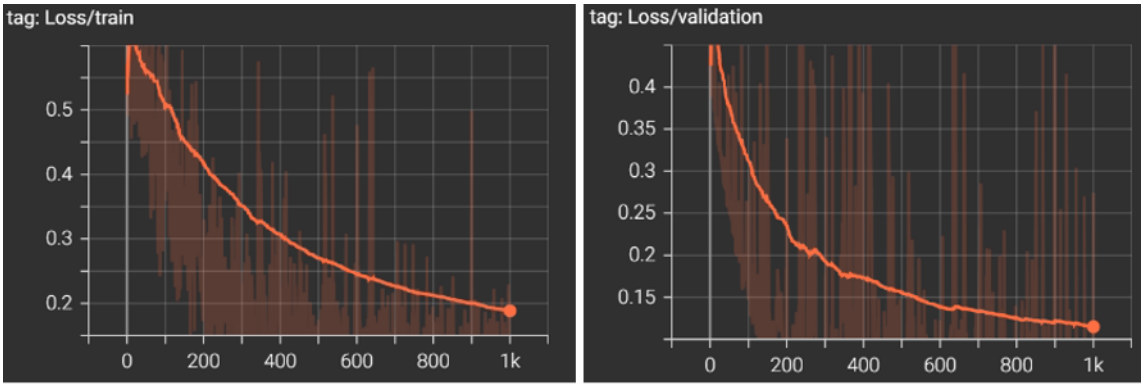


Figure 3.27: SegResNet train and validation loss function over the epochs

After all neural networks have been trained, the model produces output images for the test images that are predictions of the trained models. Here can view some test data images that were predicted using both the threshold approach and deep learning models. The networks with better training results were able to get the best metrics and segment the test images more accurately.

Figure 3.28 illustrates the performance of all approaches and allows for comparisons with both the ground truth image and the original image. RegUNET, and the threshold approach, produces the lowest results, and SegResNetVAE, which produces results that are remarkably close to the truth, is the network that performs best.

Figure 3.28: Tattoo Segmentation Results. Below each image, is presented the corresponding mean Dice value

In terms of deep learning models, U-Net was used as a base model and produced good results. It manages to segment all images reasonably well, with high Specificity and Accuracy values, and a mean Dice of 82.6%. A poor Recall value suggests that the salient image structure was not preserved, which is in accordance with the results, given that the segmentation region is correctly recognized but lacks some detail. With very similar values, DynUNet also presented good results, which was to be foreseen, given that it is a dynamic version of U-Net with minimal structural variations, displaying coincidentally

equivalent Dice and Jaccard index values. However, it has slightly higher Precision and Specificity values than U-Net but lowers Accuracy and Recall values. UNETR, on the other hand, has lower metric values across the board, with a Dice of 74.5% and an Accuracy of 90.2%. When analyzing the results, it is feasible to notice that the images have some noise because the UNETR confuses the background with the foreground, which is understandable given that the UNETR is more designed for 3D segmentation. When it comes to deep learning algorithms, only the RegUNet network performs worse since it can be used for segmentation methods but is more tailored for image registration. These metric values are low for what is intended. The SegResNetVAE network, on the contrary, produces highly satisfactory results, with a Jaccard Index of 80.3%, Dice of 88.5%, Precision of 92.7%, 85.9% Recall, 97.2% Specificity, 95.5% Accuracy, it is the best overall model. This network excelled at segmenting all tattoos; even the most complex images have a high Dice score. Following that, the SegResNet network, which is similar to SegResNetVAE but without the Variational Autoencoder, was utilized and presented lower metric values than this one. Its performance is comparable to U-Net and DynUNet. It is also the most effective network in terms of Specificity. By examining the output of this model, it is possible to notice that the images present the correct segmentation area but without much detail. The threshold technique, as expected, generates the poorest results of all methods since it cannot compete with deep learning models with such a diverse dataset. Its performance was identical to the RegUNet model. However, the threshold procedure has the highest standard deviation of all the methods, which means it can segment some images very well and others poorly. Even though the dataset has been reduced and selected, the images have a high variance, making it challenging to develop an algorithm that performs equally well for all images. Furthermore, many images are either too bright or too dark, while others contain shadows and artifacts. Even with image pre-processing, the method based on more primitive techniques is not as effective as deep learning models. Also, the threshold method identifies the region where there is skin. However, in a few cases, the skin is not correctly identified, so the tattoo within cannot be segmented as it can be seen in the first image of the threshold technique in figure III. It is necessary to reduce dataset variability, improve image brightness and contrast and make the skin region more robust, to improve this technique. When compared to deep learning algorithms, the traditional method does not generate as good results as expected, but in some cases, it can produce a good segmentation. Deep learning approaches have an advantage over threshold methods since they rely on training and validation rather than image characteristics. The SegResNetVAE network was the best of them all, reaching the expectations of what was expected. The SegResNet, U-Net, and DynUNet networks also performed well but fell short of SegResNetVAE, while UNETR performed poorly and RegUNet performed the worst of all. To sum up, the best neural network presented very good results,

with satisfactory performance. However, there is room to improve, especially when the future goal is to use it in a medical context. Even so, some neural networks showed excellent results.

Table 3.4 presents the quantitative results of the different methods.

Table 3.4: Quantitative Comparison of Different Methods

Methods	**Evaluation Metrics**					
	JI	**Dice**	**Prec**	**Rec**	**Spec**	**Acc**
U-Net	72.3± 16.9	82.6± 12.9	87.8± 17.9	82.2± 14.0	96.5± 6.8	93.7± 5.9
DynUNet	72.2± 16.7	82.6± 12.7	92.5± 7.5	77.7± 18.3	97.2± 4.9	93.1± 5.5
UNETR	61.1± 17.0	74.5± 13.6	85.2± 12.0	69.6± 20.6	94.4± 7.9	90.2± 6.1
RegUNet	54.7± 18.9	68.7± 17.2	78.5± 19.8	67.0± 22.1	93.1± 8.0	88.3± 6.6
SegResNetVAE	**80.25± 12.0**	**88.5± 8.0**	92.7± 6.7	**85.9± 12.6**	97.2± 4.9	**95.5± 4.5**
SegResNet	72.0± 17.9	82.3± 13.2	**95.1± 10.7**	75.4± 17.8	**98.0± 4.9**	92.8± 6.5
Traditional	55.5± 21.8	68.3± 21.5	74.8± 25.7	72.5± 21.3	92.7± 10.1	89.2± 7.8

3.5 Conclusion

This chapter compares tattoo segmentation strategies, six deep learning models, and a traditional threshold-based technique. Several metrics were used to assess each approach. The SegResNetVAE produces the best results, with a mean Dice value of 88.5%, surpassing the other methods. Nevertheless, it is possible to construct a more robust dataset with less variability and run numerous tests by adjusting hyperparameter values or choosing different optimizers and loss functions.

Chapter 4: Multispectral Image Influence In Tattoo Segmentation

The impact of infrared images on the segmentation process is investigated in this chapter. To do this, an image acquisition system is created to acquire real-world images of tattooed people, and the images are then studied for the previously mentioned neuronal networks and metrics. Since the previous dataset was with photos from other databases, this chapter also tries to present datasets made under defined settings and with regulated lighting, capturing images that are more identical to the clinical environment. The construction of this dataset then enables images that are closer to the final goal, as well as research into the influence of NIR photographs on tattoo segmentation.

4.1 Introduction

A technique that makes use of the near-infrared part of the electromagnetic spectrum is called NIR spectroscopy. It is possible to emphasize and enhance elements that are not as visible in the visible spectrum by using lighting and cameras designed for the infrared spectrum.

The impact of using multispectral images in the segmentation of tattoos will be examined in this chapter. Thus, this chapter presents the construction of an image acquisition setup, with a monochromatic camera (NIR) and RGB, to create three datasets composed of NIR, RGB, and NIR-RGB data. The 6 neural networks U-Net, DynUNet, UNETR, RegUNet, SegResNetVAE, and SegResNet are implemented for each of the datasets and thus understand how the different types of images behave. Therefore it was done a study on the influence of using multispectral images in the creation of a dataset for tattoo segmentation. An image acquisition setup is built, with near-infrared, white light, and ultraviolet lighting.

The benchmark for assessing each convolutional neural network is the Jaccard Index (JI), Dice (or F-1 Score), Precision (Prec), Recall (Rec), Specificity (Spec), and Accuracy (Acc) metrics.

4.2 Related Work

The NIR spectroscopy technique uses the infrared region close to the electromagnetic spectrum which is around 700 and 2.500 nanometers [89]. When NIR light is emitted whatever absorbs infrared appears

dark, while whatever reflects or transmits infrared appears lighter or translucent [90]. Therefore, it has become increasingly common to use NIR light to enhance a specific target. In medicine, infrared light can help to distinguish blood vessels and in industrial applications can be used to analyze a particular object in a non-destructive way. In this case, infrared light is used to improve tattoo detection. NIR techniques not only allow to visualize a tattoo, but they also permit to see older tattoos covered by newer tattoos. Infrared light can be used to see latent ink in the skin. The longer the wavelength, the deeper the penetration of the skin and hence the better the possibilities of reflecting off low latent ink that is not visible to the naked eye [91].

Figure 4.1 shows the difference of tattoo perception when applied different wavelengths.

Figure 4.1: Tattoo perception when applied different wavelengths [92]

In this example, the NIR light presents better distinction and the tattoo is more easily recognizable.

NIR images are applied to improve segmentation quality since they are less responsive to lighting, resulting in improved tattoo detection. To conclude, tattoos are enhanced using near-infrared images, which helps to improve tattoo detection.

Additionally, by using lights that have the same wavelength as the tattoo's colors, the acquisition setup can enhance the visibility of the tattoo since the ink's colors reflect those lights more effectively. Inks used for tattoos behave similarly to other objects in that they absorb and reflect the light of various wavelengths in different ways. A red tattoo, for instance, will only reflect red light while absorbing all other colors of visible light. Therefore, to eliminate a tattoo the laser has to be of a different wavelength so that the tattoo absorbs the laser and does not reflect it. To visualize it better, that is, for a tattoo of a specific color to be seen better, it has to reflect more, and it reflects more the more identical the wavelength of the emitted light is [93]. Figure 4.2 shows the spectrum of the wavelength of each color.

