Universiteit Leiden

Master Computer Science

Image Segmentation on the Skin Diseases

Name:	Tejan Gavandi
Student ID:	s2711036
Date:	07/05/2022
Specialisation:	Data Science
1st supervisor:	Dr. Lu Cao
2nd supervisor:	Dr. Vasileios Exadaktylos

Master's Thesis in Computer Science

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Contents

Abstract

Composed of dermis, epidermis and subcutaneous tissues, skin is the largest organ of a human body. Human skin contains seven layers and protects other organs from external invasions such as chemical damage, artificial skin damage and skin diseases. Despite of all this protection skin is not resistant to diseases. Viral, fungal and allergic are the most common types of skin diseases appear on human skin. This thesis focuses on psoriasis which is caused by an autoimmune disorder and is not contagious. People infected with psoriasis develop thick regions of red and white patches. These are also known as lesions. Commonly these lesions appear on elbows, knees, back and scalp.

Objective here is to focus on utilizing deep learning models on skin diseases. Intention is to apply image segmentation techniques to identify and extract damaged skin area from the image. Subsequently, extracted skin lesions will represent morphological and texture features of the psoriasis.

In order to use images for deep learning models, it is must that data is in the right form. Starting with standardizing patients' data acceptable for image segmentation models. This data is collected by trials performed by the patients themselves at home. Which is also called as "trials@home". Each input image is then broken into patches which serve as the input to the deep learning model. Project includes implementation and evaluation of U-Net and FCN-8 and finding the best fit for proposed problem. Size of the lesion is computed based on the tag present in each image. Evaluation of the ground truth and model is done under the guidance of dermatologists at Centre for Human Drug Research which is vital for creating the ground truth data. Trained model is deployed on the cloud platform i.e. Azure Databricks. This allows dermatologists to track the change of lesion size over the period of days. Preferably, expectation that extracted lesion size lowers over the period of time

Keywords— Image Segmentation, U-Net, Deep Learning, Psoriasis, Databricks, Azure

1 Introduction

The field of Artificial Intelligence is not new. One of the greatest architect Da Vinci already tried to automate some of his work[1] back then. Over the last decade we have seen exponential growth and self improvements in the field of Artificial Intelligence. Rapid development, efficiency and simplicity are one of the driving factors of the growth in Deep Learning in medical imaging[2]. Inspired by human visual cortex, convolutional neural networks are among widely used deep learning methods[3]. Other methods of deep learning such as image segmentation, FCN's, and artificial neural networks are practically handy and yield high performance[4]. Apart from efficient and reliable results deep learning methods also contribute to automate the entire process.

Skin is one of the largest organ of the human body and helps to cover muscles, multiple tissues and other organs from external invasions. In this process skin is exposed to outer environment and endure various infections. This demands greater attention and care for skin. The affected or damaged area on the skin is called lesion. Some of the well known and common skin diseases are Acne vulgaris, Eczema, Herpes Zoster, Sunburn and Psoriasis. Despite the fact that these diseases can be cured over the period of time they bring a great deal of unrest to patients life. Skin lesion is the first clinical sign of a skin disease. Early detection of skin disease is a very complex procedure and there is still developing research. With the help of image processing and deep Learning techniques, it becomes possible to perform preliminary diagnosis on the skin. One of the consequential factor of using Artificial Intelligence is that it is done without any physical contact with the skin. In this thesis we will be focusing only on psoriasis. Medical literature illustrates some of the physical effects of psoriasis which include, pain, itching, burning and dry skin[5]. Effects of psoriasis can not only see on persons health but also on non-health related elements. Psoriasis greatly influences patient's quality of life in terms of job, family and friends[6]. This thesis focuses on psoriasis however, developed algorithm can be used for various skin diseases provided availability of corresponding image data.

One of the key aspect of image segmentation is that it allows to look at the problem closely and in depth. This enables the possibility to extract object of interest. Segmentation plays a central role in a broad range of applications including medical image analysis (e.g., tumor boundary extraction and measurement of tissue volumes), autonomous vehicles (e.g., navigable surface and pedestrian detection), video surveillance[7], and augmented reality[8]. Numerous deep-learning based image segmentation algorithms have been developed. Some of these algorithms are based on Convolutional operations, some are based on Encoder-Decoder networks. This work focuses extensively on U-net and FCN-8. U-net is preferred specifically in medical applications because of its efficiency. The network contains encoder which is a series of Convolutional and Max-Pooling layers. Decoder is the exact mirror of encoder network and transposed convolution operations. U-Net implements skip connections which are used to copy uncompressed activations from encoder block to decoder block. These skip connections are effectively combined with different levels of abstractions from different layers[7].

