A COMPARATIVE STUDY OF DEEP LEARNING ALGORITHMS ON SKIN LESION SEGMENTATION

(DERİ LEZYONU SEGMENTASYONU ÜZERİNE DERİN ÖĞRENME ALGORİTMALARININ KARŞILAŞTIRMALI BİR ÇALIŞMASI)

by

FATİH ERGİN, B.S.

Thesis

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

in

COMPUTER ENGINEERING

in the

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

of

GALATASARAY UNIVERSITY

JUNE 2020

Approval of the thesis:

A COMPARATIVE STUDY OF DEEP LEARNING ALGORITHMS ON SKIN LESION SEGMENTATION

submitted by **FATİH ERGİN** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department, Galatasaray University** by,

Examining Committee Members:

Assist. Prof. Dr. İ. Burak Parlak
Supervisor, **Computer Engineering Department, GSU** ______________

Assist. Prof. Dr. JAMES DOE
Industrial Engineering Department, GSU ______________

Assist. Prof. Dr. JESSICA DOE
Industrial Engineering Department, University of Kentucky ______________

Date: ______________

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Computer assisted radiology becomes an interdisciplinary domain between mathematics, medicine and engineering. Tumor detection, analysis, classification are main problems in digital radiology for diagnosis and follow-up. A physician or an oncologist involves in the care of patients by regarding detailed reports of carcinoma in situ that analyze the pathology of suspicious lesions. Deep learning applied to several fields in medicine is considered as an intervention for oncology. Even if the final treatment of the lesion is decided by the oncologists or the surgeons in a case of resection, image based analysis of lesions (benign or malign) promises automated decision making for radiology. Because of the increased ability of machine learning techniques to transform input data into high level presentations, deep learning techniques used for image analysis have become important for helping physicians in the last few years. Skin lesion detection and classification are current challenges in medical image analysis. Skin diseases are difficult to recognize because of the similarity between lesions and low contrast between the lesions and the skin. Dermatologic image processing benefits from the evaluation scores of neural nets. To help the physicians for the accurate diagnosis, the medical field has an increasing interest in this technology, especially in the diagnosis of skin lesions. This paper presents a comparision between various state of the art deep learning based approaches namely Capsule Network, Generative Adversial Networks, and U-Net to solve skin lesion analysis problems using a dermatoscopic image that contains a skin tumor. Gaussian Adversarial Networks (GAN) bring a new architecture in machine learning by adding generator and discriminator steps in data analysis. U-Net is a neural network architecture which aims to overcome the insufficient data problem in medical imaging field by extracting contexual details even if the dataset is small. In this article, GAN and U-Net architectures have been implemented on two dimensional skin lesion images from the International Skin Imaging Collaboration (ISIC) 2017 Challenge. After the preprocessing, colored images have been trained in both GAN and U-Net. The experiment setup has been enriched by adding incremental noise on tumor images before models training. The evaluation has been tested through accuracy, sensitivity, specificity, Dice coefficient and Jaccard coefficient parameters.

In conclusion, test results showed that both GAN and U-Net architectures provide a robust approach in skin lesion analysis and there is no actual superiority againist each other.

Keywords : DEEP LEARNING, IMAGE PROCESSING, MEDICAL IMAGING, MACHINE LEARNING, NEURAL NETWORKS

RÉSUMÉ

La radiologie assistée par ordinateur devient un domaine interdisciplinaire entre les mathématiques, la médecine et l'ingénierie. La détection, l'analyse, la classification des tumeurs sont les principaux problèmes en radiologie numérique pour le diagnostic et le suivi. Un médecin ou un oncologue intervient dans la prise en charge des patients en consultant des rapports détaillés de carcinome in situ qui analysent la pathologie des lésions suspectes. L'apprentissage profond appliqué à plusieurs domaines de la médecine est considéré comme une intervention en oncologie. Même si le traitement final de la lésion est décidé par les oncologues ou les chirurgiens en cas de résection, l'analyse des lésions (bénignes ou malignes) basée sur l'image promet une prise de décision automatisée pour la radiologie. En raison de la capacité accrue des techniques d'apprentissage automatique à transformer les données d'entrée en présentations de haut niveau, les techniques d'apprentissage en profondeur utilisées pour l'analyse d'images sont devenues importantes pour aider les médecins au cours des dernières années. La détection et la classification des lésions cutanées sont des défis actuels dans l'analyse d'images médicales. Les maladies de la peau sont difficiles à reconnaître en raison de la similitude entre les lésions et du faible contraste entre les lésions et la peau. Le traitement d'image dermatologique bénéficie des scores d'évaluation des réseaux neuronaux. Pour aider les médecins à poser un diagnostic précis, le domaine médical s'intéresse de plus en plus à cette technologie, notamment au diagnostic des lésions cutanées. Cet article présente une comparaison entre diverses approches basées sur l'apprentissage en profondeur Capsule Network, Generative Adversial Networks et U-Net pour résoudre les problèmes d'analyse des lésions cutanées à l'aide d'une image dermatoscopique contenant une tumeur cutanée. Les réseaux adverses gaussiens (GAN) apportent une nouvelle architecture dans l'apprentissage automatique en ajoutant des étapes de générateur et de discriminateur dans l'analyse des données. U-Net est une architecture de réseau neuronal qui vise à surmonter le problème de données insuffisantes dans le domaine de l'imagerie médicale en extrayant des détails contextuels même si l'ensemble de données est petit. Dans cet article, les architectures GAN et U-Net ont été implémentées sur des images de lésions cutanées bidimensionnelles Défi 2017 de la

Collaboration internationale en imagerie cutanée (CITI). Après le prétraitement, des images en couleur ont été formées à la fois dans GAN et U-Net. La configuration de l'expérience a été enrichie par l'ajout de bruit incrémentiel sur les images tumorales avant la formation des modèles. L'évaluation a été testée par les paramètres d'exactitude, de sensibilité, de spécificité, de coefficient de dés et de coefficient de Jaccard. En conclusion, les résultats des tests ont montré que les architectures GAN et U-Net fournissent une approche robuste dans l'analyse des lésions cutanées et qu'il n'y a pas de supériorité réelle à nouveau.

Mots Clés : L'APPRENTISSAGE EN PROFONDEUR, TRAITEMENT D'IMAGE, L'IMAGERIE MÉDICALE, APPRENTISSAGE DE LA MACHINE, LES RÉSEAUX DE NEURONES

ÖZET

Bilgisayar destekli radyoloji, matematik, tıp ve mühendislik arasında disiplinlerarası bir alan haline gelmiştir. Tümör tespiti, analizi ve sınıflandırması, tanı ve takip için dijital radyolojideki temel problemlerdendir. Bir doktor veya onkolog, şüpheli lezyonların patolojisini analiz eden in situ karsinomun detaylı raporlarını dikkate alarak hastaların bakımında yer alır. Tıpta çeşitli alanlara uygulanan derin öğrenme, onkolojiye bir müdahale olarak kabul edilir. Rezeksiyon durumunda lezyonun son tedavisine onkologlar veya cerrahlar karar verse bile, görüntü temelli lezyon (iyi huylu veya kötü huylu) analizi, radyolojide otomatik karar vermeyi vaat eder. Son yıllarda makine öğrenimi tekniklerinin insan algısı için anlamsız olabilecek verileri anlamlı hale dönüştürme yeteneğinin artmasıyla beraber, görüntü analizi için kullanılan derin öğrenme teknikleri, son yıllarda hekimlere yardımcı olan önemli bir araç haline gelmiştir. Deri lezyonlarının saptanması ve sınıflandırılması tıbbi görüntü analizinde güncel zorluklardandır. Lezyonlar arasındaki benzerlik ve lezyonlar ile cilt arasındaki düşük kontrast nedeniyle cilt hastalıklarını tanımak zordur. Dermatolojik görüntü işleme, doktorların doğru tanı koymasına yardımcı olmak için bu teknolojiyle, özellikle cilt lezyonlarının teşhisiyle ilgili olan kısmına yoğun ilgi göstermektedir. Bu makale, Kapsül Ağı, Generatif Adversial Ağlar ve U-Net gibi son teknoloji derin öğrenme yaklaşımlarıyla cilt tümörü içeren dermatoskopik görüntüleri kullanarak cilt lezyonu analiz problemlerini çözmede yol göstermeyi amaçlayan bir karşılaştırma sunmaktadır. GAN, klasik CNN'lerden farklı olarak derin öğrenme modeline üretici ve ayırıcı adımlar ekleyerek yeni bir mimari sunmuştur. U-Net, medikal görüntüleme alanının temel sorunlarından biri olan yetersiz veri problemini bağlamsal detayları az veriden başarılı bir şekilde çıkararak aşmayı amaçlayan bir derin sinir ağı mimarisidir. Bu makalede, GAN ve U-Net mimarileri Uluslararası Cilt Görüntüleme İşbirliği (ISIC) 2017 Yarışması'ndan alınan iki boyutlu cilt lezyonu görüntüleri üzerine uygulanmıştır. Ön işlemeden sonra, renkli görüntüler hem GAN da hem de U-Net'te eğitilmiştir. Deney düzeneği, model eğitimi öncesinde tümör görüntülerine artımlı gürültü eklenerek zenginleştirilmiştir. Değerlendirme; doğruluk, duyarlılık, özgüllük, Dice katsayısı ve Jaccard katsayısı parametreleri ile yapılmıştır. Sonuç olarak, test sonuçları hem GAN hem de U-Net mimarilerinin cilt lezyonu analizinde tutarlı bir yaklaşım sağladığını ve farklı gürültü oranlarında birbirlerine karşı gerçek bir üstünlük kuramadıklarını göstermiştir.

Anahtar Kelimeler : DERİN ÖĞRENME, GÖRÜNTÜ İŞLEME, MEDİKAL GÖRÜNTÜLEME, MAKİNE ÖĞRENMESİ, SİNİR AĞLARI

1 INTRODUCTION

During the last few years, deep neural networks have gained considerable attention in several problems of computer vision. The new network hierarchies present complex transfer modalities to deal with adaptive learning tasks. Deep neural networks (DNN) allow machines to learn hybrid data structures of mathematical models which can be used to achieve comprehensive data analysis. Image semantics would be resolved using relevant models. The learning rate is measured in DNN by the achievement of comprehensive data analysis. Generative adversarial networks (GANs) deal with a new hierarchy in intelligent systems. The network scheme conforms to better training performance with less annotations. GAN achievement is derived through a competitive learning where the model consists of different stacks. The processing in layers is characterized with multiple levels of abstraction from high-dimensional input data.

In medicine, lesion detection becomes more efficient with new models based on deep learning networks from histological to radiological acquisitions. Recent studies reveal that detection performance of deep networks has even matched or exceeded human-level performance in several tasks such as diabetic retinopathy and tumor detection (Gulshan et al., 2016),(Işın et al., 2016). Early detection of cancer is considered as one the most complex and hard problems in radiology. The follow-ups and repeated cases are also challenges for the correct decision making. Over the last decade, the progress and the integration of DNN enable rapid diagnosis of patients at these risk groups. Even if the final medical decision must be taken with a specialist, DNNs might reduce the time for the diagnostic errors and workload of physicians. Therefore, DNN performance is not compared with physicians in our study. The evaluation of DNN based segmentation is performed through Ground Truth ; a mask which identifies the whole area or volume in target images.

The common goal of deep learning techniques is to recursively learn computational model parameters using a training data set to gradually improve the model in performing the desired purpose. Using many previously unseen data, models can also perform the same task accurately once a computer is trained for a specific task. The strong generalization ability of deep learning now distinguish it from the other techniques of machine learning. The detection and evaluation criteria result that the use of multilayered hierarchy of GAN shows valid scores. The variation of inter observer, the inhomogeneity in image scale encompass the complexity of automatic lesion detection. In skin lesion segmentation, International Skin Imaging Collaboration (ISIC) focuses on the analysis and the improvement of big datasets. Annotated image corpora is considered as a challenge in deep learning for detection and classification. Even if recent studies in data challenges promised valid results for clinical applications, the performance evaluation shows that training datasets might cause variation in skin lesion detection. In order to promote automatic analysis in this field, GAN technique that represents promising scores is preferred. Although lots of work has been proposed, there is still a margin of performance improvement for both skin lesion segmentation and classification. The International Skin Imaging Collaboration (ISIC) is a cooperation focusing on the automatic analysis of skin lesion, and has continuously expanded its datasets since 2016. In ISIC 2017, annotated datasets for three processing tasks related to skin lesion images, including lesion segmentation, dermoscopic feature extraction and lesion classification, were released for researchers to promote the accuracy of automatic melanoma detection methods.