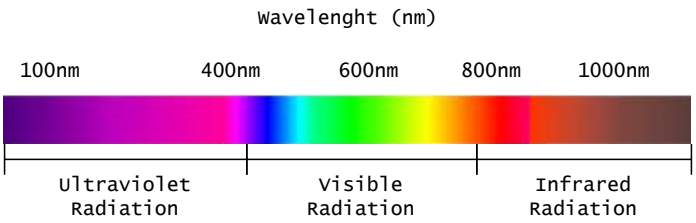


Figure 4.2: Spectrum of the wavelength of each color

So each tattoo pigment has a specific absorption and reflection length depending on the color.

In terms of acquisition setups, a typical image capture setup includes an image acquisition device, such as a camera, connected to a computer via an image acquisition board, and various light sources emitted to the target [94], as can be seen in Figure 4.3.

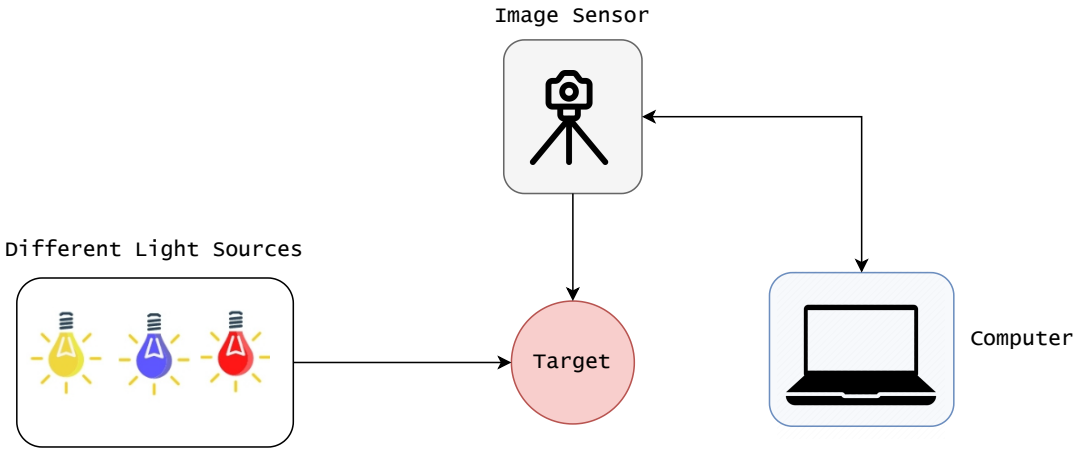


Figure 4.3: Setup for image acquisition process

In addition to the main components that make up an image acquisition setup, an imaging system also has some key aspects, including resolution, working distance, field of view, sensor size, perspective, depth of field, and contrast. Resolution is the smallest element size on the object that the imaging system can differentiate, the working distance is the length from the front of the camera lens to the element under inspection, and the field of view is the area of inspection that the camera can acquire, sensor size is the dimensions of a sensor's active area, perspective is the camera angle with relative to the item under inspection, depth of field is the deepest object that stays in focus, and the contrast represents the disparities in intensity values between the object under inspection and the background [95].

4.3 Dataset Creation

Three datasets have to be produced to study the impact of NIR images: one with only NIR images, one with only RGB images, and one with both types of images combined. However, in order to achieve this, photographs of tattooed individuals had to be acquired, which could only be done by building an image acquisition equipment. This equipment is made by combining several materials, including cameras, lights, and buttons, among others.

4.3.1 Material Specification

First, it was essential to construct an image acquisition setup made up of various components. The first setup component to be defined were the **cameras**. The images were obtained using two cameras, one of which captures monochromatic NIR images and the other RGB images. These particular cameras are:

- Teledyne Dalsa Genie Nano M2020 Mono G3-GM11-M2020
- Teledyne Dalsa Genie Nano C2020 Color G3-GC11-C2020

An illustration of the cameras used is present in figure 4.4.

Figure 4.4: Camera illustration. Two of these were required. One is suited for color images, while the other is for monochromatic images. From the outside, they are identical, but inside they are different

Table 4.1 details each camera's specifications.

Table 4.1: Cameras Specification

Camera Model	Resolution	Spectrum	Interface	Pixel Size	Max Frame Rate	Mass
Genie Nano C2020 Color	2064 x 1544	Color	GigE Vision	3.45 µm	Standard: 38 fps	46g
Genie Nano M2020 Mono	2064 x 1544	Mono	GigE Vision	3.45 µm	Standard: 38 fps	46g

After cameras, **lighting** is a crucial factor as, without the proper light, cameras cannot properly capture the tattoo. Therefore, for **illumination** was used with an adjustable beam power ring, called

EFFI-RING, that emits white light, infrared, and ultraviolet light as well as two lights that project infrared light were utilized to create various lighting configurations. Figure 4.5 shows a picture of the ring light.

Figure 4.5: Illustration of EFFI-RING [96]

The EFFI-RING light's specifications are listed in Table 4.2.

Table 4.2: EFFI-RING specification

Optics	UV wavelength White light wavelength	405 nm 465-625 nm
	IR wavelength	850 nm
Mechanics	Weight	400 g
	Width x length x height	117 mm x 151 mm x 40 mm
	Inside Diameter	58 mm
Eletronics	Connectors	M12 - 5 pins
	Power supply	24 V DC
	Power consumption	72 W
Environment	Working temperature	0°C to 40°C
	IP code	IP65 (except with polarizer accessory : IP50)

The Figure 4.6 shows the infrared, UV, and white light that the ring light emits. Attached to the ring are the two cameras in which it can be seen the lenses of them.

Figure 4.6: EFFI-RING different light modes with cameras attached

In addition to this lighting, another infrared light was also placed, to add variability of illumination. Figure 4.7, shows the NIR light turned off and on. This light has a wavelength of 760 nm.

Figure 4.7: Illustration of the infrared light (760 nm) before and after beign turned on

In addition to the lights and cameras, a computer with a Qt program was also needed to process and capture the images, as well as a board/switchboard that assembled various devices and a tripod to hold the camera.

4.3.2 Acquisition Setup

After gathering all the components, the acquisition setup was ready. The board/switchboard was made outside the scope of this scope, and was adjusted with the cooperation of another colleague for this project. The Qt program's code had already been made outside this project, with the aim of recording videos from three cameras and was modified to handle the photos from the two cameras and capture them at the same time when a button was pressed.

The Figures 4.8 and 4.9 show how the components are connected.

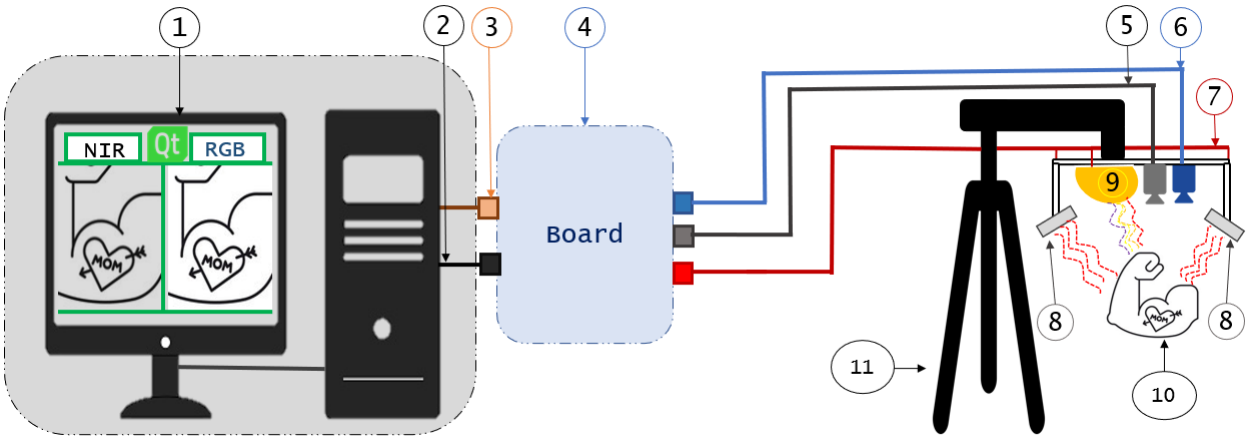


Figure 4.8: Acquisition setup design

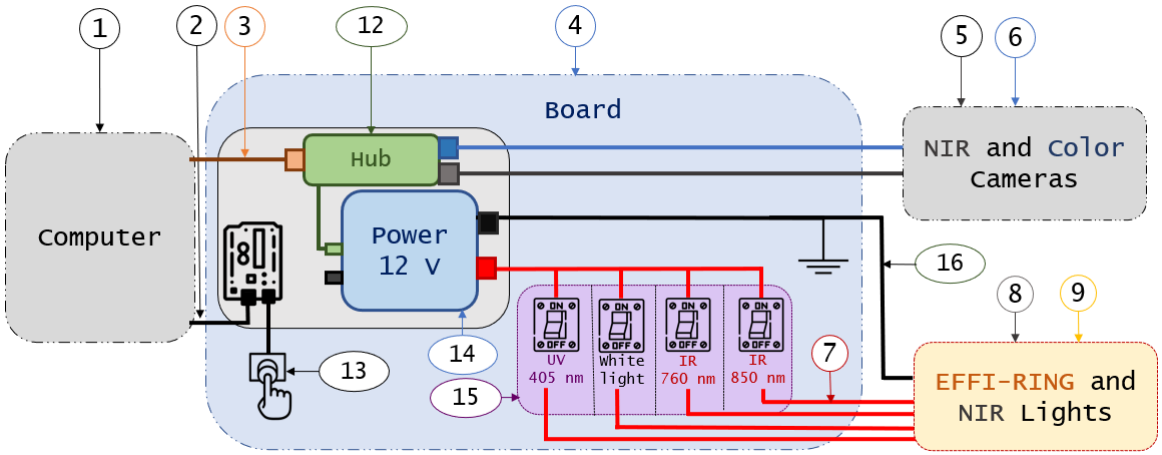


Figure 4.9: Acquisition setup design

- 1 - Qt program that receives and processes the NIR and RGB images provided by the cameras.
- 2 - USB cable that connects an Arduino (which controls the button to capture the images) to the computer.
- 3 - Ethernet cable that connects the hub from the board to the computer.
- 4 - Switchboard that connects the computer and the acquisition tripod (better representation in figure 4.9).
- 5 - NIR acquisition camera Genie Nano M2020 Mono G3-GM11-M2020 (illustrated in figure 4.4) and the corresponding wire.
- 6 - Color acquisition camera Genie Nano C2020 Color G3-GC11-C2020 (illustrated in figure 4.4) and the corresponding wire.
- 7 - Wires that are connected to a respective switch in order to turn on the corresponding light.
- 8 - Near-Infrared lights (illustrated in figure 4.7).
- 9 - Ring light that provides UV light, white light, and IR light (illustrated in figures 4.5 and 4.6).
- 10 - Tattooed person.
- 11 - Tripod holding all components.
- 12 - Hub that broadcasts data (in this case tattoo images) to the computer via Ethernet connection.
- 13 - Button that captures the tattoo images and save them in the computer storage.