Research for this thesis is conducted at Centre for Human Drug Research. The Centre for Human Drug Research (CHDR) is an independent institute that specialises in cutting-edge early-stage clinical drug research. Combining innovative methods and technologies in order to extract best possible outcome from clinical trials. With an aim to contribute to the clinical and medical research field.

Being able to quantify skin disorder conditions objectively and reliably would allow for the

assessment of pathology severity and treatment results among patients suffering from a dermatological disorder. The objective is to develop a computer vision-based machine learning algorithm that can distinguish skin disorders and identify changes after treatment (e.g., size). Specifically, the system will rely on image data to define the most significant image-based features for the creation of a reliable system.

Trial@Home is CHDR's dedicated wing which addresses off-site clinical trials. Off-site clinical trials opens up the possibilities for data collection in the comfort of the subject's home. It enables data scientists, investigators and researchers to collect valuable information as subjects go about their day-to-day lives, while reducing the number of visits to the unit or trial center. Generally, subjects are under controlled environment when they participate in the trial at center. Trial@home reflect the real-world situation of the subject better. Meanwhile, increased outpatient visits also place new demands on patients, who may not feel able to additionally participate in a clinical trial at center. Most trials can benefit from patient centric data collected outside the clinic. Trial@home can be integrated in a wide range of trials, offering a wealth of additional study data and validated digital endpoints without an increased burden for the subjects. Timely reminders are placed with an aim to inform subject about their scheduled trials.

Data collected from trials at home requires special attention. Since data collection is done by patients, most of the data has some noise which is caused by the environment of the patient. For example, in images collected by patients there will be a lot of background noise. This demands filtering and pre-processing using various image processing techniques.

Once skin lesion has been segmented we calculate its size. Finally dermatologist is able to see the extracted lesion and its size. This allows them to further analyze the disease and progress of patients. Considering that the patient is recovering well we expect extracted lesion and its size to reduce over the period of time. Further analysis of color features extraction and classification of the severity of psoriasis is discussed in section 5.

1.1 Research on Psoriasis

Psoriasis is a chronic, painful and disabling disease which affects patients quality of life with great impact. Research is ongoing on psoriasis and there is no cure on this skin disease. It can occur at any stage of life but most common affected age group is 50-70. Most of the times psoriasis occurs on the skin surface but in some cases it has also been involved on the surface of the nails. Skin lesions are mostly symmetrical and demarcated[6]. Lesion color indicates the severity of psoriasis. In most of the cases, lesion color contains light pink, red, purple, grey and dark purple. Table 1 mentions the grade used by dermatologists at Centre For Human Drug Research to determine the severity of the disease.

Apart from physical burden, psoriasis also affects the patients' emotional and social life. Dis-figuration, disability and loss of productivity are common in patients. There is also significant cost to mental well being, such as higher rate of depression which in turn affects patients' social life influencing his or her environment.

Available treatment of psoriasis is mainly about controlling visible symptoms. Well arranged

Score	Grade
0	None
1	Mild
2	Moderate
3	Severe
4	Very Severe

Table 1: ABVIE Psoriasis Severity Score

therapies along with photo-therapy are available. The need for treatment is usually lifelong and is aimed at remission. So far, there is no therapy that would give hope for a complete cure of psoriasis[6]. Additionally, care for patients with psoriasis requires treating skin lesions and joint involvement. It is also very critical to identify and manage common comorbidity that already exists or may develop, including cardiovascular and metabolic diseases as well as psychological conditions[9].