In this paper, we provide a new application area of deep neural networks in skin lesion analysis. We note that dermoscopic feature extraction is relatively a new problem in deep learning to address the detection and the classification of lesions. The following section presents the medical imaging techniques, image segmentation architectures for automatic diagnosis of skin lesions from dermoscopic images and related studies of DNN for medicine. The third section shows the basis of neural networks and convolutional neural networks. Used neural network architectures, dataset, tools and deep learning frameworks are explained in fourth section along with our corresponding formulation through computational parameters and the statistical evaluation. Our detection results are given through statistical parameters in the fifth section. Finally, the assessment of examined neural networks in skin lesion detection is concluded through the current state-of-the-art and prospective improvements.

2 RELATED WORKS

2.1 Medical Imaging Techniques

Medical Image Segmentation aims to accurately determine the location and shape of the body part or structure within a 2D or 3D image automatically or semi-automatically (Merjulah and Chandra, 2019). Medical images are created by many different modalities which will be examined below in detail. Wide modality range and the high variability of human anatomy is the major difference of medical image segmentation. Medical images are divided into several areas based on the question to achieve detecting or segmenting the tumor or mass. Irregularities, blurred vision borders, low contrast between lesion and skin, air bubles are the some of various artifacts that makes segmentation medical imaging challenging (Guo and Ashour, 2019).

Medical Imaging Techniques (MIT) are concerned to create medical images to be able to examine internal structures of body without opening up it (Kasban et al., 2015). In this section, common medical imaging techniques are being investigated.

2.1.1 Common Medical Imaging Techniques

This section represent the review of widely-used medical imaging techniques namely X-ray Radiography, Magnetic Resonance Imaging, and Computed Tomography.

X-ray Radiography is an imaging technique that uses ionizing electromagnet radiation, such as X-ray which is a type of high-energy electromagnetic radiation (Kasban

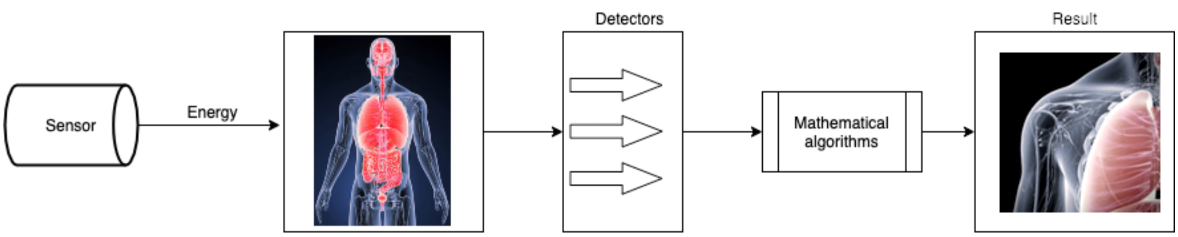


Figure 2.1: Medical Imaging Concept

Figure 2.2: Sample X-ray Images

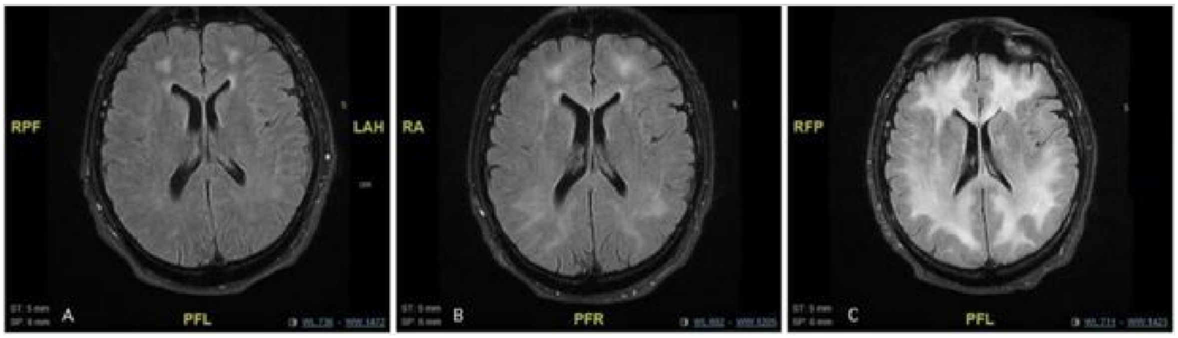


Figure 2.3: Sample MRI Images

et al., 2015). As it can be seen at Figure 2.2, there is a trade-off between radiation level and image contrast which should be chosen carefully. X-ray passes through the body and is absorbed at different levels according to several factors such as the different tissue density. Mammography which deals with the scanning of breast tissue is one of the well-known application areas of X-ray Radiography.

Magnetic Resonance Imaging (MRI) is one of the most common techniques for medical imaging which uses magnetic fields and frequencies in the radio wave spectrum to create images of body tissue (Mehmood et al., 2013). Changes in proton density and magnetic spin relaxation times can be used as distinctive in detecting abnormal tissues. MRI is based on visualizing these changes. MRI imaging can be enhanced by using a contrast solution, e.g. gadolinium, which will change the relaxation properties of some tissues under certain conditions.

Advantages of using MRI include painless, ionizing-free radiation, and high spatial resolution with operator independent usage. However, it may hard to use in people who cannot remain calm, and because of the relatively long scanning and post processing

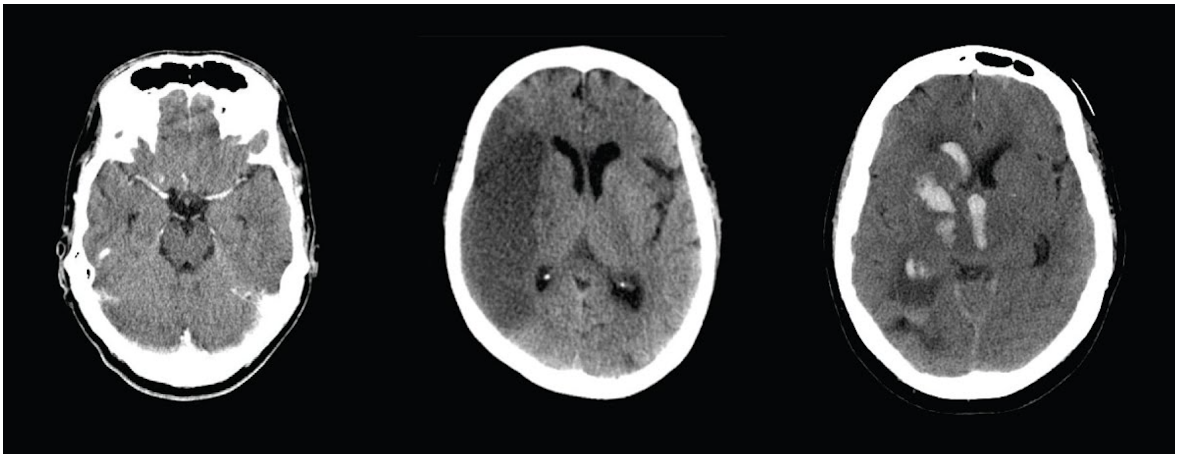


Figure 2.4: Sample CT Images

time, MRI does not offer real time results. Example of MRI image is shown in Figure 2.3.

Computed Tomography (CT) is a computerized X-ray diagnostic imaging test which is supported with a cathode ray tube used to create detailed image of parts of the human body such as internal organs, blood vessels, bones and soft tissues (*Computed Tomography (CT)*, n.d.). CT scanning is a common method in cancer diagnosis, as it is widely used to determine the size and location of a tumor. It is used to create not only for two dimensional (2D) images but also for three dimensional (3D) images using spiral CT which is basically reconstructing the collected volume data to provide 3D images. Figure 2.4 shows some examples of generated CT images.

Analyzing the parts of human body, diagnosing the abnormalities and traumas, observing the results of the cancer treatments are the common use cases of CTs. It comes with several benefits sucs as getting good spatial resolution, detecting issues quickly and painlessly. CTs, on the other hand, does not provide real time analysis and relatively useless results with the soft tissues with low contrast.

2.1.2 Skin Lesion Imaging Techniques

In this section, imaging techniques used in skin lesions are being investigated.

Table 2.1: Comparision of medical imaging techniques

Imaging Techniques	Spatial resolution	Good contrast	Cost	Real time visualization
Ultrasonography	1mm	Soft tissues	Low	Supported
X-ray	1mm	Soft tissues and fluid	Medium	Unsupported
CT	0.5mm	Hard and soft tissues	High	Unsupported
MRI	0.5mm	Hard and soft tissues	High	Unsupported

Traditional Photography (TP) is the well-known techniques which makes visualizing and monitoring the top layer of the lesion possible (Feit et al., 2004).

Dermoscopy Imaging Technique (DIT) is a real-time noninvasive diagnostic imaging technique which is more successful in distinguishing melanoma concentration than traditional photography (Aljanabi et al., 2019).

Multispectral Imaging (MI) provides information in both spectral and spatial domains. MI systems increase accuracy by calibrating image intensity, controlling exposure time automatically with the help of a multispectral camera that includes different optical filters selected by the problem definition. MI is used in medical imaging to support detecting the lesions about 2mm (Aljanabi et al., 2019). Figure 2.5 shows the images of a skin lesion taken by using different optical filters.

Confocal Laser Scanning Microscopy (CLSM) is an imaging technique that provides real-time details of skin morphology and provides images with the same resolution as traditional microscopes (Gerger et al., 2005). CLSMs are very sensitive for clinical applications but they are relatively expensive to use in there. From right to left clinical, dermoscopical, confocal images of a skin lesion is shown on Figure 2.6.

Ultrasonography which is also known as diagnostic sonography is another imaging technique that is used to create medical imaging to create internal body parts using high frequency broadband sound waves. Because different tissues behave differently under these sound waves, the images generated using the waves reflected by tissue

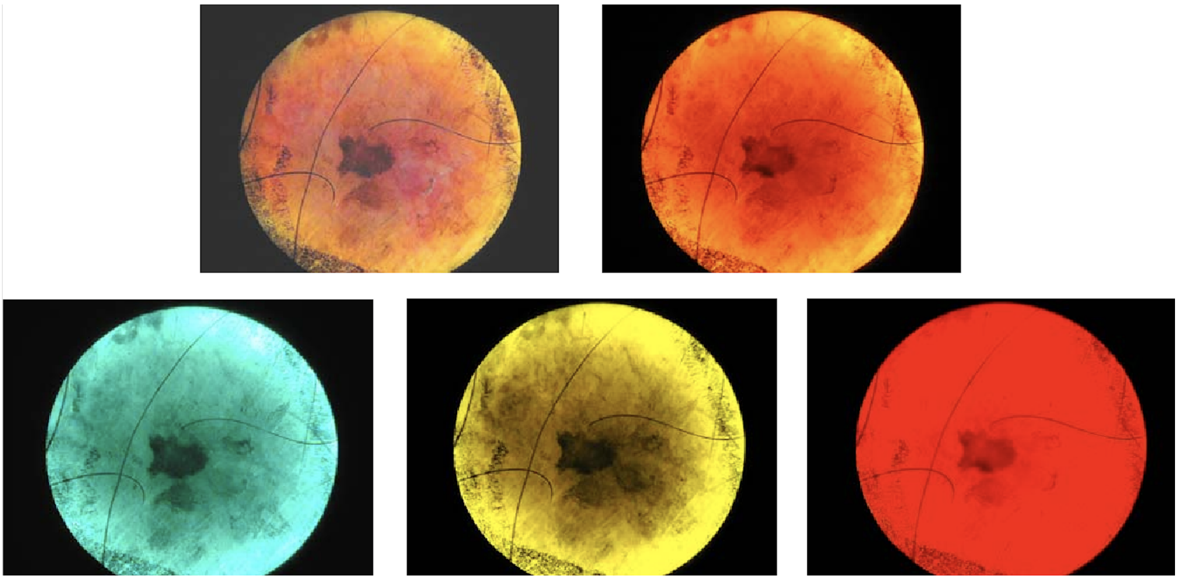


Figure 2.5: Sample Multispectral Images (Dhawan et al., 2009)

Figure 2.6: From right to left clinical, dermoscopical, confocal images of a skin lesion (Ruini et al., 2016)

(Sahuquillo et al., 2013). Calculating the depth of skin cancer is the focused usage of Ultrasonography for this project.