- 14 - Power supply of 12 V.
- 15 - Switches that control the lighting environment by turning on and off the lights.
- 16 - Ground (GND) of the lights.

Figures 4.10, and 4.11 show the real structure of the acquisition setup.

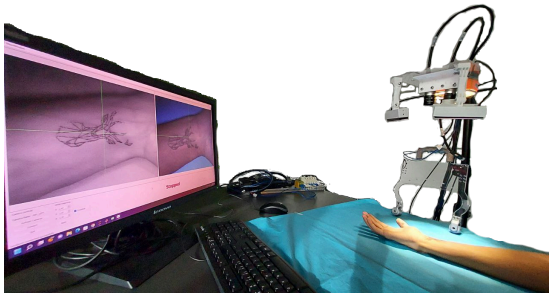


Figure 4.10: Acquisition setup implementation - perspective one

Figure 4.11: Acquisition setup implementation - perspective two

4.3.3 Acquisition Protocol

After the setup is ready and with all the components connected, it was moved to a private space so the dataset participants feel comfortable. People with tattoos were needed to construct the dataset, thus individuals having tattoos or willing to obtain temporary tattoos were asked to participate in the tattoo acquisition.

Each tattoo was captured in two different positions by each light combination. 8 lighting combinations were used. Therefore, 16 images were captured for each tattoo, and there was variability in lighting and position/angle. The light combination protocol is described in Table 4.3

Table 4.3: Light combination protocol

Capture	White Light	UV	IR (760 nm)	IR (850 nm)
1st	X			
2nd	X	X		
3rd			X	
4th			X	X
5th				X
6th	X		X	X
7th		X	X	X
8th	X	X	X	X

By pressing the button, the images that each camera sees are captured and saved in the corresponding folder. The capture from the NIR camera is saved in the NIR folder, and the RGB camera's capture is saved

in the RGB/color folder. Both images are saved at the same time. The stages of picture acquisition are shown in Figure 4.12.

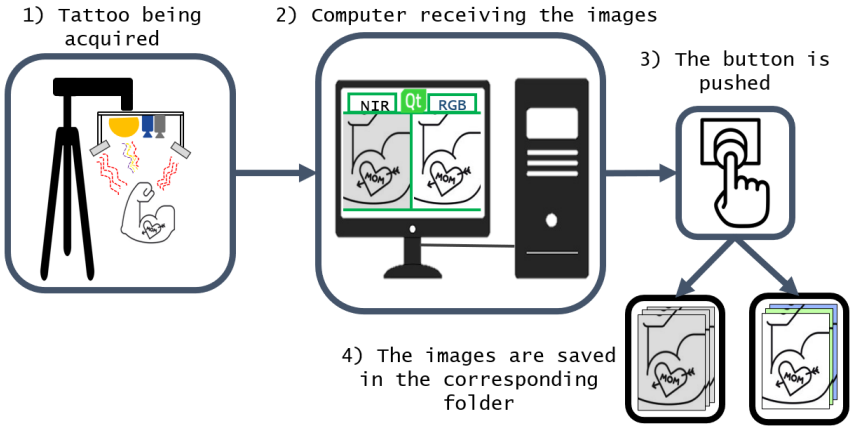


Figure 4.12: Process of saving the images

The process of saving images to the computer is depicted in Figure 4.12. The computer is connected to a button that, when pressed, starts the image recording process. The computer receives the images from the tripod holding the cameras. Images are saved in the path defined by the Qt software.

A total of 1109 images of tattooed people were acquired for each camera. Later, the images were selected and the respective ground truth was performed.

4.3.4 Ground Truth Generation

To generate the ground truth, 4 simple steps were implemented:

- **Select the appropriate images:** RGB and monochromatic images were selected to generate the three datasets mentioned earlier. In other words, photos that didn't fit the intended dataset were taken out. For instance, pictures taken with visible light but without infrared light are unwanted for the NIR image dataset because the NIR camera can't "see" them. The same thing occurs from the opposite perspective; since the color camera cannot capture an accurate image when only infrared lights are present, these images should not be included in the dataset of color camera images.

- **Create the bounding boxes manually:** Once only the appropriate images are selected for each dataset, a manual bounding box was made, setting and painting the area where the tattoo is in white and leaving the rest of the image (irrelevant part) in black, turning the image into a binary image. This process was done both on the dataset of images taken in visible light and NIR light.

- **Application of an adaptive threshold:** Automatically, an adaptive threshold was applied to each image, converting them to binary format, where anything below a predetermined value appeared as 0's or black and anything above appeared as 1's, in this case in white.

- **Implementation of an AND Logic between the bounding boxes and the thresholded images:** After creating the bounding boxes manually, and applying an adaptive threshold to the images, it is possible to obtain the ground truth by implementing an "And logic" between them, which means that the two binary images will form the segmentation since the pixels that are at logical level 0 (black pixel) will remain at 0 if the correspondents of the other image are also at 0. The same happens for logical level 1 (white pixel), which is only maintained if the pixel of the bounding box image and the image that was applied to the threshold is at 1. When there is a pixel at logical level 0 (black pixel) and a pixel at logic level 1 (white pixel), the output pixel will be black - logic level 0. In this way, it is possible to eliminate noise from the segmented image. Figure 4.13 can illustrate this process:

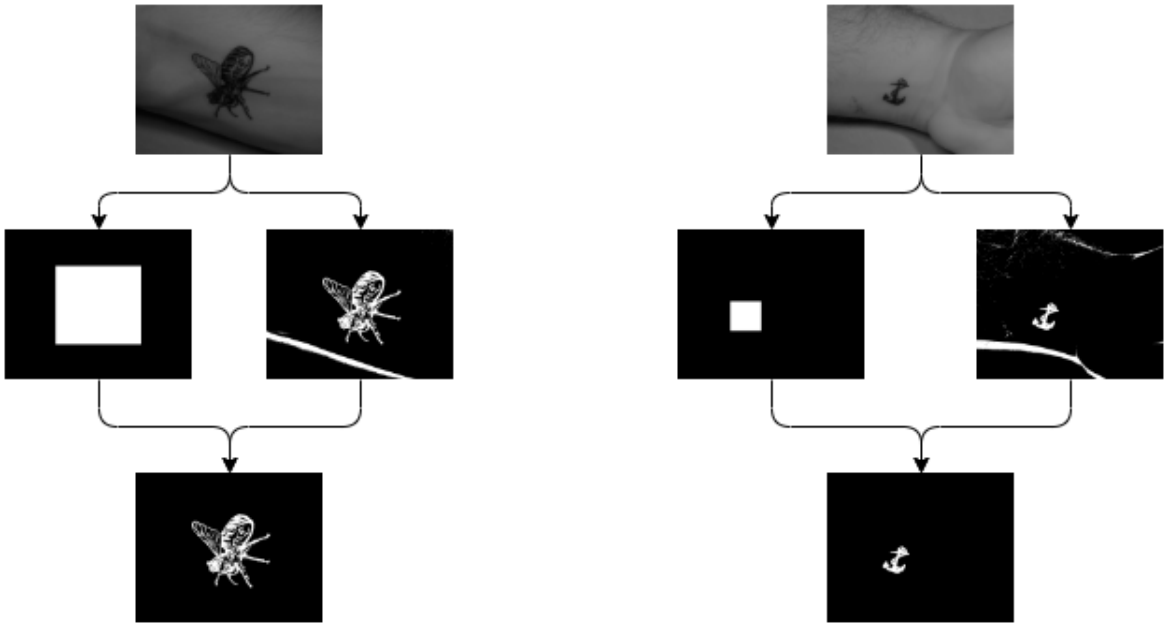


Figure 4.13: Ground truth creating process

Even though the procedure is not as precise and accurate as the manual segmentation carried out in the preceding chapter, it is undoubtedly faster and meets the needs of obtaining the ground truth.

4.3.5 Final Datasets

After obtaining the captures' ground truth, an image division procedure is used to create a dataset that is divided into training, validation, and test sets.

The correct captures were used for the NIR photos; for instance, the first and second captures in table 4.3 are not advantageous because no light-emitting in the infrared spectrum is on, therefore they were not chosen. Additionally, because the camera's focus is fixed and tattooed people occasionally don't have the best focal point, some pictures might not be in focus. Consequently, the division was performed as depicted in figure 4.14.

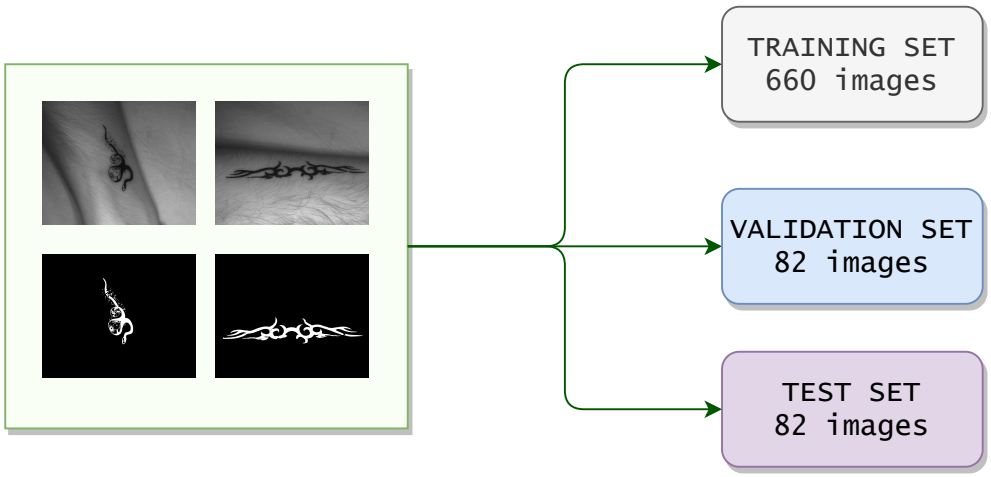


Figure 4.14: Splitting monochromatic images into training, validation, and test sets

The same process was used for the RGB images, that is, only images whose capture lighting was adequate were selected, and the image was not blurred, which resulted in 580 training images, 73 images for the validation folder, and 72 for the test set, as can be seen in figure 4.15.