2 Related Work

2.1 Understanding Deep Learning Techniques for Image Segmentation

In terms of Computer Vision, image segmentation is a technique which divides an image into several meaningful classes[10]. At the pixel level it is nothing but a classification technique. The most classic version of image segmentation is the semantic segmentation. It is a process where each pixel is classified into one of the predefined set of classes, such that pixel belonging to the same class belong to a unique entity[10]. It is widely used in most of the real world problems. Semantics which are being segmented here are not only depend on the data but also on the type of the problem. For example, in self driving, cars, roads, pedestrians etc. will be segmented. There are various supervised[11], unsupervised[11], semi-supervised[12], weakly supervised segmentation[13] algorithms have been developed. Most of these depend on the area of interest of the segmentation. These methods are suitable for binary or multi class segmentation. However, choice of the method is highly influenced by objective of segmentation task. For example, Encoder-Decoder networks such as U-Net[14] is best suited for medical imaging, SharpMask[15] which is based on redefined Deep Mask with multi layer feature fusion was developed by Facebook research AI team(FAIR). Models such as DeepMask and SharpMask are more suitable for the tasks in which targeted segmented area involves human beings. Segnet[16] is more suited for segmenting multiple classes. Segnet[16] contributes a lot when it comes to self driving cars. Following figure 1 illustrates operations, activation functions, inputs and outputs mentioned in the architectures of a image segmentation network. Almost every major image segmentation network is drawn using the components mentioned in the figure1

2.2 Region Growing Segmentation based on Image Color

Segmentation should stop when objects of interest in an image or video have been isolated[17]. In the proposed work[17] authors have used region growing segmentation method which is also known as pixel-based segmentation[18].

The general idea of background segmentation is to automatically generate a binary mask which divides the set of pixels into the set of foreground and the set of background pixels. Statistical measures which can represent the detailed information of the disease are extracted. These are as follows:

1. Mean: Represents average color pixels present in the image.

$$M = \frac{1}{N}\sum_{n=1}^{N} P \tag{1}$$

2. Standard Deviation: Represents square root of variance of the distribution.

$$\sigma = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(P-\mu)^2} \tag{2}$$