Ultrasonography offer painless real time visualization without ionized radiation in high resolution. But it is a time consuming and operator dependendent imaging technique.

2.2 Image Segmentation Architectures

Fully convolutional network (FCN) is a CNN variant which is a turning point for semantic segmentation literature (Long et al., 2015). After that, many variants of CNN for segmentation has been developed. In this section, commonly used convolutional neural networks which are used for image segmentation starting from FCN are examined.

2.2.1 Fully Convolutional Network

Fully convolutional networks (FCN) indicate that the convolutional neural networks are obtained by dismantling the fully connected layers from deep CNNs (Ulku and Akagunduz, 2019). FCNs are built on traditional classification networks such as AlexNet (Krizhevsky et al., 2012), VGG(Simonyan and Zisserman, 2014), GoogLeNet(Szegedy et al., 2014), and ResNet(He et al., 2016).

Convolutional layers are used instead of fully connected layers to produce outputs with the same size of inputs instead of classification scores which are the outputs of CNNs. FCNs consist of two units encoding and decoding. Convolution and subsampling operations are performed in the encoding unit to encode the lower dimensional latent space. Deconvolution and upsampling are performed in the decoding unit which guarantee the obtaining the same size of output with the input. Because FCNs do not include fully connected layers, it is faster to inference an image if they are compared with the classical CNNs.

Besides the including convolutional layers, skip architecture is one of the main reasons that makes FCNs faster over CNNs. Skip architectures help to prevent losing some information which can be lost because of the dropout or any other architectural decisions which may cause losing information. They provide flowing the summed or concatenated data between downsampling and upsamling blocks. Skip connections are also preserve

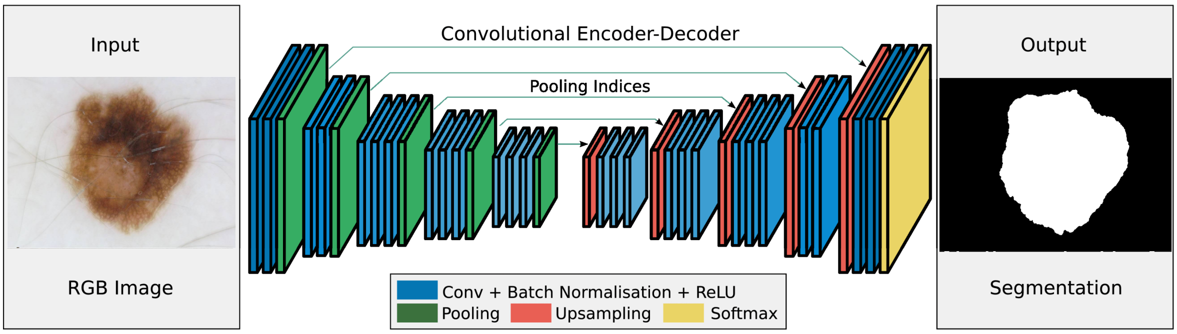


Figure 2.7: SegNet architecture (Badrinarayanan et al., 2017)

the localised information which may lose on pooling layers with bypassing them.

2.2.2 SegNet

Badrinarayanan et al. (2017) proposed a FCN based network architecture, called SegNet, aiming to increase the accuracy of segmentation tasks. As it can be seen in Figure 2.7, the encoder network of the proposed method is consist of 13 convolutional layers

of VGG16 network instead of the original fully connected layers of FCN. A pixel-wise classification layer is added to helps the upsampling on the lower resolution images in the decoder network. The upsampling part is the novel improvement of SegNet.

Encoder is not fully connected in Segnet causes the train parameters to decrease by 90%. There is a corresponding decoder for every 13 encoders and they are responsible for upsampling of the feature map.

2.2.3 U-Net

Ronneberger et al. (2015) proposed a new convolutional neural network namely U-Net which is designed for medical imaging. Medical image segmentation suffers from lack of large dataset which makes capturing image context with localized lesions harder. U-Net aims to achieve competitive results even if the training data are relatively small.

Classical feed-forward CNNs can learn many small information with details with the help of the fully connected layers because large datasets provide a number of parameters to train. On the contrary, large datasets oftenly are not exist or not accesible in medical imaging. Thus, if the neural network aims to accurate results with them, each image in

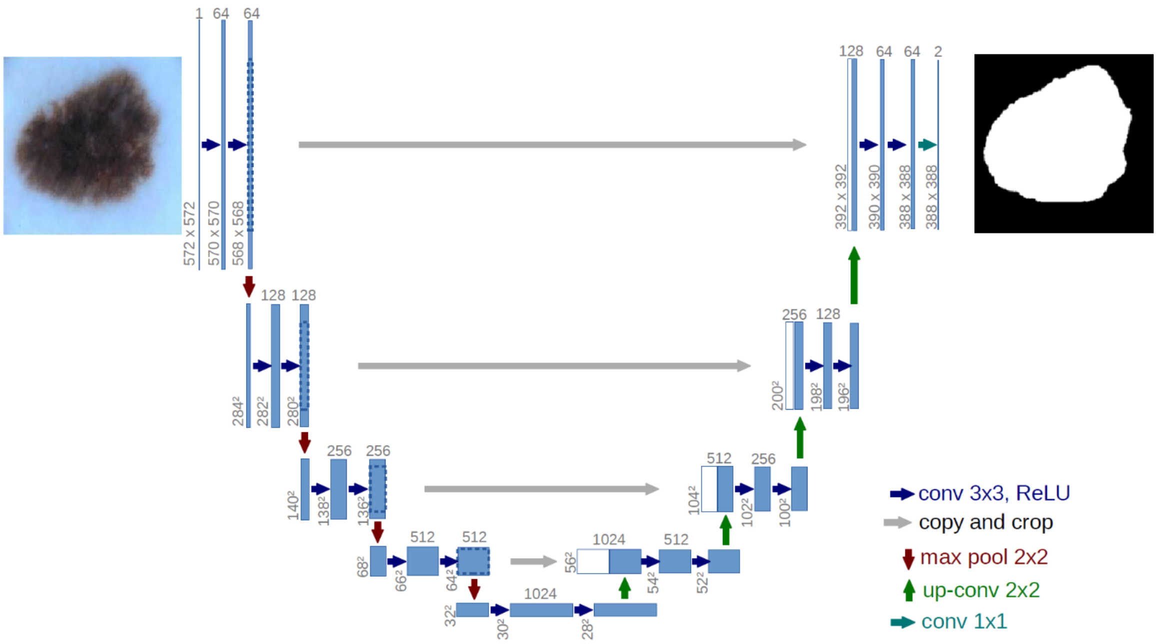


Figure 2.8: U-Net architecture (Ronneberger et al., 2015)

dataset needs to be learnt deeply to extract the features. Up convolutions of the decoder unit which is a replacement of fully connected layers still have trainable parameters and this makes U-Net relatively more successful with smaller datasets by capturing context in detail. As it can be shown in Figure 2.8, U-Nets are made up of contracting and expansive paths on the left and right respectively.

The purpose of the contracting path is to increase resolutions and learn features to capture context while the role of the expanding path is to aid in precise localization with a series of upsampling operations. The contracting path consist of two three-by-three convolutions followed by a ReLU and two-by-two max pooling layers. On the other side, up convolution layers exist to upsample the outputs. Skip connections help to prevent to lose the spatial context combining with upsampled outputs by transfering the low resolution features to expanding path. The authors used a large-weighted loss function to separate boundaries of background labels and touching segments which is a known problem of medical image segmentation.

2.2.4 Generative Adversarial Network

Goodfellow et al. (2014) proposed a deep learning framework which is called generative adversial network (GAN) consisting of two neural networks namely generator and

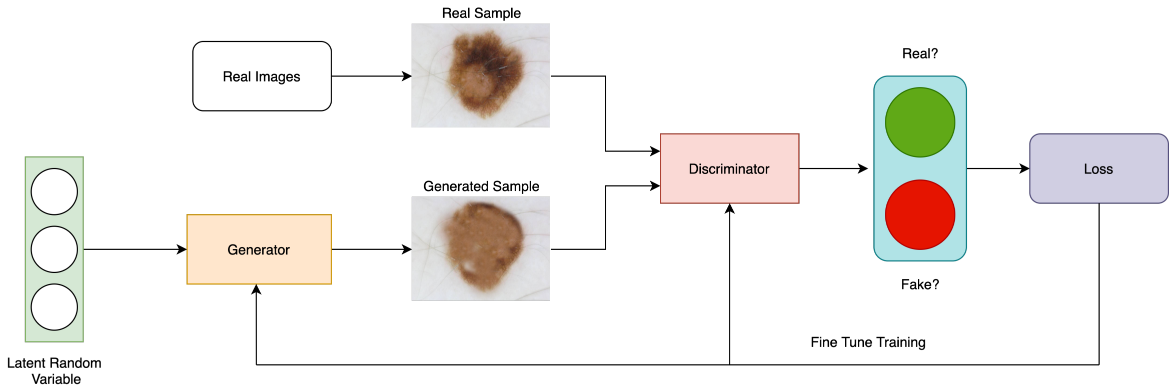


Figure 2.9: GAN architecture

discriminator. The proposed network can be considered as an autoencoder trying to produce a fake version of the real data.

The generator which is the first part of the GANs, generates a sample and the discriminator interprets the sample as a real or fake. The 'real' means that whether the source of the data is training set. The flow can be seen in Figure 2.9.

It looks like a game where the generator tries to fool the discriminator with the samples it creates. The generator are update itself using the output of the discriminator on each iteration and gives better results. GANs have been successfully applied for many image analysis tasks, e.g. inducing realism in synthetic images (Shrivastava et al., 2017), domain adaption (Bousmalis et al., 2017) and data completion (Yeh et al., 2017). Such successful applications of GANs to image processing tasks.

2.3 Related Works

Publication of AlexNet in 2012 have triggered a paradigm change in image segmentation, and then deep learning methods have provided prominent results and became the state-of-the-art in this area in recent years (Quang et al., 2017). In this section, the studies that propose deep architectures for skin lesion segmentation are discussed. Table 2.3 shows the summary of the discussed surveys.

Long et al. (2015) proposed a fully convolutional network from the convolutional networks known to be successful in semantic segmentation. They adapted well-known classification networks such as AlexNet, VGG, GoogleLeNet to fully convolutional networks. Then, to create a successful segmentation, they combined semantic information from a deep layer and the appearance information from a shallow layer to define a new skip architecture. The proposed architecture achieved state-of-the-art segmentation results on PASCAL VOC.

Ronneberger et al. (2015) built a new neural network aimed to be able to get accurate results with insufficient data by using them more effectively. U-Net, the proposed network, is based on classical fully convolutional neural networks and consist of two symmetric paths namely contracting and expanding which is responsible for capturing the context and enabling precise localization respectively. The new neural network proved its success with very few images by winning the International Symposium on Biomedical Imaging (ISBI) 2015 cell tracking challenge. In addition to being able to work with insufficient data, U-Net offers prominent results for training duration with images with relatively higher resolutions such as 512x512. In the following years, new studies showed that the proposed U-shaped network is more successful than C-Means Clustering in ISBI 2017 challenge dataset (Lin et al., 2017).

Yuan et al. (2017) introduced an improved version of FCN model using Jaccard distance as loss function. The aim of this network is increasing segmentation accuracy with solving common dermoscopic image problems such as imbalanced skin and lesion pixels, the existence of various artifacts, and irregular lesion borders. The proposed network achieved better results than the other state-of-the-art networks in ISBI 2016 challenge and PH2 databases.

Yuan (2017) proposed a new skin lesion segmentation network base on Fully Convolutional-Deconvolutional Neural Networks (CDNN). Their main focus is to improve network architecture rather than pre and post processings. Rectified Linear Unit (ReLU) is used as the activation of each layer in the network except to output layer. Internal covariate shift is reduced by adding batch normalization to the output of CD layers. The proposed CDNN architecture took first place in ISBI 2017 challenge.