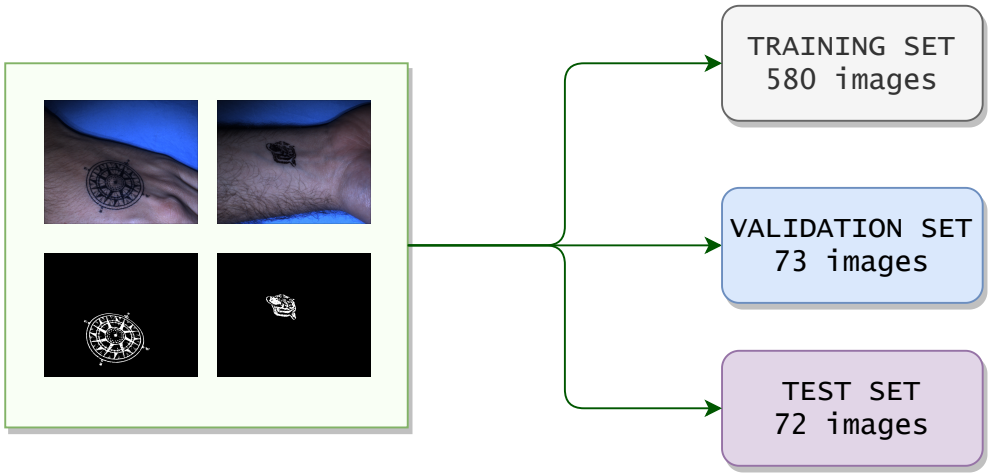


Figure 4.15: Splitting RGB images into training, validation, and test sets

Finally, a dataset containing RGB and monochromatic (NIR) images was organized. For this, the sixth capture was used whose illumination is composed of white light, and two lights in the infrared spectrum with wavelengths 760 nm and 850 nm. In this way, this dataset was short. The training set is composed of 115 monochromatic images that match 115 RGB images, which forms 115 images resulting from the concatenation of the two formats. The validation set is formed by 14 concatenated NIR-RGB images, as well as the test set. Figure 4.16 illustrates this division of the dataset.

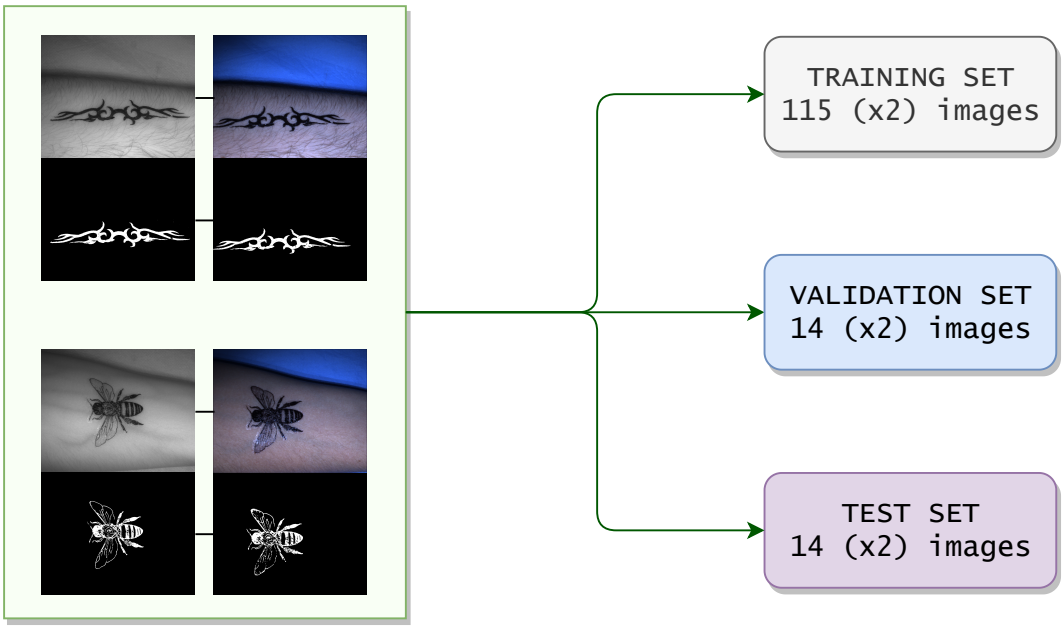


Figure 4.16: Splitting NIR-RGB images into training, validation, and test sets

4.4 Experimental Procedure

To evaluate these created datasets, new evaluations were made on the neural networks and the same hyperparameters. Therefore, a new experimental test was carried out.

4.4.1 Hyperparameters and Neural Networks

The implemented neural networks were the same as in the previous chapter, and are listed in Table 3.2.

For comparative analysis, the described networks ran on the same **hyperparameters** mentioned in the previous chapter, namely:

- Number of epochs: 1000

- Batch Size: 6
- Optimizer: Adam
- Learning Rate: 1e-5
- Weight Decay: 1e-5

The study was approached using the same **loss function** - Dice loss (more specifically, the MONAI DiceLoss function). To study the influence of multispectral light, two experiments were carried out:

- A comparison of the outcomes using the standard procedure for each spectrum under study, which entails rerunning the neural networks with the same code and distinct datasets. first with the dataset of color images, then with the dataset of monochromatic images.
- A comparison of the outcomes utilizing the dataset created by concatenating the two image types, which consist of NIR-RGB pictures.

The code has to be modified to concatenate the RGB photos and the monochromatic images for the dataset that mixes RGB-NIR images. The RGB and NIR pictures, as well as their masks, are required to match for it to happen. In addition, careful consideration had to be given to how the data augmentation was carried out to ensure that the images, which are not the same since they came from different cameras, would maintain their similarity as much as possible. For this, a concatenation was made in the second channel that corresponds to the dimension responsible for the color. In RGB images, in this dimension, there are color channels (Red, Green, Blue): 3, while in monochrome images (NIR) there is only one color since it is a grayscale and therefore only has 1 channel. By concatenating this dimension, the resulting image presents this dimension with a value of 4, as shown in Figure 4.17.

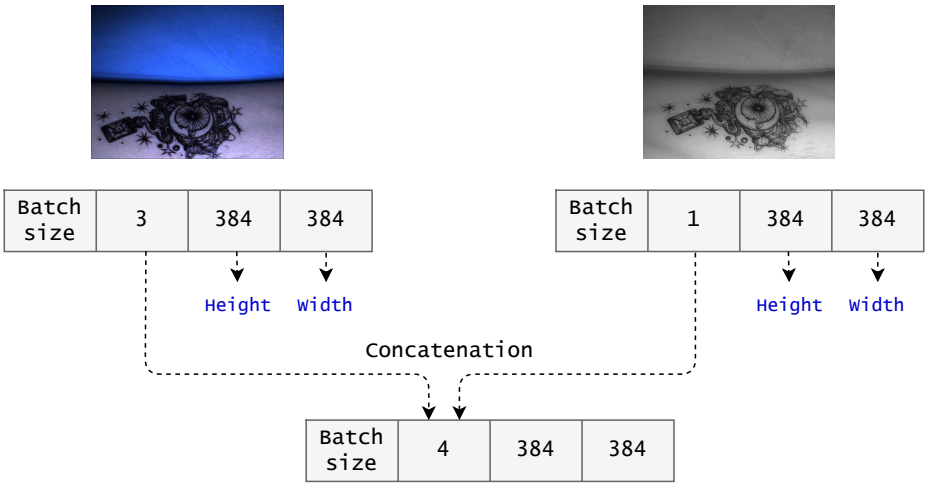


Figure 4.17: Concatenation in color channel dimension for the normal images

This process was also done in masks, as can be seen in Figure 4.18, during the training, validation, and testing sets.

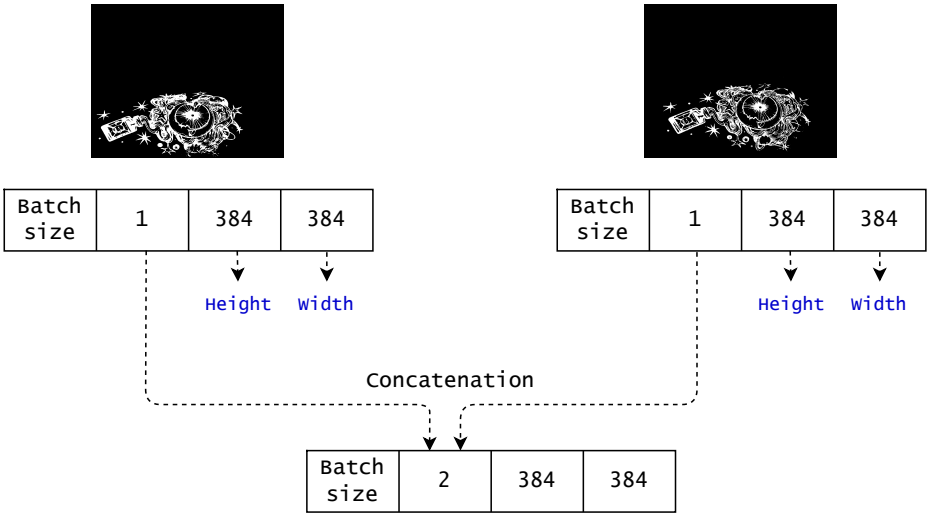


Figure 4.18: Concatenation in color channel dimension for the masks

4.4.2 Metrics

To evaluate the neural networks, the following metrics were used: **Jaccard Index** (JI), **Dice** (or F-1 Score), **Precision** (Prec), **Recall** (Rec), **Specificity** (Spec), and **Accuracy** (Acc). These metrics are explained in the section 3.2.3.

4.5 Results and Discussion

After all the neural networks run for the three datasets, the results were obtained, using the mentioned metrics. The study of the influence of NIR images on the segmentation of tattoos started exactly with the dataset that corresponds to the monochromatic images. In Figure 4.19 it is possible to observe the segmentation of neural networks.