3. Skewness: It is a measure which represents degree of the asymmetry in the distribution.

$$S = \sqrt[3]{\frac{1}{N}(\sum_{n=1}^{N} P-\mu)^3} \tag{3}$$

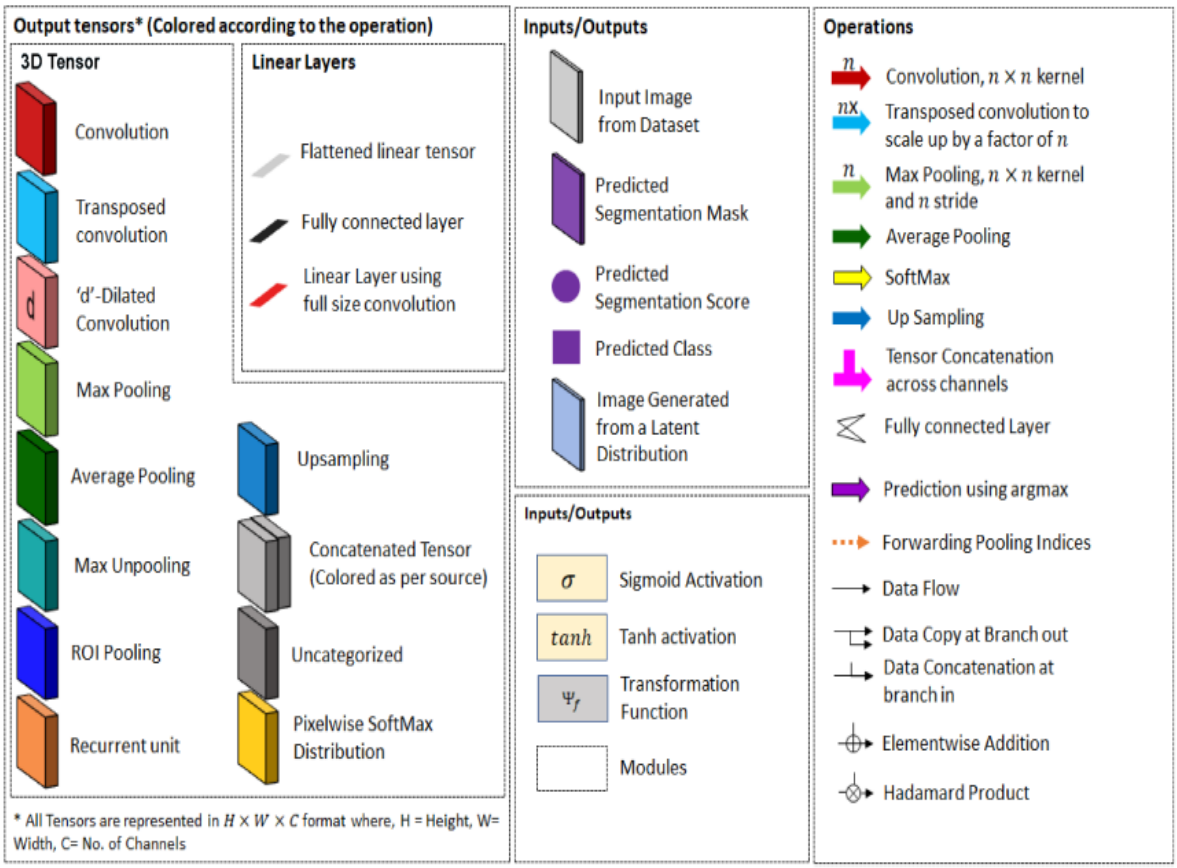


Figure 1: Legends for subsequent diagrams of popular deep learning architectures [Understanding Deep Learning Techniques for Image Segmentation]

4. Variance: Represents variation present in the different pixel values of colors.

$$\sigma = \frac{1}{N}\sum_{n=1}^{N}(P-\mu)^2 \tag{4}$$

These measures help to quantify colors present in an image. To estimate texture features statistical texture descriptor based on Grey level co-occurrence matrix (GLCM) was employed. RGB image histogram plots pixels for each tonal value. This enables us to analyze extracted features closely. For example, color histogram plots the intensity of pixel values present in the lesion. Horizontal axis represents tonal variations whereas vertical axis represents number pixels in that particular tone[17]. Guided setup under the expertise of experienced dermatologist was made to evaluate the output of the segmentation. Further classification of the lesion severity was done. For this, authors used SVM and K-NN along with OR rule[17].

2.3 Skin Lesion Segmentation Using Deep Learning

There have been numerous successful attempts of applying Deep Learning on Skin Diseases. In the proposed method[19] authors experimented segmenting images of Melanoma with Deep CNN model. Melanoma is one of the aggressive form of skin cancer which is on the rise in recent years. Input images are pre-processed in order to handle noisy artifacts[19]. Paper mentions application of the guided filter which helps to reduce effects of the disturbing elements such as body hair. With the help of an annotation tool, ground truth mask is created. Input image and its corresponding mask is the input of the Deep CNN network.

Each block of Deep CNN consists of four layers which are convolutional layer, Max Pooling layer, another set Conv2D and Max Pool. First convolutional layer has a feature of size 6 * 6 * 3. Kernel size in the second convolutional layer is 5 * 5 * 60. Kernel size of Maxpooling layer is 2 * 2 and 3 * 3 respectively. Each convolutional layer consists 60 feature maps[19]. Since the dataset size was small containing 126 images in total, authors also used data augmentation. Evaluation of the segmented lesion was done by a domain expert, in this case a senior dermatologist.

3 Methodology and Experiments

3.1 Dataset Description

Trial@Home data is collected by off-site trials. Meaning patients collect data at the comfort of their home. This enables us to collect images more frequently. In addition, each set of images allows us to look more deeper into smallest changes in psoriasis. Patients take these images using their phone and upload to the platform. Standard dimensions of all images captured with smartphone. Image dimension varies depending on the smartphone however, approximately this is in the range of 2000 by 4000(width * height). Each patient is asked to use a tag near the lesion. This tag is useful for color calibration and to calculate size of the lesion.

The dataset used for this project includes total of 1600 images. This includes total of 18 different patients. All of these patients are in the age group of 28-70. These images are collected by patients respectively using their mobile phones. Each patient has a schedule in which he or she is asked to take pictures daily or weekly depending on the severity of the Psoriasis. This data is stored on the Microsoft Azure cloud. Trained network is stored on the cloud. Using databricks platform we perform analysis and prediction on the patient data. Segmented/predicted mask is stored on the cloud along with patient ID and date on which image was taken. This allows dermatologists to track the progress of the lesion. Along with this we also calculate size of the lesion with the help of a physical tag which each patient uses in each image.

In order to avoid any data leakage we will be working on Virtual Machine and Databricks in Azure platform. Once pre-processing of the images is done, these images will be uploaded to Databricks Azure cloud where further implementation will be carried out. Most of the patients are of light skin tone and this will indeed make model a biased towards light skin tones.