Yuan and Lo (2017) improved their other skin lesion segmentation architectures by using smaller kernels to optimize the discriminant capacity of their newly proposed neural network. The improved version of the previous work is evaluated on the ISBI 2017 challenge dataset and placed among the top 21 in the ranking.

Bi et al. (2017) proposed a multistage FCN to increase segmentation accuracy of classical FCNs. In this network, first stage FCN focused on learning localization information and coarse appearance, whereas second stage FCN focused on subtle characteristics of the lesion boundaries. A parallel integration method is also introduced to combine the results of the first and second stage FCNs.

Yu et al. (2018) presented an end-to-end deep neural network which consists of two stages called segmentation and classification. The network combines a deep learning method with a local descriptor encoding strategy for dermoscopy image recognition. A pretrained large image dataset is used to extract deep representations of a rescaled image. After that, extracted descriptors are aggregated and encoded with a Fisher Vector to get global features. At the end, the global features are used to classify images with the help of a support vector machine. The proposed network is a fully convolutional residual network (FCRN) and took second place in the segmentation category of the ISBI 2016 challenge.

Al-Masni et al. (2018) developed a framework for skin lesion segmentation via full resolution convolutional networks (FrCN). This method eliminates subsampling layers and learns the full resolution features directly. It is tested with ISBI 2017 challenge and PH2 datasets and has achieved better results againist the well-known state-of-the-art segmentation networks such as U-Net, SegNet and FCN.

Li et al. (2018) introduced a new dense deconvolutional network (DDN) for skin lesion segmentation. The proposed network is based on residual learning and consist of three main parts namely dense convolutional layer, hierarchical supervision (HS), and chained residual pooling (CRP). Dimensions of the input and output images remain unchanged in DDLs. CRP helps to capture contextual background features while HS is responsible for improving the prediction mask. They tested the network with the ISBI 2017 dataset and it achieved 86.6% Dice coefficient indices.

Xue et al. (2018) proposed an Adversial Neural Network (GAN), called SeGAN, based deep neural network aimed to increase accuracy of medical image segmentation. Classical GANs are not as good as expected in providing gradient feedback to the network, because their output is single which may not represent pixel level details of images. Segmentation label maps are created with the help of newly created FCN based segmentor network with a new activation function. Another significant improvement in the proposed network is multi-scale L1 loss function aimed to extract both local and global features which represent the relations between pixels.

Peng et al. (2019) introduced a new adversial network based segmentation architecture consisting of a CNN based discrimination and a U-Net based segmentation networks. This utilized generative adversial network is evaluated on the ISBI 2016 challenge dataset and achieved 97.0% Accuracy rate.

Tu et al. (2019) proposed an adversial network based deep learning framework focused on solving the imbalanced lesion-background problem. The segmentation block of the proposed network is an encoder-decoder network with Dense-Residual block. Deep supervision is utilized with a multi-scale loss function. The network is evaluated on the ISBI 2017 challenge dataset and gained better segmentation results than the other state-of-the-art methods participating in that challenge.

Tschandl et al. (2019) introduced a new FCN where pretrained ImageNet weights are being used to feed the network on ResNet34 layers which are reused as encoding layers. The evaluation results showed that using pretrained weights improved the segmentation score on the ISBI 2017 challenge dataset.

Ninh et al. (2019) proposed a SegNet architecure based FCN framework which aims to decrease the number of upsampling and downsampling layers of classical SegNet architecture to reduce the learned parameters. The proposed network is evaluated on the ISBI 2017 challenge dataset and gained sufficient results in terms of Jaccard Index and Dice coefficient.

Mirikharaji et al. (2019) proposed a deep convolutional neural network framework focused on segmenting skin lesions. The main contribution of this network is the use of two different annotation set consisting of reliable and unreliable annotations. The reliable annotations are marked by experts and have more effect on segmentation result. This reweighting is done by a newly deployed meta-learning approach. The proposed network shows that using different levels of annotation noise on weighting affects the segmentation results and model robustness positively.

Sarker et al. (2019) proposed a lightweight GAN framework, called MobileGAN, aiming to reduce the number of training parameters while keeping the segmentation accuracy high. They combined the channel attention module with the 1D non-bottleneck factorization networks for the generator part of the GAN. MobileGAN is trained with ISIC 2018 training dataset and evaluated with ISBI 2017 challenge dataset. Compared to state-of-the-art models such as FCN, U-Net, or SegNet, the results show that the proposed network has fewer parameters, about 2.3 million, and achieved considerable scores.

Lei et al. (2020) proposed a GAN framework aiming to increase skin lesion segmentation accuracy and win the first part of ISBI 2017 challenge. The segmentation part of the proposed GAN is construct with a skip connection and dense convolution U-Net while the discrimination part is consist of a dual discriminator module. One of the discriminators is responsible for increasing the detection of boundaries while the other one is responsible for learning the contextual informations.

Zafar et al. (2020) proposed an automated neural network architecture aimed to segment skin lesion accurately. Res-Unet, the proposed network, is a combination of two well-known neural networks in image segmentation namely U-Net and ResNet. The other major improvement in this network is using image inpainting for hair removal. It is evaluated on the ISBI 2017 challenge and PH2 datasets and gained Jaccard Index of 77.2% and 85.4% respectively.

Xie et al. (2020) introduced a CNN variant, called MB-DCNN, which consist of three sub CNNs namely coarse segmentation network, mask guided segmentation network, and enhanced segmentation network respectively. The first network is responsible for creating coarse masks which is used on the next network to classify the lesions. The third network is a segmentation network fedded from the second classification network. There are learning transfer between networks to increase the segmentation accuracy. MB-DCNN is tested with the ISBI 2017challenge and PH2 datasets and it achieved Jaccard index of 80.4% and 89.4%.

Table 2.2: Summary of related skin lesion segmentation surveys

Publication	Architecture	Title	Highlights
Long et al. (2015)	FCN	Fully convolutional networks for semantic segmentation	The first FCN implementation for semantic segmentation
Ronneberger et al. (2015)	U-Net	U-net : Convolutional networks for biomedical image segmentation	A new architecture focused on medical image segmentation

Yuan et al. (2017)	FCN	Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance	Jaccard distance bases loss function
Yuan (2017)	CDNN	Automatic skin lesion segmentation with fully convolutional-deconvolutional networks	Adding batch normalization to the output of CD layers
Yuan and Lo (2017)	CDNN	Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks	Discriminant capacity is optimized with smaller kernels
Bi et al. (2017)	FCN	Dermoscopic image segmentation via multistage fully convolutional networks	Multistage FCN with localized responsibilities
Yu et al. (2018)	FCRN	Melanoma recognition in dermoscopy images via aggregated deep convolutional features	A pretrained dataset is used to extract deep representations of images
Al-Masni et al. (2018)	FrCN	Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks	Eliminates subsampling layers and learns the full resolution features directly
Li et al. (2018)	DDN	Dense deconvolutional network for skin lesion segmentation	Residual learning based 3 layered network
Xue et al. (2018)	SeGAN	Adversarial learning with multi-scale loss for skin lesion segmentation	GAN based network with multi-scale loss function

Peng et al. (2019)	GAN	Segmentation of dermoscopy image using adversarial networks	Consist of A CNN based discrimination and a U-Net based segmentation networks
Tu et al. (2019)	GAN	Segmentation of Lesion in Dermoscopy Images Using Dense-Residual Network with Adversarial Learning	Dense-Residual Network based segmentation block
Tschandl et al. (2019)	FCN	Domain-specific classification-pretrained fully convolutional network encoders for skin lesion segmentation	Encoding layers are fed with pretrained weight
Ninh et al. (2019)	SegNet	Skin Lesion Segmentation Based on Modification of SegNet Neural Networks	Reduces the training parameters while keeps the accuracy
Mirikharaji et al. (2019)	CNN	Learning to segment skin lesions from noisy annotations	Reliable and unreliable annotation sets are used together
Sarker et al. (2019)	MobileGAN	MobileGAN : Skin Lesion Segmentation Using a Lightweight GAN	Reduces the training parameters while keeps the accuracy
Lei et al. (2020)	GAN	Skin Lesion Segmentation via Generative Adversarial Networks with Dual Discriminators	Dual discriminator module are used for discrimination block
Zafar et al. (2020)	Res-Unet	Skin Lesion Segmentation from Dermoscopic Images Using Convolutional Neural Network	Combination of U-Net and ResNet

Xie et al. (2020)	MB-DCNN	A Mutual Bootstrapping Model for Automated Skin Lesion Segmentation and Classification	Consists of three sub CNNs with different responsibilities

3 NEURAL NETWORKS IN TUMOR DETECTION

3.1 Artificial Neural Networks

Artificial Neural Networks (ANN) are a category of supervised machine learning algorithms whose design has been inspired by the neurophysiological workings of the human brain (Hill et al., 1994). These networks consist of several layers mainly first layer, last layer and middle layer(s). The first layer is known as the input layer, middle layer which is called as hidden layer and the last layer is the output layer where each layer has several artificial neurons. An example of ANN with a single hidden layer can be seen at Figure 3.1. The most common layer organization is the fully connected layer, where each neuron is fully paired with adjacent neurons.

An ANN transform the inputs into outputs using the activation function, bias and weights. To explain, the sum of inputs is multiplied by weights ; the deviation is added and the result is passed through the activation function. Deciding whether the neuron is active is determined by the activation function. There are different activation functions for different purposes.

During the training, training samples are sent one by one through the network. The output value is calculated for each sample sent. Output values are compared to the target with the help of a loss function to minimize the error rate. At the backpropagation step, the network is updated by propagating errors backwards through the network (LeCun et al., 1988).

3.1.1 Weight Update

Gradient descent is used to minimize the loss function of the neural network. The first-order derivative of the loss function, namely gradient is computed at the current point and it is used to increase the slope in the opposite direction by moving in by the value of self. These two steps are applied to the weights in each cycle. Batch gradient descent (BGD), stochastic gradient descent (SGD), and mini-batch gradient descent are

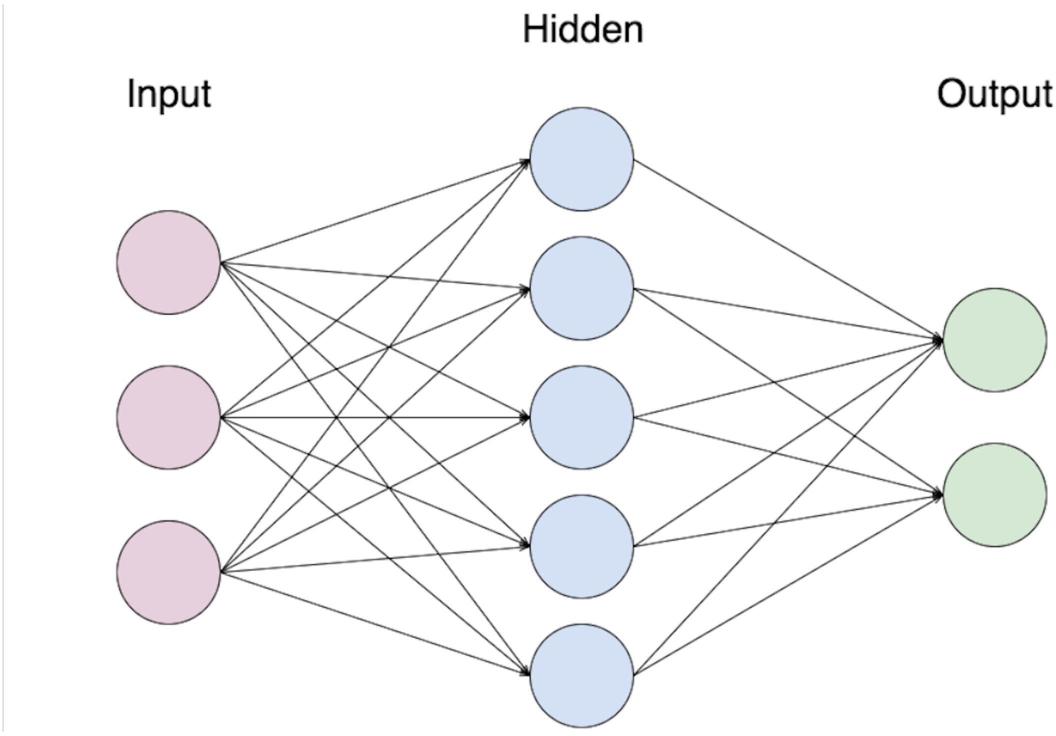


Figure 3.1: Sample ANN with three inputs, one hidden layer, and one output

some of the commonly used weight update methods. In SGD, the training samples are randomly shuffled, to put it another way, the weights are updated after each training sample (Bottou, 2010). On the other hand, all the training samples are used at weight update in batch gradient descent. SGD requires more calculations than batch gradient descent and it is more sensitive than the other. Because SGD is suitable for larger datasets and batch gradient descent is for the smaller datasets, mini-batch gradient descent which is a mixture of SGD and batch gradient descent is developed. It use a batch of a fixed number of samples to update the weights.