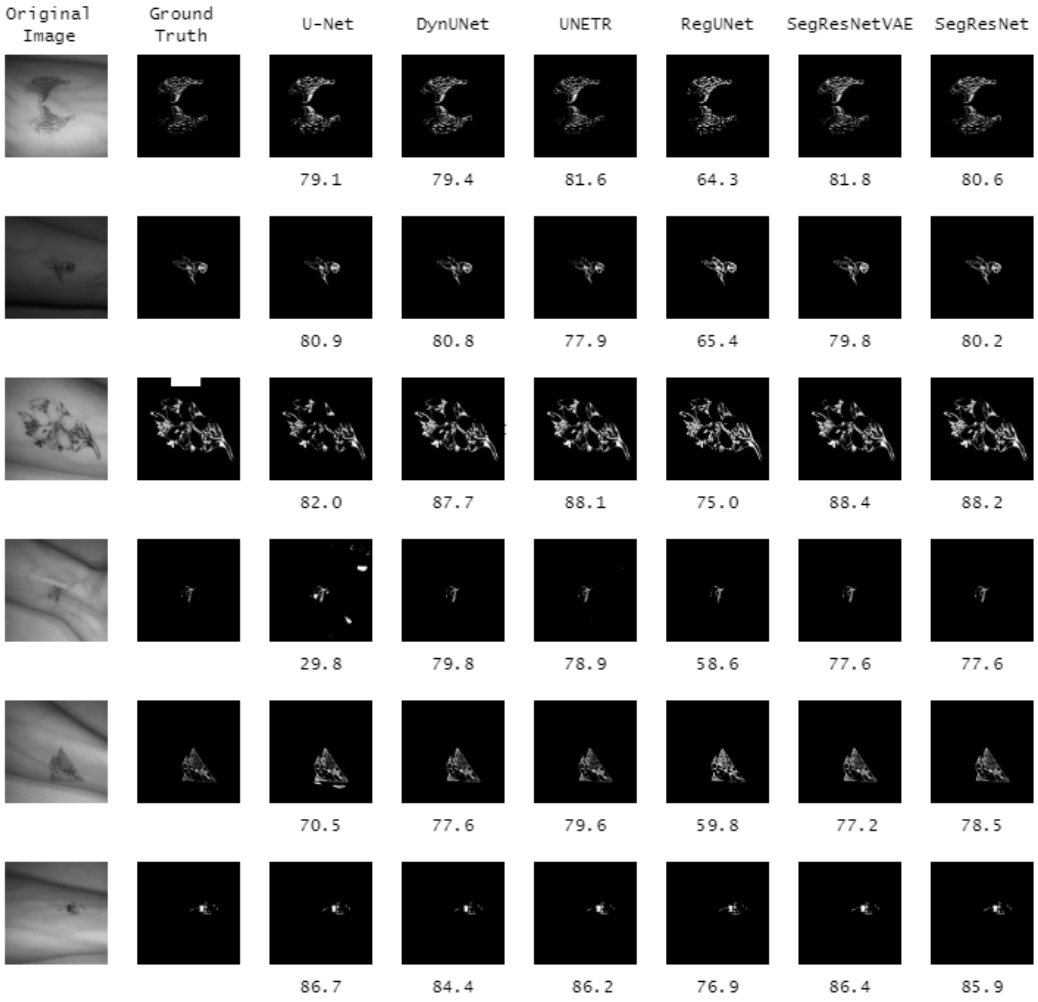


Figure 4.19: Tattoo segmentation results of the NIR dataset. Below each image, is presented the mean Dice value

Initially, the dataset containing monochromatic images in the NIR spectrum was used. It started with the U-Net convolutional neural network, which presented 67.8% of the Jaccard Index, with a standard deviation of 4.9, with Dice 79.9%, Precision 85.2%, Recall 75.4%, and Specificity 99.8%, Accuracy 99.4%. These results were reasonably good, however, there were other networks such as DynUNet which obtained 69.7% of Jaccard Index, 81.9% of Dice, 86.7% of Precision, 78.1% of Recall, and Specificity of 99.8% and 99.5% of Accuracy. This network also segmented the test images relatively well. Similarly, the DynUNet

network has the UNETR network, which also segmented satisfactorily, with 68.8% of the Jaccard Index, 81.2% of Dice, Precision of 84.6%, and Recall of 78.7% (best Recall value obtained comparing with the other networks), 99.7% of Specificity e 99.6% of Accuracy (also the best value when compared with the other networks). Contrary to the other networks, RegUNet presented a worse performance, with 50.1% of Jaccard Index, 66.4% of Dice and 75.8% of Precision, 59.8% of Recall, 99.6% Specificity, 98.9% of Accuracy. In addition to these, the SegResNetVAE neural network obtained a Jaccard Index value of 69.7%, Dice of 81.9%, Precision of 89.9%, Recall of 75.5%, Specificity of 99.8%, and Accuracy of 99.4%. Finally, the neural network that presented the best performance was SegResNet, where it obtained the best values in the metrics Jaccard Index, Dice, Precision, and Specificity, 70.2%, 82.3%, 90.3%, and 99.8%, respectively. In conclusion, the worst neural network was RegUNet, which presents worse segmentations in figure 4.19, and lower values in the metrics of table 4.4. On the other hand, SegResNet achieved a good performance, as shown in Figure 4.19 and Table 4.4.

Table 4.4: Quantitative Comparison of Different Networks for the NIR Dataset

Networks	**Evaluation Metrics**					
	JI	**Dice**	**Prec**	**Rec**	**Spec**	**Acc**
U-Net	67.8 ± 4.9	79.9 ± 3.5	85.2 ± 3.6	75.4 ± 4.8	99.8 ± 0.2	99.4 ± 0.6
DynUNet	69.7 ± 6.7	81.9 ± 4.7	86.7 ± 4.6	78.1 ± 7.9	99.8 ± 0.2	99.5 ± 0.5
UNETR	68.8 ± 7.8	81.2 ± 5.6	84.6 ± 5.4	**78.7 ± 8.6**	99.7 ± 0.2	**99.6 ± 0.5**
RegUNet	50.1 ± 7.8	66.4 ± 6.9	75.8 ± 5.7	59.8 ± 9.9	99.6 ± 0.3	98.9 ± 1.0
SegResNetVAE	69.7 ± 7.4	81.9 ± 5.3	89.9 ± 3.5	75.7 ± 8.5	99.8 ± 0.2	99.4 ± 0.4
SegResNet	**70.2 ± 6.7**	**82.3 ± 4.6**	**90.3 ± 2.9**	75.8 ± 7.2	**99.8 ± 0.1**	99.4 ± 0.4

Next, the neural networks were evaluated for the RGB dataset. Like the monochromatic dataset, this dataset also achieved good results in general. The Figure 4.20 shows the segmentation of networks for different test images.

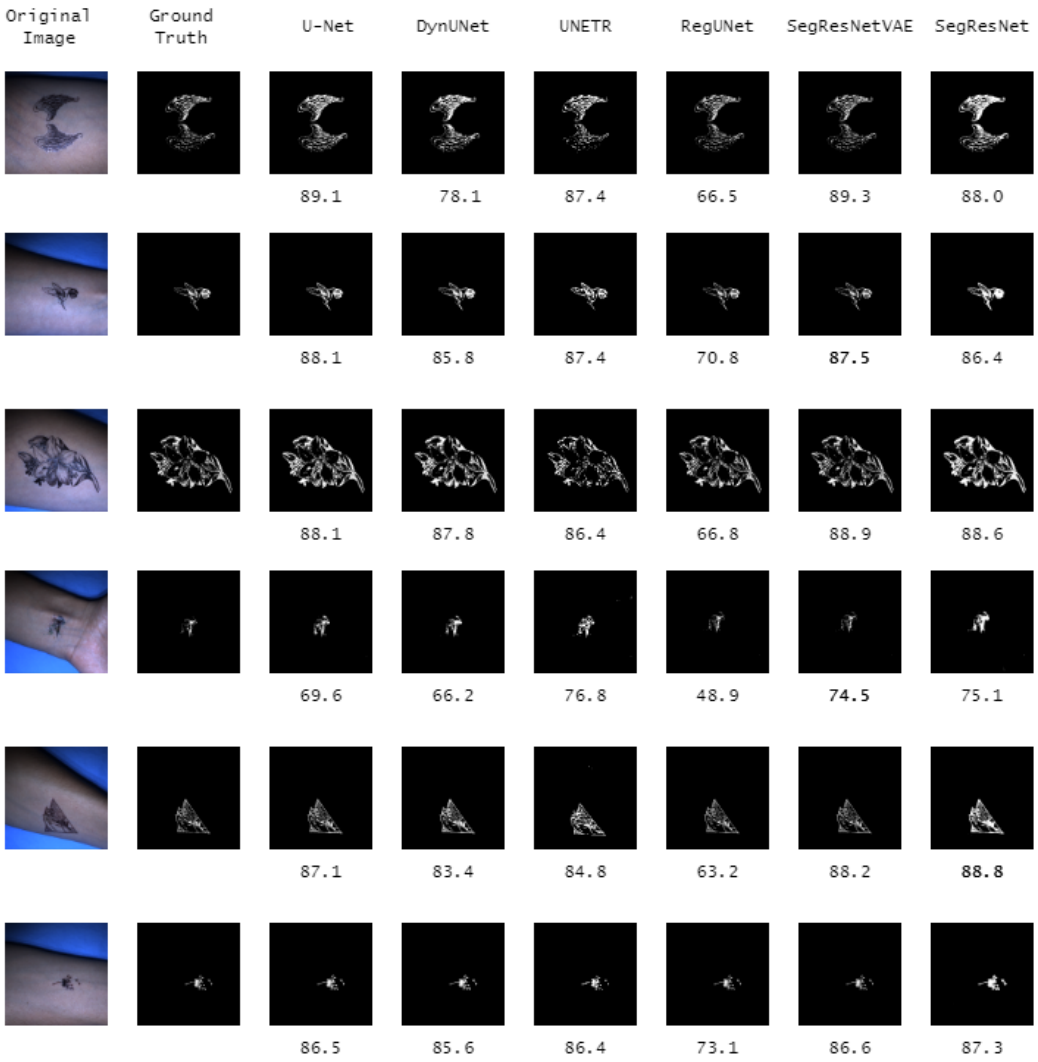


Figure 4.20: Tattoo segmentation results of the RGB dataset. Below each image, is presented the mean Dice value

The dataset with RGB images was also evaluated with the networks already mentioned. The first neural network was the U-Net, which obtained satisfactory results, presenting a Dice of 83.4%, a Jaccard Index of 72.3%, and with high values of Specificity and Accuracy. Next, the DynUNet neural network was tested, which despite having a lower Dice value (80.1%) compared to the U-Net, obtained the highest Precision and Specificity values, 96.3% and 99.9% precisely, of all the networks for this dataset. Next, UNETR was evaluated, which reached a Dice of 85.3%, which is quite good since it was the third-best network in this metric, and was the best in terms of Recall, obtaining 87.5%, with a standard deviation of 7.5. This was followed by RegUNet, which fell short of expectations, obtaining 44.9% and 60.8% in the Jaccard Index and Dice metrics, respectively. After RegUNet, the performance of the SegResNetVAE neural network was studied, which was the best segmented in the initial dataset, from chapter 3. This neural network obtained the highest values for the metrics Jaccard Index and Dice, with 77.0% and 86.8% respectively.