For deep learning and computer vision tasks, we need to have input images in the standard sizes and it's burdensome to work on images with variable size dimensions. Keeping this in mind, dimensions of all input images are reduced to 512 * 512 without losing any proportion ratio and adding padding on the sides. Figure 2 and figure 3 illustrates this operation. It is also important to mention that by reducing image dimensions we are also reducing the the total size of the dataset by almost 70%. However, from figure2 and figure3 it is safe to say that we do not miss any information represented in the image. That is details which are useful to define a lesion are preserved. This reduces cloud storage and computational cost significantly.

3.2 Pre-processing

In order to train Image Segmentation network, we need annotated masks. Image annotation is a method used to mark part of the image which is our object of interest. The task of labelling images for computer vision task is quite challenging, as there are various segmentation algorithms with varied demands for training data. However, for the medical imaging with specific focus on skin diseases semantic segmentation[10] approach have been more successful in recent years. Semantic segmentation allows to annotate right object of interest with pixel-level accuracy.

The technique of semantic segmentation is vital for achieving high accuracy on the lesion. But its very important to select the right tool to maintain the accuracy. There are several tools

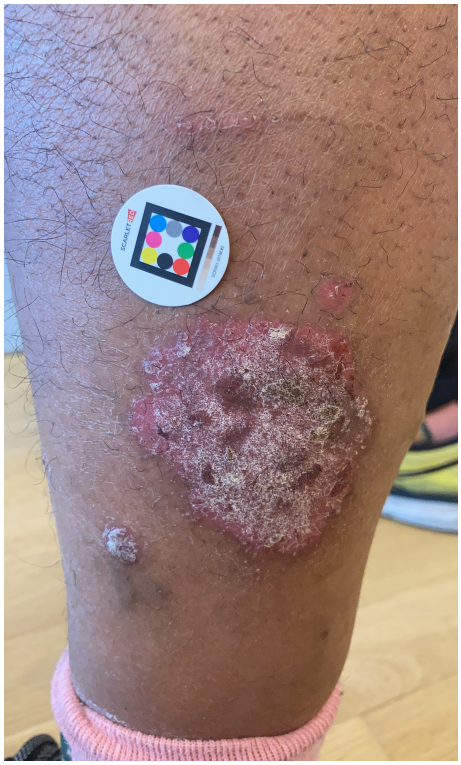


Figure 2: Original Sample Image

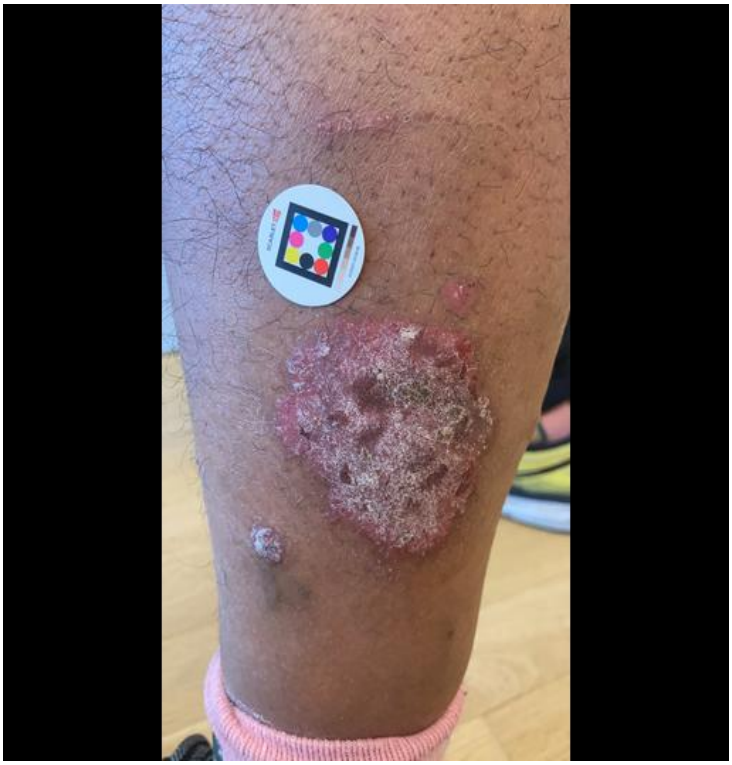


Figure 3: Reshaped Image

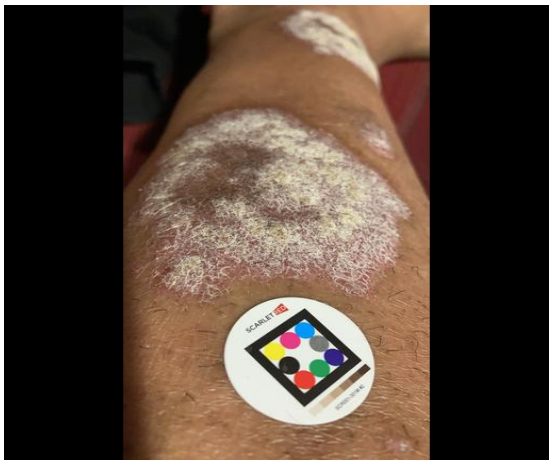
which are freely available to label data for semantic segmentation such as Labelme [20] and Label Studio[21]. Resulting mask in Labelme is a JSON file which represents polygon co-ordinates for each lesion. Whereas, Label Studio is more flexible. It allows users to store resulting mask as a black and white image representing lesion. When it comes to precisely annotating edges, label studio enables you to zoom in to specific area. More in depth comparison between these tools and why we chose Label Studio is discussed later in this section. With the help of brush tool provided by label studio we annotated images. Since there are experienced domain specialists available at CHDR we decided to divide the process of annotation into iterations. In the first round twenty images were annotated. Subsequently, we discussed annotations with the experienced dermatologists at CHDR to make sure that annotations are accurate. Constant feedback from dermatologists was useful to precisely annotate the object of interest, which in turn would result in more robust model.