3.1.2 Activation Functions

An activation function is used to decide whether an artificial neuron should be activated by calculating the weighted sum of its input. To explain, it basically decides whether the information that the neurons receive is relevant, and ignores it if not. The activation functions can be basically divided into 2 groups as linear and non-linear activation functions. Because linear functions have constant derivatives, there is no relation between the derivative of the linear function and the input value of x. Therefore, the output of functions will not be limited across any range.

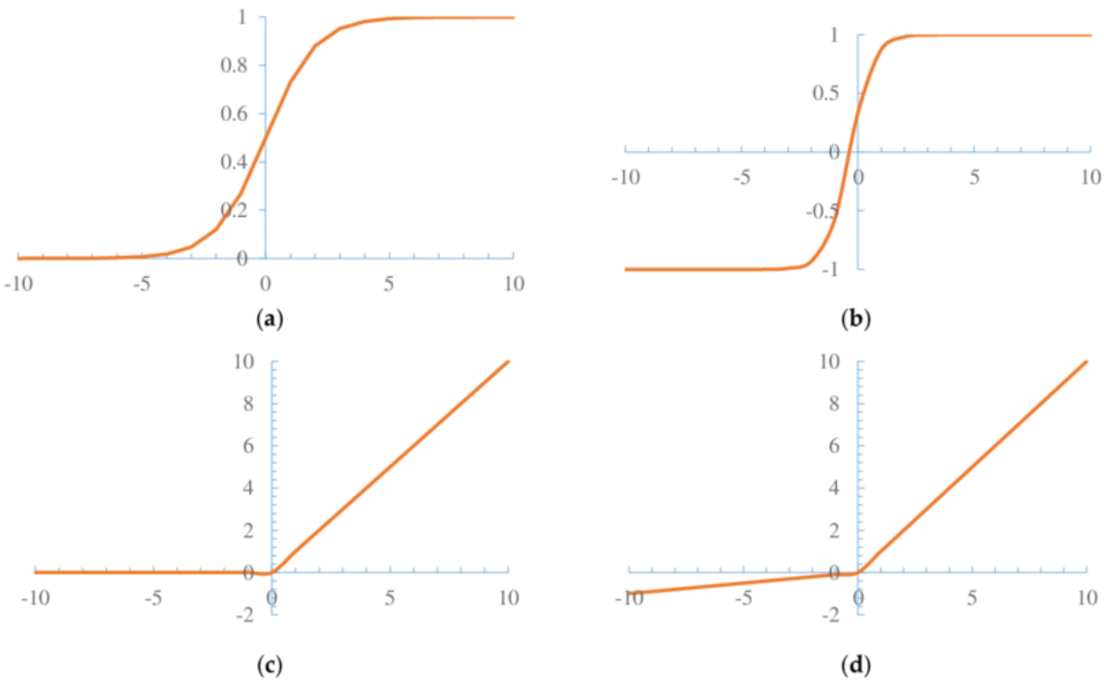


Figure 3.2: Nonlinear activation functions **(a)** Sigmoid, **(b)** Tanh, **(c)** ReLU, and **(d)** Leaky ReLU (Yang and Yang, 2018)

The non-linear activation functions more preferred over the linear activation functions. It makes it easy for the model to generalize or adapt with variety of data and to differentiate between the output. They are basically grouped by their curves and ranges. Commonly used activation functions namely Sigmoid, Softmax, Tanh, ReLU, Leaky ReLU and PReLU are examined in this section.

Sigmoid functions are smooth and is continuously differentiable which means the slope of the Sigmoids can be found for any two points. The sigmoids are monotonic but their derivatives are not. In Sigmoids, the Y values tend to respond very less to changes in X as it can be seen in Equation (3.1). It means that small changes in the X values will cause larger changes in the Y values in this range. So the purpose of this function is to try to keep the Y values to the extremes. This is a very desirable quality when were trying to classify the values to a particular class. If they are compared to linear functions, the outputs are always stay in a fixed range (0,1) unlike the linear functions whose outputs can be in the range of infinity. It is clearly understood that the sigmoid function is not symmetric around the origin and the values received are all positive.

$$h_\theta(x) = \frac{1}{1 + e^{-}\theta^T x} \tag{3.1}$$

Softmax is a Sigmoid function derivative which gives remarkable results in multi variant image classification.

$$y_i(z_i) = \frac{e^{z_i}}{\sum_{k=1}^{k} e^{z_k}} \tag{3.2}$$

Tanh function is an activation function which is scaled version of the Sigmoid function. The main difference between the Tanh and the Sigmoid is that the Tanh is symmetric over the origin while the Sigmoid is not symmetric, so the range of the Tanh is between -1 and 1. Continuity and differantiability of the Tanh is similiar to the Sigmoid, it is continuous and differentiable at all points.

$$tanh(x) = \frac{2}{(1 + e^{(} - 2x)) - 1} \tag{3.3}$$

It is preferred to Softmax if there are no more than two classes in a classification problem. The advantage is that the negative inputs will be mapped strongly negative and the zero inputs will be mapped near zero in the Tanh graph.

Rectified Linear Unit (ReLU) is an activation function which is non-linear and the most widely used activation function while designing neural networks today.

$$f(x) = x^+ = \max(0, x) \tag{3.4}$$

The non-linearity makes backpropagate the errors and have multiple layers of neurons being activated by ReLU function easy. The main advantage of using ReLU function over other activation functions is that it does not activate all the neurons at the same time. As it can be seen in Equation (3.4), if the input is negative it will convert it to zero and the neuron does not get activated. In this activation function, only a few neurons are activated at a time. This makes the network sparse means efficient and easy for computation. The maim problem of ReLU is possibility of creating dead neurons which never get activated.

Leaky ReLU function is an improved version of the ReLU function. The gradient which is *0* for $x < 0$ in ReLU make the neurons die for activations in that region. Leaky ReLU is focused on solving the dead neurons problem. Instead of defining the ReLU function as *0* for x less than *0*, it is defined as a small linear component of x.

$$f(x) = \begin{cases} x & \text{if } x > 0, \\ 0.01x & \text{otherwise.} \end{cases} \tag{3.5}$$

PReLU which is also known as Parameterised ReLU is very similar to the Leaky ReLU.

$$f(x) = \begin{cases} x & \text{if } x > 0, \\ ax & \text{otherwise.} \end{cases} \tag{3.6}$$

In this context, a is a trainable parameter which its values is learnt the network for faster and more optimum convergence. PReLU is preferred when Leaky ReLU fails to passing the relevant information to the next layer with solving the dead neurons problem.

3.1.3 Loss Functions

Loss function which is also called cost function evaluates the penalty between the prediction and the ground truth label during the training process to detect how well neural network models the dataset. The output of the loss function is inversely proportional to the success of the model, meanly higher numbers in the output are the sign of the unsuccessful model. Loss function is used the error which is calculated during backpropagation to update the weights in the negative direction of its derivative.

In this section, some of the commonly used loss functions for image segmentation such as weighted cross entropy, balanced cross entropy, and Dice loss are examined.

Weighted Cross Entropy (WCE) weights the classes based on the fraction of the respective class in the total dataset as it can be seen in Equation (3.7) which is the definition of WCE for prediction p and label $\hat{p}$. Thus, a class with a low fraction of the pixels in the dataset will get a high weighting. This is particularly interesting when the dataset contains unbalanced classes.

$$\mathrm{WCE}\,(p,\hat{p}) = -\left(\beta p \log\left(\hat{p}\right) + (1-p)\log\left(1-\hat{p}\right)\right) \tag{3.7}$$

$\beta >$ *1* should be setted to reduce the false negative rates while $\beta <$ *1* should be setted to reduce the false positive rates.

Balanced Cross Entropy (BCE) another variant of cross entropy which differs from WCE is that weighting the negative examples also.

$$\mathrm{BCE}\,(p,\hat{p}) = -\left(\beta p \log\left(\hat{p}\right) + (1-\beta)(1-p)\log\left(1-\hat{p}\right)\right) \tag{3.8}$$

Dice Loss

The values of Dice ranges from zero to one, and high value means high similarity between the segmentations. For two overlapping regions, the Dice is defined as two times the intersection over the union. Dice loss is defined in Equation (3.9) for prediction p and label $\hat{p}$.

$$\mathrm{DL}\,(p,\hat{p}) = 1 - \frac{2\sum p_{h,w}\hat{p}_{h,w}}{\sum p_{h,w} + \sum \hat{p}_{h,w}} \tag{3.9}$$

3.2 Convolutional Neural Networks

A convolutional neural network (also known as CNN) is a deep neural network inspired in the behavior of biological systems through artificial neurons with learnable weights and biases for image recognition tasks. D.H Hubel and T.N Wiesel discovered that the visual cortex consists of receptive fields that detect light in overlapping subregions. It is the entry point of modeling CNNs (Hubel and Wiesel, 1968). Every neuron responds to stimuli in a restricted region, as in visual cortex of the human brain, and that the overlapping regions of the neurons together cover the entire visual area.

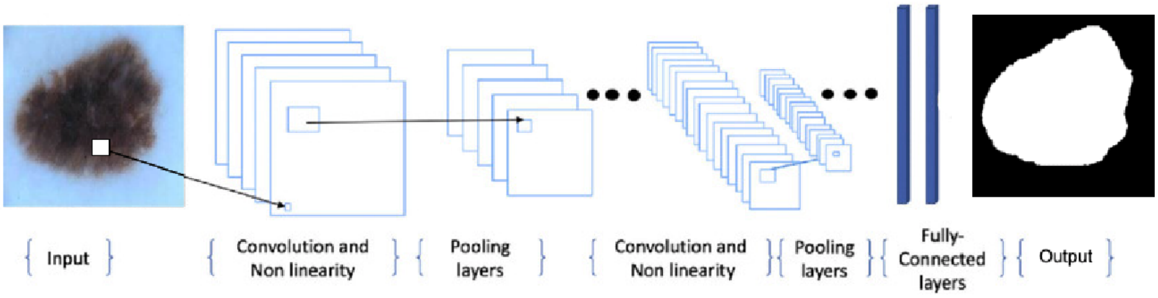


Figure 3.3: A CNN architecture with the result classes of a skin lesion classification

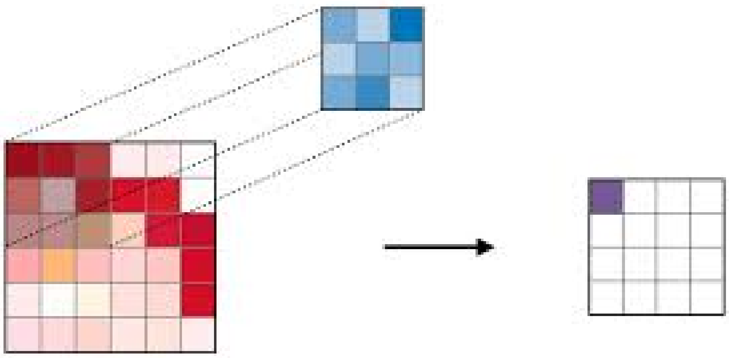


Figure 3.4: Visualization of a convolutional layer of a CNN

A CNN mainly consist of different types of layers as it can be seen in Figure 3.3 including input, convolutional, non-linearity, pooling, and fully connected layers.

Input layer contains input images as a matrix of the raw pixel values.