In the other metrics, it presented equally high values, which can be seen in Table 4.5. Finally, SegResNet had a satisfactory performance, with 86.2% of Dice and with the highest Accuracy of all neural networks for this dataset with 99.6%.

Table 4.5: Quantitative Comparison of Different Networks for the RGB Dataset

Networks	**Evaluation Metrics**					
	JI	**Dice**	**Prec**	**Rec**	**Spec**	**Acc**
U-Net	72.3 ± 7.8	83.4 ± 5.6	80.7 ± 5.4	87.8 ± 8.6	99.7 ± 0.2	99.5 ± 0.4
DynUNet	67.6 ± 10.8	80.1 ± 8.1	**96.3 ± 4.2**	69.4 ± 11.9	**99.9 ± 0.1**	99.3 ± 0.6
UNETR	74.8 ± 6.1	85.3 ± 4.1	83.9 ± 7.3	**87.8 ± 7.5**	99.7 ± 0.3	99.5 ± 0.4
RegUNet	44.9 ± 12.1	60.8 ± 13.2	74.8± 10.4	54.6 ± 17.6	99.4 ± 0.8	98.6 ± 1.0
SegResNetVAE	**77.0 ± 6.4**	**86.8 ± 4.2**	88.5 ± 3.9	85.6 ± 7.1	99.8 ± 0.2	99.5 ± 0.7
SegResNet	76.0 ± 6.3	86.2 ± 4.2	87.6 ± 4.8	85.5 ± 8.1	99.8 ± 0.1	**99.6 ± 0.1**

The results for dataset that contains only NIR images did not have better results than the dataset that contains only RGB images, nor that the dataset initially made with random images, which was not to be expected. One reason that could explain this, is the fact that the creation of ground truth for these last datasets was done semi-automatically and not manually as it was done before. The manual segmentation for creating the ground truth is much more perfect since each pixel is evaluated individually and with more care in defining whether it is a pixel of interest (tattoo) or an irrelevant pixel (background), but the creation of ground-truth of this dataset was done semi-automatically and therefore some pixels may have been poorly defined since most of the tattoos are not real tattoos but temporary and reflect much more light, as you can see in Figure 4.21.

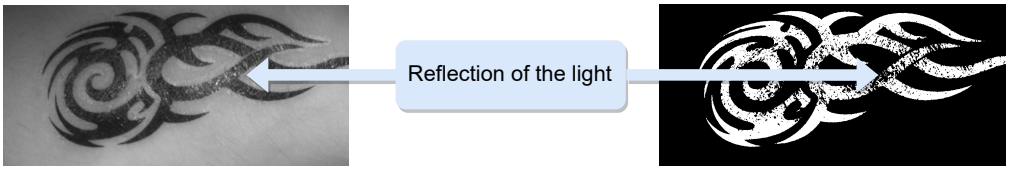


Figure 4.21: Reflection of the light

In a few cases, it happened that the segmentation in the creation of the ground truth was not so perfect, because a temporary tattoo, as it is inserted in the skin with water, has much more reflected brightness, than a real tattoo. As ground truth is not so accurate, both the training of the neural networks and the final evaluation made through the output predicted with the ground truth is not the most correct,

which could indicate that the dataset with NIR images could have better results, if the segmentation was manual.

The dataset made up of NIR-RGB pictures was also assessed. The outcomes, however, did not match expectations. There are a few things that might have affected this outcome, such the cameras' imperfect alignment despite pointing at the same target, which could have affected the concatenation process, and the hyperparameters' lack of adjustment for the dataset's size. With a few samples from the test set, it is feasible to see the segmentation that all neural networks performed in figure 4.22.

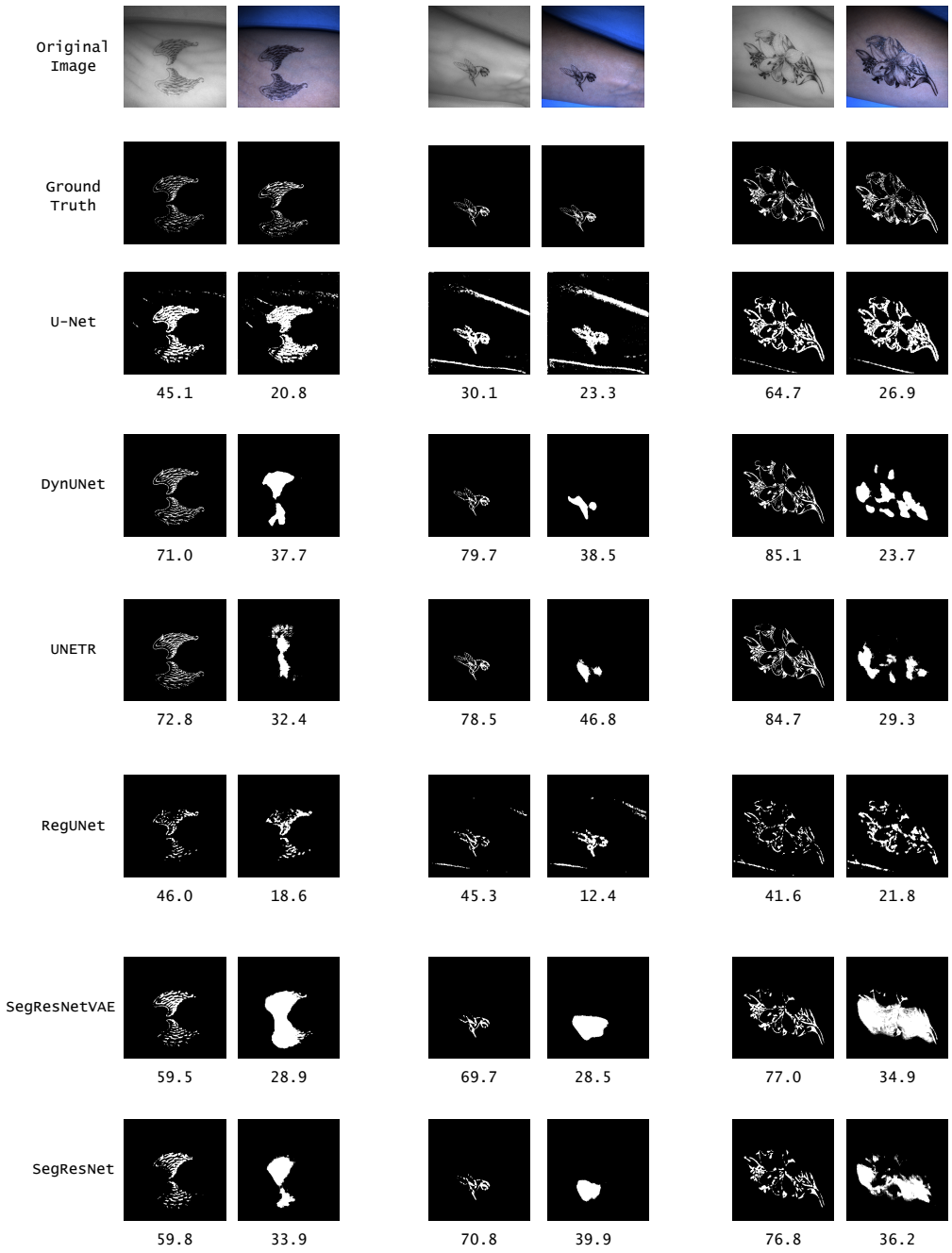


Figure 4.22: Tattoo segmentation results of the RGB and NIR concatenated dataset. Below each image, is presented the corresponding mean Dice value

Finally, the dataset including NIR images and RGB images that have been combined, was used to analyze its performance. However, due to the short dataset, the results were not satisfactory. Techniques like data augmentation were used to address this dataset issue, but they were insufficient because this dataset was much smaller than the others. The RGB images and the monochromatic images were combined, or two images were combined into one, as well as their masks, producing a final result that was then split into two parts: the result/segmentation of the NIR image and the result/segmentation of the RGB image. In Figure 4.22, it is possible to observe that all the networks found it easier to segment the part related to the monochromatic image. The neural network U-Net only had a Dice of 17.75% and a Jaccard Index of 10.9%. In Figure 4.22, it is possible to observe that the tattoos are noticeable in the result of the segmentation of this network, but contain some noise. DynUNet had the best results of all networks for this dataset with 57.5% of Dice and 43.2% of Jaccard Index. The result of the NIR image part had a reasonable segmentation, however, the RGB image part could not segment correctly. Nevertheless, it manages to identify the region where the tattoo is located, despite the little detail. The UNETR neural network had similar results to DynUNet, with 53.7% of Dice and 40.4% of Jaccard Index, and presented a similar segmentation considering Figure 4.22. RegUNet once again performed poorly compared to other neural networks. Analyzing the results, it performed very similarly to the U-Net neural network. The Dice and Jaccard Index values were 22.8% and 13.8%, respectively. Finally, the neural networks SegResNetVAE and SegResNet presented very similar performances in the segmentation of these images. SegResNetVAE got a Dice of 45.5% while SegResNet got 45.3%, in terms of the Jaccard Index, neural network SegResNetVAE got 31.6% and SegResNet got 30.9%. Table 4.6 presents the quantitative comparison of different networks for the NIR-RGB dataset.

Table 4.6: Quantitative Comparison of Different Networks for the NIR-RGB Dataset

Networks	**Evaluation Metrics**					
	JI	**Dice**	**Prec**	**Rec**	**Spec**	**Acc**
U-Net	10.9 ± 10.2	17.7 ± 14.6	63.9 ± 8.6	11.7 ± 11.5	99.3 ± 0.8	91.2 ± 1.0
DynUNet	43.2 ± 7.3	57.5 ± 7.3	74.9 ± 4.2	48.2 ± 9.4	99.4 ± 0.8	97.9 ± 1.8
UNETR	40.4 ± 8.2	53.7 ± 8.5	73.6 ± 10.7	73.6 ± 9.5	99.3 ± 0.9	97.9 ± 1.7
RegUNet	13.8 ± 6.3	22.8 ± 9.5	38.6 ± 14.2	19.9 ± 10.9	98.4 ± 1.5	96.4 ± 2.3
SegResNetVAE	31.6 ± 7.7	45.5 ± 8.3	77.9 ± 14.4	36.8 ± 10.5	99.4 ± 0.8	96.7 ± 2.4
SegResNet	30.9 ± 8.2	45.3 ± 7.8	78.6 ± 12.1	34.6 ± 11.1	99.4 ± 0.8	97.1 ± 2.1

Analyzing the three datasets presented in this chapter, it can be noted that they are not large datasets and therefore, the addition of more data could greatly improve all datasets, especially the last one which is quite short and combines two different types of images. Both monochromatic and RGB images show good results in the first and second datasets of this chapter, respectively. The acquisition of these images allowed the creation of datasets with satisfactory results and with a low standard deviation, which means that the results are generally good. Regarding the concatenation of NIR and RGB images, the results were not so good and the output segmentation was not as satisfactory as the others, which can be due to the dataset being short or the incorrect alignment of the two images. However, it can be seen from the result of this dataset that it may be easier to learn to segment NIR images than RGB since in their concatenation, the NIR part is better segmented than the RGB part, which was the study what it was intended to do.