Semantic segmentation can be multi-class or single class. This depends on the objects of interest. In this project we will be dealing with binary image segmentation. Figure 4a and figure 4b describes how binary masks are created for each input image. Output of the annotated image is a grey-scale 512 * 512 image. Figure 5 represents the output of the image annotated in label studio.

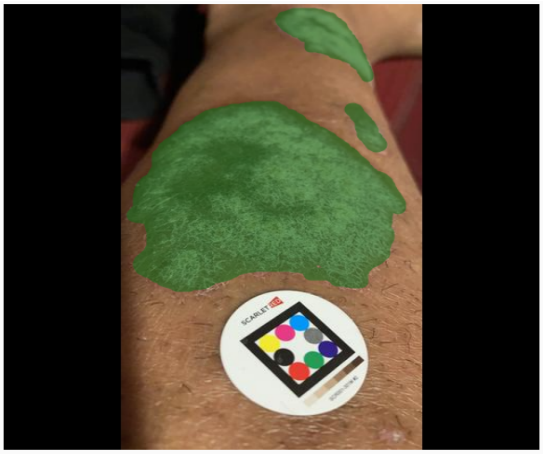
Once annotations for all input images are performed, we create patches of input images and its corresponding masks. Usage of creating patches of the images for segmentation is specifically done in aerial image segmentation[22] tasks. Cutting images and corresponding labels into non-overlapping patches allows deep learning model to learn features closely[22]. With respect to this project each input image(512 * 512 * 3) and mask of size (512 * 512 * 1) will be broken down into four non-overlapping patches. Each patch is of size 256 * 256. Some benefits of using patches for training are:

1. Breaking input image and masks into smaller non-overlapping patches enables network to learn features of the lesion closely.
2. Earlier when developing a U-Net model for 512 * 512 images approximated parameters for training would be 7 million. When network is reduced to 256 * 256 this size comes down 1.7 million. This would reduce required computational power and GPU.
3. Smaller network takes less time to train. This would allow us to tweak certain parameters more frequently in order to perform some experiments. This enables us to come up with best parameter setting required for the model to perform at its peak.
4. When we break image into patches, some patches will be empty. Meaning in some patches there won't be any lesion. This indirectly enables network to learn what is "not a psoriasis".

Figure 6a and figure 6b is an example representing patches of input image and its corresponding mask. Each non-overlapping patch of an image is named by binary digits i.e. [image1_00, image1_01, image1_10, image_11]. Binary representation allows us to keep the track of the mask and its corresponding image and to avoid any confusion when dividing data for training. The whole pre-processing of the data is carried out on the virtual machine where data is stored. This is because the freely available version of the label-studio is compatible with local data and does not provide cloud connectivity. Images and their corresponding masks are then moved to databricks azure platform where we will apply deep learning models for further analysis.