Convolutional layer is used to extract features of the input data. Each neuron has a local receptive field which means it is not fully connected, but connected to some section of the input to provide abstractions of small sections of the input data. Convolution layers calculate a dot product between the receptive field and the filter by performing convolution. The result of this convolution is a single integer which will be used as the input of the next layer.

The filter is slided over the next receptive fields of the input image repeatedly until there is no unconvolved receptive field left.

Non-linearity layer consists of an activation function, which applies an elementwise activation by thresholding at zero, creates and activation map with taking the output of the convolutional layer in CNNs.

Pooling layer apply a spatial downsampling along the output volume. Pooling layers are commony used to reduce the computational requirements of the neural networks progressively and minimize the overfitting (*Layers of a Convolutional Neural Network - Convolutional Neural Networks for Image and Video Processing - TUM Wiki*, n.d.).

Fully connected layer mainly computes the class scores based on the training dataset. They connect the neurons in layers to each other. The last fully connected layer classifies the generated features with the help of an activation function.

3.3 Regularization Techniques

3.3.1 Data Augmentation

Data augmentation is artificially boosting the diversity and number of training examples by performing random transformations to existing images to create a set of new variants without altering the meaning of the data. Flipping, rotating, adding noise are the some of the commonly used data augmentation techniques.

Data augmentation is used to prevent overfitting and especially useful when the training dataset is relatively small. While some augmentation increases the robustness of the algorithm, irrelevant transformations might make the task hard to learn, and adding new data to the training set will increase the model complexity and reqiured time to build the model.

3.3.2 Dropout

Overfitting is a common problem which is not limited only deep neural networks but includes the different disciplines such as several supervised and unsupervised methods in machine learning. Neural networks can be used to create relation between their input and output to predict the newly added input with acceptible result. It can be said that

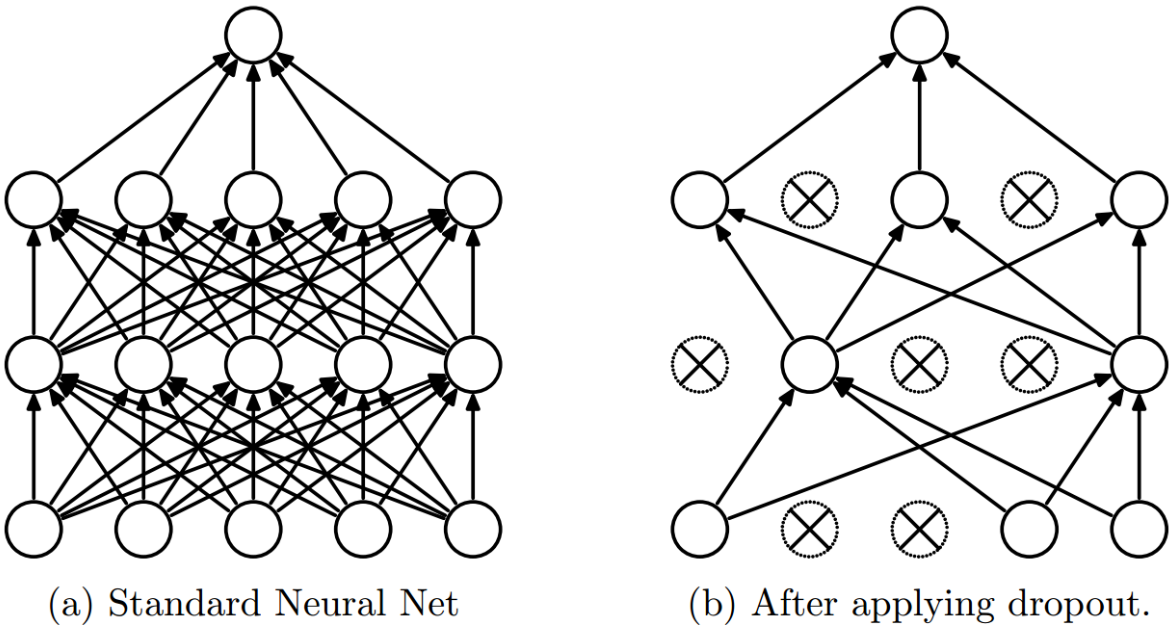


Figure 3.5: Dropout neural network model

there is an overfitting if the results are not good for the unknown test data but good for the training data.

Feeding the neural network with more training data is the simplest way which can be tried to prevent overfitting. This may effective if the newly added training data bring about new features which may increase the representativeness of the model. On the other hand, more training data will require more training time because it increase the model complexity. Bootstrap aggregating is another method which increase the network success (Breiman, 1996). This method classify different subsets of the training data, and fit a model based on these subsets.

Srivastava et al. (2014) said that feature vectors should be combined instead of a single feature detector in order to describe meaningful features. They found out that individual feature detectors start to detect helpful features afte dropping units from the neural network randomly.

Dropout is a method of improvement aims to increase the performance of a neural network by reducing the overfitting (Srivastava et al., 2014). It's not only for CNNs but also all neural networks. At each training step, a new subset is excluded to improve the network's ability to generalize. The amount of exclusion is regulated by the dropout rate. Figure 3.5 shows a regular neural network (a) and a thinned network by applying dropout (b).

3.3.3 Weight Decay

Weight decay is another technique used to prevent overfitting by adding a regularization term such as L1 or L2 to the loss function. L1 regularization is the sum of the absolute value of the weights and produces sparse weight matrices while L2 regularization is the sum of the squares of all the feature weights and make the calculation more computationally efficient.

$$L_2 \text{ regularization term} = ||\boldsymbol{w}||_2^2 = w_1^2 + w_2^2 + ... + w_n^2$$

In L2 regularization, model complexity is dramatically affected by the outlier weights.

3.3.4 Early Stopping

Early stopping is a technique to reduce overfitting using the some part of the training data as a validation set. Training process does not include this data. If the error of the validation set reaches a certain amount, training is stopped at the training phase. It can be said that there is an overfitting exists in the current neural network for the training data.

A significant point of eary stopping is selection of the validation set. It should represent the all data. It can be understood how well the model is generalizing beyond the training data.

4 METHODOLOGY

4.1 Dataset

The first part of ISBI Challenge 2017 (Codella et al., 2018) - Skin Lesion Analysis Towards Melanoma Detection : Lesion Segmentation dataset is used in this project. This dataset has train, validation and test data separately. The training dataset consist of 2000 dermoscopic .jpg images and the related masks with .png format. The dataset include various type of lesions namely malignant melanoma, nevus and seborrhoeic keratosis .
There are also validation and test datasets which contain 150 and 600 images respectively which are provided by the organizers. The results are based on several common image similarity metrics which are given related section.

The images are of various dimensions and the all used neural networks can't handle relatively big images because of their different internal architectures and memory constraints. We also had to resize all images into same dimension to reduce the memory consumption and increase the accuracy as a preprocessing stage. As it can be stated at Figure 4.1, arrays of mask files converted to uint8 to reduce the size of the masks.

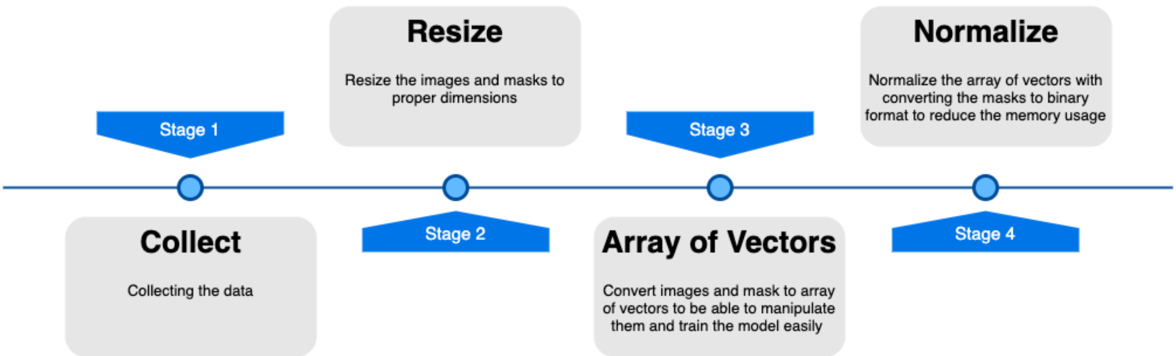


Figure 4.1: Data preparation process

4.2 Networks

Definition of hyper-parameters

— Epoch

— Batch size

— Decay

— Learning rate

— Optimizer

4.2.1 U-Net

4.2.2 SegAN

4.2.3 SegCaps

4.3 Tools and Frameworks

This section presents tools used for development and testing during the project. Python is selected as the main programming language for this project.

4.3.1 Tools

Numpy (Numerical Python) is a scientific computing library for the Python that allows us to perform scientific calculations quickly (Oliphant, 2006). Numpy arrays form the basis of Numpy. Numpy arrays are similar to python lists, but are more useful in terms of speed and functionality than python lists.

Scipy is a package for scientific computing which includes functionality several clustering algorithms, Fourier transforms, linear algebra, interpolation, regression, image and signal processing for the Python programming language (**?**).

Python Imaging Library (PIL) is a free Python library which supports several widely-used image manipulation procedures like per-pixed manipulating, image filtering, image enhancing, masking etc (Anjal and Patil, 2019).

ImageMagick is a open-source, free image editing tool that makes many morphological operation easy for more than 200 image format with its built-in features like resize, flip, transform, or special filters (*ImageMagick - Convert, Edit, or Compose Bitmap Images*, n.d.). It runs on multiple thread to increase performance and supports command-line usage that makes image editing possible for scripting languages.

Jupyter Notebook is an open source web application that allows editing and running code which can be used with over 40 different programming languages (Kluyver et al., 2016). It is a Json based document that has ordered cells which can be live code, equations, visualizations or narrative text.

4.3.2 Deep Learning Frameworks

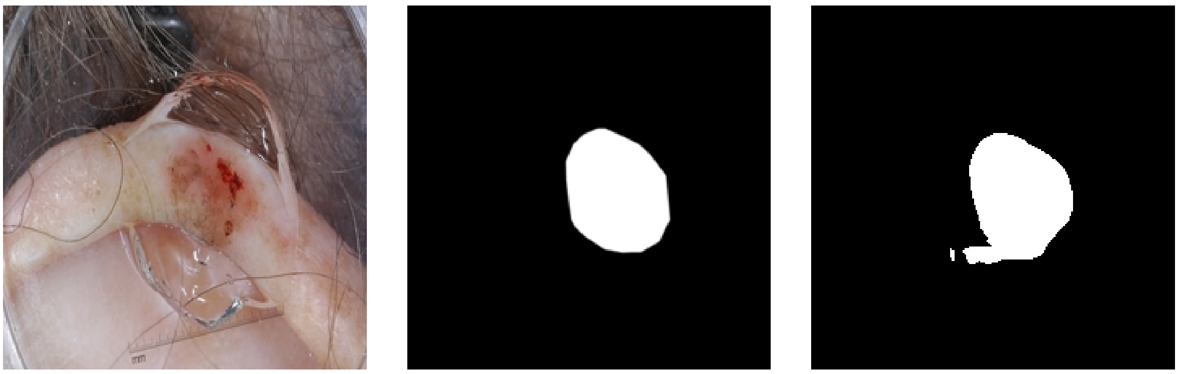
TensorFlow is an open source library for performing numerical computations. Although it can be used for computations in general, it is most commonly used as a tool for machine learning research. TensorFlow can be interfaced using Python and is then translated to a computational graph (Abadi et al., 2015). The computational graph can be fed with the tensors by launching a TensorFlow session which are generalization of N-dimensional arrays. The graph performs a series of mathematical operations on the data. Weight matrices and biases are trainable variables in the TensorFlow graph during a session. Loss functions and optimization algorithms for backpropagation exist in TensorFlow (Johansen and Pedersen, 2019). That makes training a model is as simple as specifying an objective function to optimize for, as well as running the optimizer with a batch of data inside a session.

Keras is a neural networks API for Python (Chollet, n.d.). It runs on top of TensorFlow or Theano (Mohan and Subashini, 2019) which is used as the main neural network framework. Keras is user-friendly and allows for complex models to be created with relatively few lines of code. Keras consists of many commonly used building blocks of neural networks. These are parts as layers, objectives, activation functions and optimizers. The components include parts for convolutional and recurrent neural networks as convolutions, pooling, dropout and batch normalization.