Comparing the results of this chapter to the results of the previous chapter, it was expected that these results would be better because they were carried out in a controlled environment, however, this is not the case. A difference that may have been influential for the results not being better was the fact that the resize of the images are 384 by 348. This value was chosen to take into account the images of the databases found in the construction of the first dataset. In other words, the image dimensions used were the same for both datasets since it was intended to analyze the performance according to the same parameters. However, the second dataset's images were captured with much larger dimensions. Furthermore, the manual segmentation of the ground truth in the first dataset was more thorough than the second one, which was semi-automated and therefore less accurate, not only in the training but also in the performance evaluation.

4.6 Conclusion

Finally, it was investigated how NIR pictures affected the segmentation of tattoos. To do this, an image acquisition system was built, enabling the photography of tattooed individuals (with temporary and some real tattoos). This picture acquisition framework was made to be flexible and used by other applications, which is quite valuable. With this setup, three datasets were produced: one with monochromatic images, one with RGB images, and one with both types. Six neural networks—U-Net, DynUNet, UNETR, RegUNet, SegResNetVAE, and SegResNet—were then put into use to test their performance. Networks performed well with datasets that only include images in a single format. The SegResNet network performed best on the NIR dataset, scoring 82.3% of Dice, while the SegResNetVAE neural network performed best on the RGB dataset, scoring 86.8% of Dice. Although in this context the RGB dataset obtained slightly better

results than the NIR dataset, when the two formats are combined, neural networks find it easier to segment NIR images, which may indicate some advantage in this format. This dataset presented weaker results, the fact that the dataset is smaller and that the cameras are not perfectly aligned may have prejudiced its performance. Other factors such as obtaining the ground truth semi-automatically and captures being mostly in temporary tattoos, which reflect more light when compared to real tattoos, may have worsened the learning of networks and also the quantitative analysis of the metrics used.

However, all these aspects enriched the project, not only due to the construction of the setup but also due to the addition of more images (through the creation of datasets) and through the analysis of the influence of NIR images.

Chapter 5: Final Remarks

The overall project's final conclusions are presented in this chapter, as well as the possible future work to continue this project.

5.1 Conclusion

This work proposed the segmentation of tattoos using Deep Learning techniques. Six neural networks were used, namely U-Net, DynUNet, UNETR, RegUNet, SegResNetVAE, and SegResNet. To train these networks it was necessary to build a dataset from other datasets found on the Internet, such as Flickr, WebTattoo, and DeMSI. The dataset consisted of 770 images with their respective ground truth. To evaluate the neural networks, six metrics served as a benchmark, namely, Jaccard Index, Dice, Precision, Recall, Specificity, and Accuracy. The SegResNetVAE neural network achieved the best results with a Jaccard Index of 80.25%, Dice of 88.5%, and Accuracy of 95.5%. A segmentation technique was used for a comparative study that does not rely on modern techniques like deep learning. Then a segmentation approach was put into practice based on morphological operations and thresholds. However, this performed on par with the poorest neural network (RegUNet), demonstrating the advantages of incorporating artificial intelligence and studying different neural networks.

In a second phase, a study was carried out on NIR images in segmentation, for which a very versatile image acquisition setup was built, which allowed the acquisition of RGB images and NIR images. With this, three datasets were built - a dataset with only RGB images, another with only NIR images, and a dataset that combines the two formats. The six neural networks were again tested under the same metrics. In the NIR dataset, the best network was SegResNet, while in the RGB dataset, it was the neural network SegResNetVAE. In the last dataset, which concatenates two types of images, the results were not so good, which may be because the dataset is very short, and also that the ground truth generation is done semi-automatically and hence is not as perfect as the first built dataset, however according to the results the neural networks appear to be easier to segment the NIR images than the RGB images given this last dataset. This dissertation's completion marks the start of a project by 2Ai Laboratory that will support laser tattoo removal therapy with Artificial Intelligence.

5.2 Limitations

There were some limitations during this project, namely, the lack of datasets of tattoos already directed towards clinical treatment, and with ground-truth, which led to a time-consuming work of building a dataset for the intended purpose. In addition, some datasets that were used to create the dataset for this project required licenses and terms of conditions, which made it difficult to obtain them. Another limitation is computational power, to run neural networks a computer capable of doing so is needed. Finally, the dataset created with the acquisition setup was made with temporary tattoos which reflect much more light in comparison with real tattoos. Furthermore, the number of participants in the creation of this dataset was reduced.

5.3 Future Work

Despite all the work done throughout this project, there are some points that can be improved, such as:

- **Building a more robust dataset:** Although various tattoo datasets were built and created during the course of this project, an artificial intelligence system depends on the quality and quantity of data. Therefore, even though creating the dataset is a time-consuming and hard task, it is crucial to the project's success. So there is still room to make the dataset more robust.
- **Alignment of multispectral and RGB images before concatenating:** Images not being correctly aligned is harmful to the neural networks learning process.
- **Implementation of other multi-task architectures to learn from images of different spectrums.**
- **Starting with the collaborative robotic phase:** Many neural networks, parameters, metrics, and outcomes were investigated during the project's development; this might be a significant step forward for the second phase, which entails utilizing a collaborative robot to assist the physician in removing the tattoo. The segmentation performed by artificial intelligence will serve as the foundation for the collaborative robot.

References

[1] A. Laser, "Tattoo removal by the numbers: Trends on the rise." [Online; Accessed on January 2023].

[2] Google, "tattoo removal, tattoo removal - explorar - google trends." [Online; Accessed on January 2023].

[3] V. Khosla, V. Joseph, and H. Gordon, "Tattoos: What is their significance?," *Advances in Psychiatric Treatment*, vol. 16, pp. 281–287, 2010.

[4] D. Dreamstime, "Tattoo process. the tattooing process causes damage to the epidermis and dermis. stock vector - illustration of info, medical: 112780948.," 2021. [Online; Accessed on August 2022].

[5] S. M. Ferguson-Rayport, R. M. Griffith, and E. W. Straus, "The psychiatric significance of tattoos," *The Psychiatric Quarterly*, vol. 29, pp. 112–131, 2005.

[6] J. Levy, M. Sewell, and N. Goldstein, "II. A Short History of Tattooing," *The Journal of dermatologic surgery and oncology*, vol. 5, pp. 851–856, 1979.

[7] V. M. Hsu, A. S. Aldahan, S. Mlacker, V. V. Shah, and K. Nouri, "The picosecond laser for tattoo removal," *Lasers in Medical Science*, vol. 31, pp. 1733–1737, 2016.

[8] G. B. Palermo, "Tattooing and tattooed criminals," *Journal of Forensic Psychology Practice*, vol. 4, pp. 1–25, 2004.

[9] A. J. W. Taylor, "A Search Among Borstal Girls For The Psychological and Social Significance Of Their Tattoos," *The British Journal of Criminology*, vol. 8, pp. 170–185, 04 1968.

[10] L. Manuel and P. D. Retzlaff, "Psychopathology and tattooing among prisoners," *International Journal of Offender Therapy and Comparative Criminology*, vol. 46, pp. 522–531, 2002.

[11] M. Atkinson, "Pretty in Ink: Conformity, Resistance, and Negotiation in Women's Tattooing," *Sex Roles*, vol. 47, pp. 219–235, 01 2002.

[12] N. Scheinfeld, "Tattoos and religion," *Clinics in Dermatology*, vol. 25, pp. 362–366, 7 2007.

[13] S. Wohlrab, J. Stahl, and P. M. Kappeler, "Modifying the body: Motivations for getting tattooed and pierced," *Body Image*, vol. 4, pp. 87–95, 3 2007.

[14] IBISWorld, "Industry market research, reports, and Statistic." [Online; Accessed on August 2022].

[15] B. Wire, "Global tattoo removal market - drivers and forecast from Technavio. ," 2017. [Online; Accessed on August 2022].

[16] B. Wire, "Revitalize laser care brings leading-edge technology and extensive medical expertise to tattoo removal in Denver. ," 2018. [Online; Accessed on August 2022].

[17] S. Varma, S. W. Lanigan, and S. Varma, "The psychological, social and financial burden of tattoos," *British Journal of Dermatology*, vol. 140, pp. 37–45, 1999.

[18] "Home tattoo removal – get the lowdown | Laser Clinic Bournemouth." [Online; Accessed on September 2022].

[19] "Judy She-nanigans: April 2012." [Online; Accessed on September 2022].

[20] S. Tattoo, "Tattoo Removal Procedures." [Online; Accessed on September 2022].

[21] J. Mohammad, "Layered dermabrasion: A simple controlled method for the treatment of tattoos," *The Internet Journal of Plastic Surgery*, vol. 3, 2006.

[22] M. Sepehri and B. Jørgensen, "Surgical treatment of tattoo complications," *Current Problems in Dermatology (Switzerland)*, vol. 52, pp. 82–93, 2017.

[23] C. for Surgery, "Surgical Tattoo Removal London." [Online; Accessed on September 2022].

[24] R. Saini, S. M. Winhoven, and J. Kaufman, "Hypertrophic scar after chemical tattoo removal," *Dermatologic Surgery*, vol. 34, pp. 1599–1602, 11 2008.

[25] E. Bernstein, "Laser tattoo removal," *Seminars in Plastic Surgery*, vol. 21, pp. 175–192, 8 2007.

[26] I. Robotics, A. Society, R. S. of Japan, I. of Electrical, and E. Engineers, *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* 2019.

[27] J. K. Henley, F. Zurfley, and M. L. Ramsey, "Laser tattoo removal," *StatPearls*, 7 2022.

[28] Google, "Tattoo process," 2021. [Online; Accessed on August 2022].

[29] T. D. Lab, "How ink colour affects laser tattoo removal," 2022. [Online; Accessed on January 2023].

[30] L. I. Naga and T. S. Alster, "Laser Tattoo Removal: An Update," *American Journal of Clinical Dermatology*, vol. 18, pp. 59–65, 2 2017.

[31] Fotona, "Tattoo Removal." [Online; Accessed on August 2022].

[32] M. Dermatology, "What are the steps of laser tattoo removal?." [Online; Accessed on January 2023].

[33] "Potential risks of laser tattoo removal: Orange coast aesthetics: Cosmetic specialists." [Online; Accessed on January 2023].

[34] IBM, "Artificial intelligence in medicine | ibm." [Online; Accessed on January 2023].

[35] S. C. Society., I. of Electrical, E. E. S. A. Section, I. of Electrical, E. E. R. 8, I. of Electrical, and E. Engineers, *2nd International Conference on Computer Applications Information Security (ICCAIS' 2019) : 01-03 May, 2019 Riyadh, Kingdom of Saudi Arabia*. 2019.

[36] A. A. Aly, S. B. Deris, and N. Zaki, "Research review for digital image segmentation techniques," *International Journal of Computer Science and Information Technology*, vol. 3, pp. 99–106, 10 2011.

[37] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G. Z. Yang, "Deep learning for health informatics," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, pp. 4–21, 1 2017.

[38] S. A. Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh, "Deep semantic segmentation of natural and medical images: A review," *Artificial Intelligence Review*, vol. 54, pp. 137–178, 10 2019.

[39] G. W. Lindsay, "Convolutional neural networks as a model of the visual system: Past, present, and future," *Journal of Cognitive Neuroscience*, vol. 33, pp. 2017–2031, 1 2020.

[40] C. M. Portugal, "Dispositivos de remoção de tatuagens para a sua clínica." Online; Accessed on August 2022].

[41] DataToBiz, "Importance of datasets in machine learning and ai research." [Online; Accessed on January 2023].

[42] D. D. Lab, "What is ground truth in machine learning? | domino data lab." [Online; Accessed on January 2023].

[43] DLabs.AI., "How to implement Artificial Intelligence in your company?," 2021. [Online; Accessed on January 2022].

[44] M. Ngan, P. Grother, and K. Hanaoka, "Tattoo Recognition Technology - Evaluation (Tatt-E) performance of tattoo identification algorithms," 10 2018.

[45] H. Han, J. Li, A. Jain, S. Shan, and X. Chen, "Tattoo Image Search at Scale: Joint Detection and Compact Representation Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, pp. 1–1, 1 2019.

[46] X. Di and V. M. Patel, "Deep Tattoo Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 119–126, 2016.

[47] I. A. for Pattern Recognition, Robotics, and N. C. M. A. for Computer Vision, *2016 23rd International Conference on Pattern Recognition (ICPR) : 4-8 Dec. 2016.* 2016.

[48] M. J. Wilber, E. Rudd, B. Heflin, Y. M. Lui, and T. E. Boult, "Exemplar codes for facial attributes and tattoo recognition," pp. 205–212, IEEE Computer Society, 2014.

[49] M. Martin, J. Dawson, and T. Bourlai, "Large Scale Data Collection of Tattoo-Based Biometric Data from Social-Media Websites," in *2017 European Intelligence and Security Informatics Conference (EISIC)*, pp. 135–138, 2017.

[50] "Tattoo detection based on cnn and remarks on the nist database," *2016 International Conference on Biometrics, ICB 2016*, 2016.

[51] T. Hrkac, K. Brkic, and Z. Kalafatic, "Tattoo Detection for Soft Biometric De-identification Based on Convolutional Neural Networks," 2016.

[52] H. Han and A. K. Jain, "Tattoo Based Identification: Sketch to Image Matching." Online; Accessed on August 2022].

[53] P. Duangphasuk and W. Kurutach, "Tattoo skin detection and segmentation using image negative method," pp. 354–359, 2013.

[54] D. hai shi da xue, J. da xue, M. I. Systems, M. I. Systems, B. I. F. of Automatic Control. Technical Committee on Economic, I. of Electrical, and E. Engineers, *IEEE ICCSS 2017 : 2017 International Conference on Information, Cybernetics, and Computational Social Systems : Dalian, Liaoning, China, July 24-26, 2017*. 2017.

[55] J. Kim, A. Parra, H. Li, and E. J. Delp, "Efficient graph-cut tattoo segmentation," vol. 9410, p. 94100H, SPIE, 3 2015.

[56] J. D. Allen, N. Zhao, J. Yuan, and X. Liu, "Unsupervised tattoo segmentation combining bottom-up and top-down cues," vol. 8063, p. 80630L, SPIE, 5 2011.

[57] A. T. Sturgeon and S. Kumar, *Getting started with Paint.NET : create amazing images easily and professionally with one of the best free photo editors available*. Packt Pub, 2013.

[58] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523–3542, 2022.

[59] S. Manuela and G. Dixe, "Deteção de danos no interior de um carropor visão por computador."

[60] Prabhu, "Understanding of convolutional neural network (CNN) - deep learning," 2019. [Online; Accessed on January 2022].

[61] L. Pauly, H. Peel, S. Luo, D. Hogg, and R. Fuentes, "Deeper networks for pavement crack detection," pp. 479–485, International Association for Automation and Robotics in Construction I.A.A.R.C), 2017.

[62] S. Saha, "A comprehensive guide to Convolutional Neural Networks," 2017. [Online; Accessed on January 2022].

[63] X. Di and V. M. Patel, "Deep learning for tattoo recognition," *Advances in Computer Vision and Pattern Recognition*, vol. PartF1, pp. 241–256, 2017.

[64] S. Sinha, "Alexnet architecture," 2021. [Online; Accessed on January 2022].

[65] G. R. Koch, "Siamese neural networks for one-shot image recognition," 2015.

[66] S. Jadon, "An overview of deep learning architectures in few-shot learning domain," *ArXiv*, vol. abs/2008.06365, 8 2020.

[67] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *LNCS*, vol. 9351, pp. 234–241, 10 2015.

[68] M. J. Willemink, W. A. Koszek, C. Hardell, J. Wu, D. Fleischmann, H. Harvey, L. R. Folio, R. M. Summers, D. L. Rubin, and M. P. Lungren, "Preparing medical imaging data for machine learning," *Radiology*, vol. 295, pp. 4–15, 2020.

[69] T. Shah, "About Train, Validation and Test Sets in Machine Learning." [Online; Accessed on September 2022].

[70] K. Kamnitsas, W. Bai, E. Ferrante, S. McDonagh, M. Sinclair, N. Pawlowski, M. Rajchl, M. Lee, B. Kainz, D. Rueckert, and B. Glocker, "Ensembles of multiple models and architectures for robust brain tumour segmentation," vol. 10670 LNCS, pp. 450–462, Springer Verlag, 2018.

[71] S. Valderrama, "Considerations when choosing a machine learning model." [Online; Accessed on September 2022].

[72] J. Brownlee, "How to Choose Loss Functions When Training Deep Learning Neural Networks." [Online; Accessed on September 2022].

[73] A. Mishra, "Metrics to evaluate your machine learning algorithm. Medium.," 2020. [Online; Accessed on January 2022].

[74] E. Tiu, "Metrics to Evaluate your Semantic Segmentation Model." [Online; Accessed on September 2022].

[75] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and Flexible Image Augmentations," *Information*, vol. 11, no. 2, 2020.

[76] A. Anaya-Isaza, L. Mera-Jimenez, J. M. Cabrera-Chavarro, L. Guachi-Guachi, D. Peluffo-Ordonez, and J. I. Rios-Patino, "Comparison of current deep convolutional neural networks for the segmentation of breast masses in mammograms," *IEEE Access*, vol. 9, pp. 152206–152225, 2021.

[77] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. Roth, and D. Xu, "Unetr: Transformers for 3d medical image segmentation," 3 2021.

[78] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, and K. H. Maier-Hein, "nnu-net: Self-adapting framework for u-net-based medical image segmentation," 9 2018.

[79] "Deepreg: a deep learning toolkit for medical image registration," *Journal of Open Source Software*, vol. 5, p. 2705, 11 2020.

[80] A. Myronenko, "3D MRI Brain Tumor Segmentation Using Autoencoder Regularization," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, (Cham), pp. 311–320, Springer International Publishing, 2019.

[81] upGrad blog, "Types of Optimizers in Deep Learning Every AI Engineer Should Know." [Online; Accessed on October 2022].

[82] A. Topiwala, L. Al-Zogbi, T. Fleiter, and A. Krieger, "Adaptation and evaluation of deep learning techniques for skin segmentation on novel abdominal dataset," pp. 752–759, Institute of Electrical and Electronics Engineers Inc., 10 2019.

[83] MATLAB, "MATLAB - MathWorks - MATLAB Simulink." [Online; Accessed on October 2022].

[84] Anaconda, "Anaconda | The World's Most Popular Data Science Platform." [Online; Accessed on October 2022].

[85] NVIDIA, "CUDA Toolkit - Free Tools and Training | NVIDIA Developer." [Online; Accessed on October 2022].

[86] Python, "Welcome to Python.org." [Online; Accessed on October 2022].

[87] PyTorch, "PyTorch." [Online; Accessed on October 2022].

[88] "Project Jupyter | Home." [Online; Accessed on October 2022].

[89] M. Panalytical, "Near Infrared Spectroscopy (NIRS)." [Online; Accessed on October 2022].

[90] D. Bryson, J. Wright, and K. Barker, "The identification of tattoo designs under cover-up tattoos using digital infrared photography," *Journal of Visual Communication in Medicine*, vol. 36, pp. 104–110, 12 2013.

[91] H. Clarkson and W. Birch, "Tattoos and human identification: Investigation into the use of x-ray and infrared radiation in the visualization of tattoos," *Journal of Forensic Sciences*, vol. 58, pp. 1264–1272, 2013.

[92] M. Martin and T. Bourlai, "Enhanced Tattoo Image Quality Assessment Through Multispectral Sensing," *IEEE Sensors Letters*, vol. 1, no. 6, pp. 1–4, 2017.

[93] A. Laser, "Why laser wavelengths matter for removing tattoos." [Online; Accessed on January 2023].

[94] M. . Simulink, "Setting up image acquisition hardware - matlab simulink." [Online; Accessed on January 2023].

[95] NI, "Ni vision image acquisition system concepts - ni." [Online; Accessed on January 2023].

[96] effiLux, "Adjustable beam power ring - datasheet effi-ring." [Online; Accessed on October 2022].