PyTorch is a machine learning framework developed by Facebook which has relatively advantages over TensorFlow in terms of simplicity and usability. It implements dynamic computational graphs which makes dynamic changes on the networks possible with a little effort. Debugging is relatively easy with Pytorch.

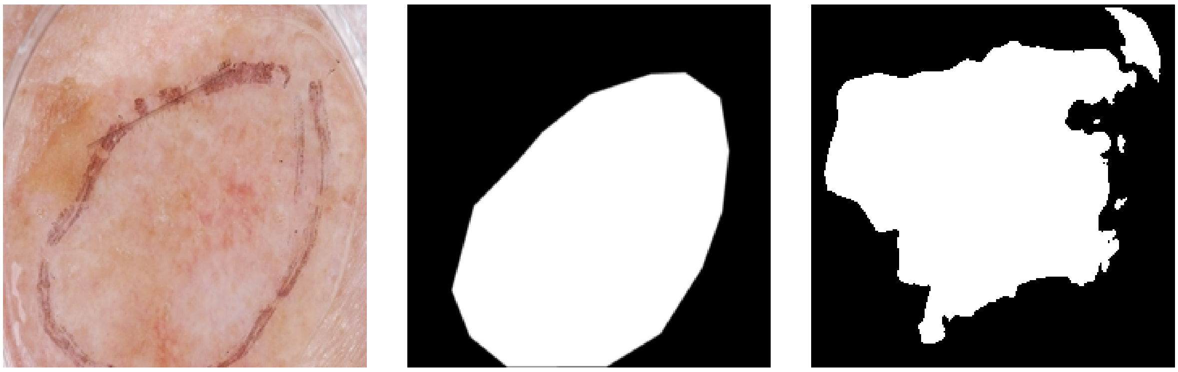
4.3.3 Hardware Requirements

Google Colab which is also known as Colaboratory that requires no setup and runs entirely in the cloud is used in this paper. Colab is a Jupyter Notebook environment aims to support Machine Learning and Artificial Intelligence researches for free, because this kind of process requires serious computational power. Many deep learning projects can be developed with Google Colaboratory on the default GPU processor of it, Tesla K80, using common Deep Learning frameworks and tools like Keras, TensorFlow and PyTorch. Google Colab runs on a connected Google Drive accounts. All models were trained and tested using Tesla K80 GPU which has 25 GB of video memory on a Ubuntu 18.04.



(a) Original image (b) Ground truth (c) Result(dice=0.82)

Figure 4.2: SegAN result with average score at 0% of Gaussian noise



(a) Original image (b) Ground truth (c) Result(dice=0.50)

Figure 4.3: SegAN result with low success compared to average at 0% of Gaussian noise

(a) Original image (b) Ground truth (c) Result(dice=0.91)

Figure 4.4: SegAN result with high success compared to average at 0% of Gaussian noise

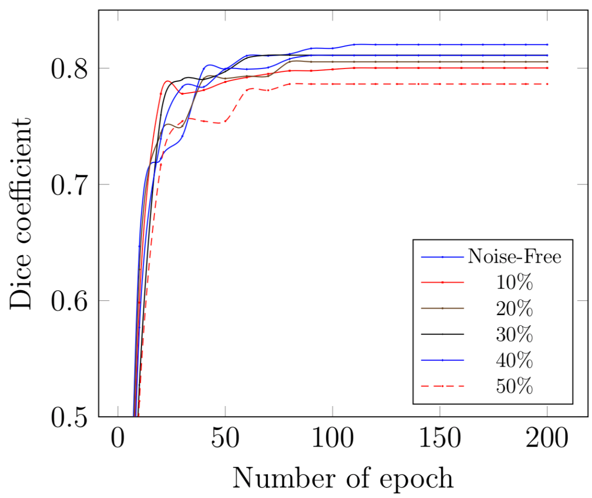
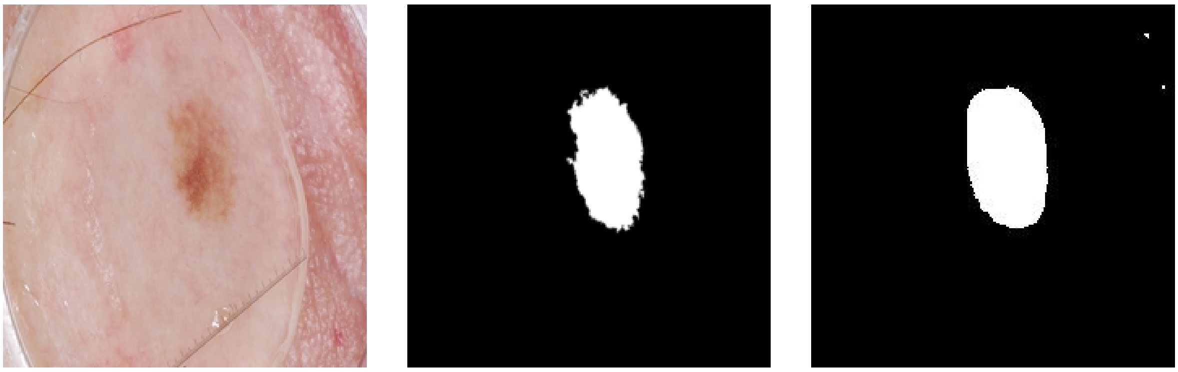
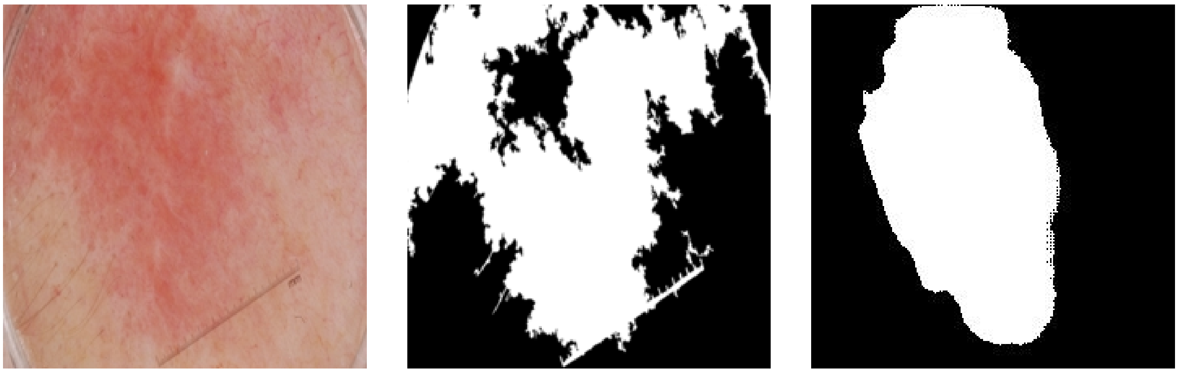


Figure 4.5: Dice results for SegAN at different Gaussian noises by number of epochs



(a) Original image (b) Ground truth (c) Result(dice=0.80)

Figure 4.6: MultiResUNet result with average score at 0% of Gaussian noise



(a) Original image (b) Ground truth (c) Result(dice=0.54)

Figure 4.7: MultiResUNet result with low success compared to average at 0% of Gaussian noise

4.4 Experiments

4.4.1 SegAN with Different Gausian Noises

4.4.2 MultiResUNet with Different Gausian Noises

(a) Original image (b) Ground truth (c) Result(dice=0.90)

Figure 4.8: MultiResUNet result with high success compared to average at 0% of Gaussian noise

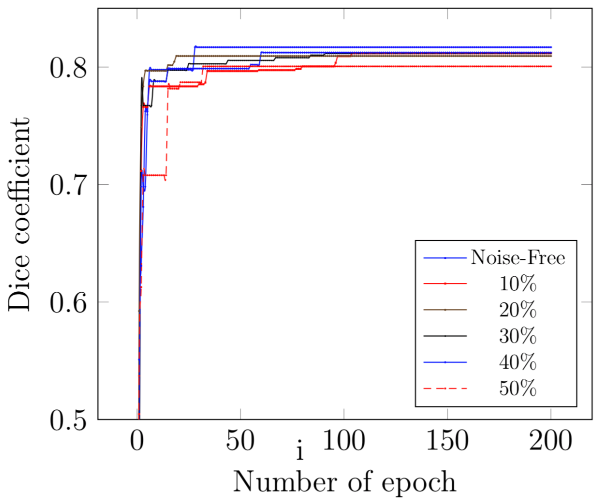


Figure 4.9: Dice results for MultiResUNet at different Gaussian noises by number of epochs

4.5 Evaluation

Several evaluation metrics were used to determine the quality of the models. Dice coefficient, Jaccard index, Accuracy, Sensitivity and Specificity were used to compare the target and predicted segmentation mask. The true positives (TP) determine pixels (or voxels) correctly classified as being part of the segmentation, a false positive (FP) is a pixel incorrectly classified as being part of the segmentation, and a false negative (FN) is a pixel which should have been part of the segmentation but was not.

4.5.1 Metrics

Dice coefficient which is also known as similarity coefficient or F1 score is a similarity metrics computed by comparing the pixel-wise agreement between the groundtruth and its predicted segmentation. Specially, this metric is just used to evaluate the segmentation performance of the model.

$$Dice = \frac{2 * TP}{2 * TP + FN + FP} \tag{4.1}$$

Jaccard index, also known as the Jaccard similarity coefficient, is a similarity metrics which compares predictions with the ground truths by dividing the size of the intersections by size of the unions.

$$Jaccard = \frac{TP}{TP + FN + FP} \tag{4.2}$$

Accuracy measures the proportion of true positives and true negatives whose are correctly segmented instances to the total number of instances. It is derived from sensitivity and specificity which are given below.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4.3}$$

Sensitivity and Specificity are the other metrics used in this project. Sensitivity aims to measure correctly segmented instance ratio while specificity measures incorrectly segmented instance ratio.

$$Sensitivity = \frac{TP}{TP + FN} \tag{4.4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{4.5}$$

5 RESULTS

Table 5.1: Comparision of segmentation results of SegAN at different levels of Gaussian noise with evaluation metrics

Gaussian noise(%)	Dice	Jaccard	Accuracy	Sensitivity	Specificity
0	0.8203	0.7076	0.9264	0.9178	0.9312
10	0.8002	0.6845	0.9192	0.9167	0.9291
20	0.8055	0.6894	0.9248	0.903	0.9294
30	0.8112	0.6958	0.9226	0.9376	0.9263
40	0.811	0.6969	0.9227	0.9081	0.9285
50	0.7864	0.6625	0.9134	0.9282	0.9208

Table 5.2: Comparision of segmentation results of MultiResUNet at different levels of Gaussian noise with evaluation metrics

Gaussian noise(%)	Dice	Jaccard	Accuracy	Sensitivity	Specificity
0	0.8169	0.7221	0.922	0.964	0.9482
10	0.8118	0.7175	0.9172	0.9539	0.9582
20	0.8094	0.7158	0.9192	0.9534	0.9424
30	0.8115	0.7165	0.9186	0.9646	0.9356
40	0.8123	0.72	0.9201	0.9553	0.9384
50	0.8006	0.7057	0.9136	0.9501	0.9368

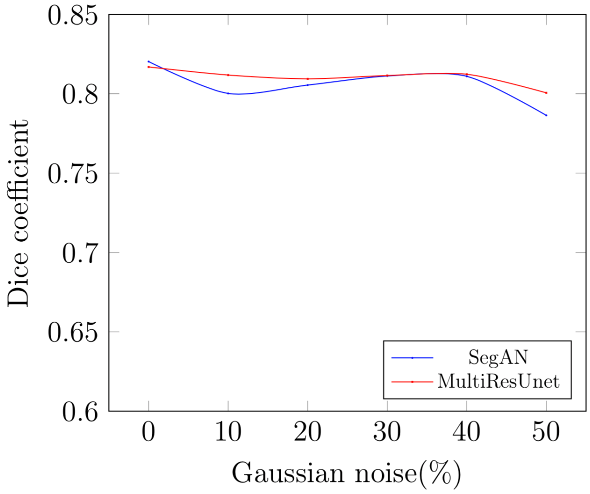


Figure 5.1: Comparision of results of models at different noise levels

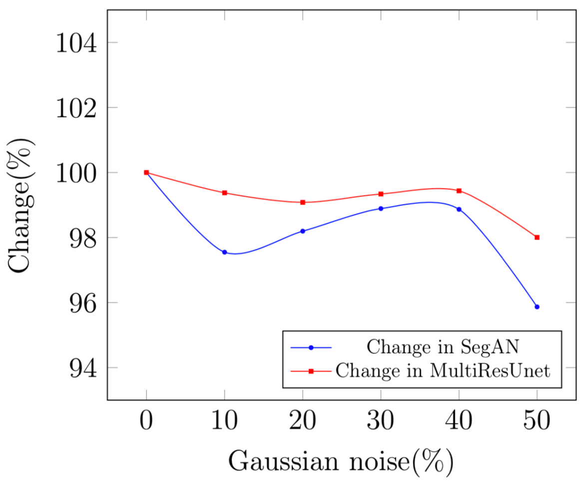


Figure 5.2: Change of success of the models by noise level

6 DISCUSSION

7 CONCLUSION

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M. et al. (2015). Tensorflow : Large-scale machine learning on heterogeneous systems.

Al-Masni, M. A., Al-antari, M. A., Choi, M.-T., Han, S.-M. and Kim, T.-S. (2018). Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks, *Computer methods and programs in biomedicine* **162** : 221–231.

Aljanabi, M. H., Jumaa, F. A., Aftan, A. O., Alkafaji, M. S. S., Alanı, N., Al-Tameemi, Z. H. and Al-Mamoori, D. H. (2019). Various types of skin tumors lesion medical imaging (stlmi) of healthy and unhealthy moles a review and computational of : Segmentation, classification, methods and algorithms, *IOP Conference Series : Materials Science and Engineering*, Vol. 518, IOP Publishing, p. 052014.

Anjal, S. and Patil, A. (2019). Roi based automated meter reading system using python.

Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017). Segnet : A deep convolutional encoder-decoder architecture for image segmentation, *IEEE transactions on pattern analysis and machine intelligence* **39**(12) : 2481–2495.

Bi, L., Kim, J., Ahn, E., Kumar, A., Fulham, M. and Feng, D. (2017). Dermoscopic image segmentation via multistage fully convolutional networks, *IEEE Transactions on Biomedical Engineering* **64**(9) : 2065–2074.

Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent, *Proceedings of COMPSTAT'2010*, Springer, pp. 177–186.

Bousmalis, K., Silberman, N., Dohan, D., Erhan, D. and Krishnan, D. (2017). Unsupervised pixel-level domain adaptation with generative adversarial networks, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3722–3731.

Breiman, L. (1996). Bagging predictors, *Machine learning* **24**(2) : 123–140.

Chollet, F. (n.d.). et al. 2015. keras.

Codella, N. C., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H. et al. (2018). Skin lesion analysis toward melanoma detection : A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic), *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, IEEE, pp. 168–172.

Computed Tomography (CT) (n.d.). `https://www.radiologyinfo.org/en/submenu.cfm?pg=ctscan`. (Accessed on 05/01/2020).

Dhawan, A. P., D'Alessandro, B., Patwardhan, S. and Mullani, N. (2009). Multispectral optical imaging of skin-lesions for detection of malignant melanomas, *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, pp. 5352–5355.

Feit, N. E., Dusza, S. W. and Marghoob, A. A. (2004). Melanomas detected with the aid of total cutaneous photography, *British Journal of Dermatology* **150**(4) : 706–714.

Gerger, A., Koller, S., Kern, T., Massone, C., Steiger, K., Richtig, E., Kerl, H. and Smolle, J. (2005). Diagnostic applicability of in vivo confocal laser scanning microscopy in melanocytic skin tumors, *Journal of Investigative Dermatology* **124**(3) : 493–498.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014). Generative adversarial nets, *Advances in neural information processing systems*, pp. 2672–2680.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J. et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, *Jama* **316**(22) : 2402–2410.

Guo, Y. and Ashour, A. S. (2019). Neutrosophic sets in dermoscopic medical image segmentation, *Neutrosophic Set in Medical Image Analysis*, Elsevier, pp. 229–243.

He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Hill, T., Marquez, L., O'Connor, M. and Remus, W. (1994). Artificial neural network models for forecasting and decision making, *International journal of forecasting* **10**(1) : 5–15.

Hubel, D. H. and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex, *The Journal of physiology* **195**(1) : 215–243.

ImageMagick - Convert, Edit, or Compose Bitmap Images (n.d.). `https://imagemagick.org/index.php`. (Accessed on 06/06/2020).

Işın, A., Direkoğlu, C. and Şah, M. (2016). Review of mri-based brain tumor image segmentation using deep learning methods, *Procedia Computer Science* **102** : 317–324.

Johansen, J. S. and Pedersen, M. A. (2019). *Medical image segmentation : A general u-net architecture and novel capsule network approaches*, Master's thesis, NTNU.

Kasban, H., El-Bendary, M. and Salama, D. (2015). A comparative study of medical imaging techniques, *International Journal of Information Science and Intelligent Systems* **4**(2) : 37–58.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J. B., Grout, J., Corlay, S. et al. (2016). Jupyter notebooks-a publishing format for reproducible computational workflows., *ELPUB*, pp. 87–90.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, pp. 1097–1105.

Layers of a Convolutional Neural Network - Convolutional Neural Networks for Image and Video Processing - TUM Wiki (n.d.). `https://wiki.tum.de/display/lfdv/Layers+of+a+Convolutional+Neural+Network/`. (Accessed on 05/10/2020).

LeCun, Y., Touresky, D., Hinton, G. and Sejnowski, T. (1988). A theoretical framework for back-propagation, *Proceedings of the 1988 connectionist models summer school*, Vol. 1, CMU, Pittsburgh, Pa : Morgan Kaufmann, pp. 21–28.

Lei, B., Xia, Z., Jiang, F., Jiang, X., Ge, Z., Xu, Y., Qin, J., Chen, S., Wang, T. and Wang, S. (2020). Skin lesion segmentation via generative adversarial networks with dual discriminators, *Medical Image Analysis* p. 101716.

Li, H., He, X., Zhou, F., Yu, Z., Ni, D., Chen, S., Wang, T. and Lei, B. (2018). Dense deconvolutional network for skin lesion segmentation, *IEEE journal of biomedical and health informatics* **23**(2) : 527–537.

Lin, B. S., Michael, K., Kalra, S. and Tizhoosh, H. R. (2017). Skin lesion segmentation : U-nets versus clustering, *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, pp. 1–7.

Long, J., Shelhamer, E. and Darrell, T. (2015). Fully convolutional networks for semantic segmentation, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440.

Mehmood, I., Ejaz, N., Sajjad, M. and Baik, S. W. (2013). Prioritization of brain mri volumes using medical image perception model and tumor region segmentation, *Computers in biology and medicine* **43**(10) : 1471–1483.

Merjulah, R. and Chandra, J. (2019). Classification of myocardial ischemia in delayed contrast enhancement using machine learning, *Intelligent Data Analysis for Biomedical Applications*, Elsevier, pp. 209–235.

Mirikharaji, Z., Yan, Y. and Hamarneh, G. (2019). Learning to segment skin lesions from noisy annotations, *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data*, Springer, pp. 207–215.

Mohan, G. and Subashini, M. M. (2019). Medical imaging with intelligent systems : A review, *Deep Learning and Parallel Computing Environment for Bioengineering Systems*, Elsevier, pp. 53–73.

Ninh, Q. C., Tran, T.-T., Tran, T. T., Tran, T. A. X. and Pham, V.-T. (2019). Skin lesion segmentation based on modification of segnet neural networks, *2019 6th NAFOSTED Conference on Information and Computer Science (NICS)*, IEEE, pp. 575–578.

Oliphant, T. E. (2006). *A guide to NumPy*, Vol. 1, Trelgol Publishing USA.

Peng, Y., Wang, N., Wang, Y. and Wang, M. (2019). Segmentation of dermoscopy image using adversarial networks, *Multimedia Tools and Applications* **78**(8) : 10965–10981.

Quang, N. H. et al. (2017). Automatic skin lesion analysis towards melanoma detection, *2017 21st Asia Pacific Symposium on Intelligent and Evolutionary Systems (IES)*, IEEE, pp. 106–111.

Ronneberger, O., Fischer, P. and Brox, T. (2015). U-net : Convolutional networks for biomedical image segmentation, *International Conference on Medical image computing and computer-assisted intervention*, Springer, pp. 234–241.

Ruini, C., Hartmann, D., Saral, S., Krammer, S., Ruzicka, T. and von Braunmühl, T. (2016). The invisible basal cell carcinoma : how reflectance confocal microscopy improves the diagnostic accuracy of clinically unclear facial macules and papules, *Lasers in medical science* **31**(8) : 1727–1732.

Sahuquillo, P., Tembl, J. I., Parkhutik, V., Vázquez, J. F., Sastre, I. and Lago, A. (2013). The study of deep brain structures by transcranial duplex sonography and imaging resonance correlation, *Ultrasound in medicine & biology* **39**(2) : 226–232.

Sarker, M., Kamal, M., Rashwan, H. A., Abdel-Nasser, M., Singh, V. K., Banu, S. F., Akram, F., Chowdhury, F. U., Choudhury, K. A., Chambon, S. et al. (2019). Mobilegan : Skin lesion segmentation using a lightweight generative adversarial network, *arXiv preprint arXiv :1907.00856* .

Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W. and Webb, R. (2017). Learning from simulated and unsupervised images through adversarial training, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2107–2116.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv :1409.1556* .

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014). Dropout : a simple way to prevent neural networks from overfitting, *The journal of machine learning research* **15**(1) : 1929–1958.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. (2014). Going deeper with convolutions. corr, vol. abs/1409.4842.

Tschandl, P., Sinz, C. and Kittler, H. (2019). Domain-specific classification-pretrained fully convolutional network encoders for skin lesion segmentation, *Computers in biology and medicine* **104** : 111–116.

Tu, W., Liu, X., Hu, W., Pan, Z., Xu, X. and Li, B. (2019). Segmentation of lesion in dermoscopy images using dense-residual network with adversarial learning, *2019 IEEE International Conference on Image Processing (ICIP)*, IEEE, pp. 1430–1434.

Ulku, I. and Akagunduz, E. (2019). A survey on deep learning-based architectures for semantic segmentation on 2d images, *arXiv preprint arXiv :1912.10230* .

Xie, Y., Zhang, J., Xia, Y. and Shen, C. (2020). A mutual bootstrapping model for automated skin lesion segmentation and classification, *IEEE Transactions on Medical Imaging* .

Xue, Y., Xu, T. and Huang, X. (2018). Adversarial learning with multi-scale loss for skin lesion segmentation, *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, IEEE, pp. 859–863.

Yang, J. and Yang, G. (2018). Modified convolutional neural network based on dropout and the stochastic gradient descent optimizer, *Algorithms* **11**(3) : 28.

Yeh, R. A., Chen, C., Yian Lim, T., Schwing, A. G., Hasegawa-Johnson, M. and Do, M. N. (2017). Semantic image inpainting with deep generative models, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5485–5493.

Yu, Z., Jiang, X., Zhou, F., Qin, J., Ni, D., Chen, S., Lei, B. and Wang, T. (2018). Melanoma recognition in dermoscopy images via aggregated deep convolutional features, *IEEE Transactions on Biomedical Engineering* **66**(4) : 1006–1016.

Yuan, Y. (2017). Automatic skin lesion segmentation with fully convolutional-deconvolutional networks, *arXiv preprint arXiv :1703.05165* .

Yuan, Y., Chao, M. and Lo, Y.-C. (2017). Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance, *IEEE transactions on medical imaging* **36**(9) : 1876–1886.

Yuan, Y. and Lo, Y.-C. (2017). Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks, *IEEE journal of biomedical and health informatics* **23**(2) : 519–526.

Zafar, K., Gilani, S. O., Waris, A., Ahmed, A., Jamil, M., Khan, M. N. and Sohail Kashif, A. (2020). Skin lesion segmentation from dermoscopic images using convolutional neural network, *Sensors* **20**(6) : 1601.

APPENDIX A PROOF OF SOME THEOREM

BIOGRAPHICAL SKETCH

Write your curriculum vitae here.

PUBLICATIONS

— If you have publications you must write there.