(a) Raw image



(b) Annotated image in label studio

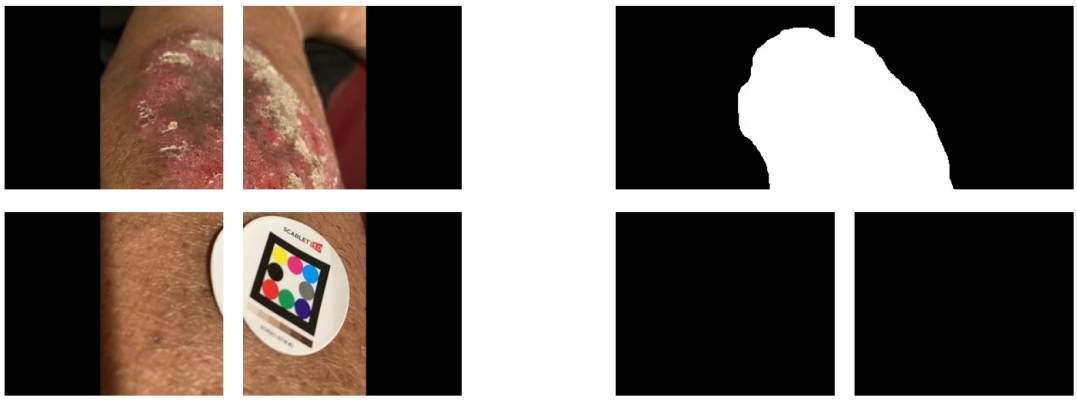
Figure 4: Label Studio Sample

Figure 5: Annotated image output

3.3 Proposed Models

3.3.1 U-Net

U-Net is a convolutional neural network which was specifically designed for image segmentation tasks in 2015 by Pascal Kaiser et al.[23]. Classical convolutional neural networks are used for classification tasks where output of an image is a single class label. In order to converge such network to a reasonable performance we need to train it for significant amount of data. That is not the case in segmentation task, particularly when talking about bio-medical data. Generally there are not enough images available for training in this domain. Localization and quantification of the object of interest is our goal when it comes to applying segmentation on bio-medical images. U-Net is designed elegantly for such tasks. Since it is considered that we deal with images in bio-medical domain, the network is designed to work with very few images. Moreover, it is expected to yield more precise segmentation output. Each label is assigned to a pixel of the feature map which makes it possible to output such precise segmentation result. Figure 7 demonstrates overall U-Net architecture for image size [256 * 256 * 3] as the input which was proposed by by Pascal Kaiser et al.[23].



(a) Input image patches (b) Input image masks

Figure 6: Input image and mask example

Theoretically entire architecture of U-Net can be divided into encoder and decoder blocks. From figure 7 left(contracting part) is the encoder network where as right (expansive part) is the decoder network. To simplify, encoder network encoders input image and decoder decodes encoded image to its segmented mask. It is important to mention that, in the expansive or upsampling part of the network, large number of feature channels have been used. This enables model to propagate weights to higher layers. As a consequence, upsampling part exacts mirrored image of contracting part. This makes network looks like U shaped and hence called U-Net. This symmetry in encoder and decoder part makes U-Net more robust when segmenting crucial edges and borders of an image. This strategy is very important and has been proved revolutionary when it comes to segmenting large images [23].

U-Net relies heavily on image augmentation because it is designed for small number of datasets. This is not the case in this project since we have divided images into patches which makes our total size of the data multiplied by four. In this case it is 1600 images. Output image is one channel grey-scale representation of the skin lesion. Each block of the network contains two convolutional layers with kernel of size 3 * 3 and MaxPooling layer with kernel of size 2 * 2. This will help to reduce features(by 50%) at each layer. MaxPooling is only used in the encoder block when we are transforming the image to a single vector to be used by decoder block.

During the upsampling process we use concatenation at each layer with its corresponding counterpart. For example, when the network is building up in the decoder part at the layer (128 * 128) we will concatenate it with the (128 * 128) block of encoder. In the work proposed by Pascal Kaiser et al. [23], larger input tiles were preferred over large batch size. This will minimize the overhead and maximize the GPU. Smaller batch size may make gradient descent less accurate at each step and will take longer than expected but model will converge. But choosing larger batches will harm the network because of its inability to generalize. In our case since we have enough data samples for each patient, we prefer the first approach. That is for each epoch we will be training network on batch size of 16. Considering total training size of 1440 images with the batch size of 16 for 100 epochs, we will have approximately 90 iterations of back propagation at each epoch.

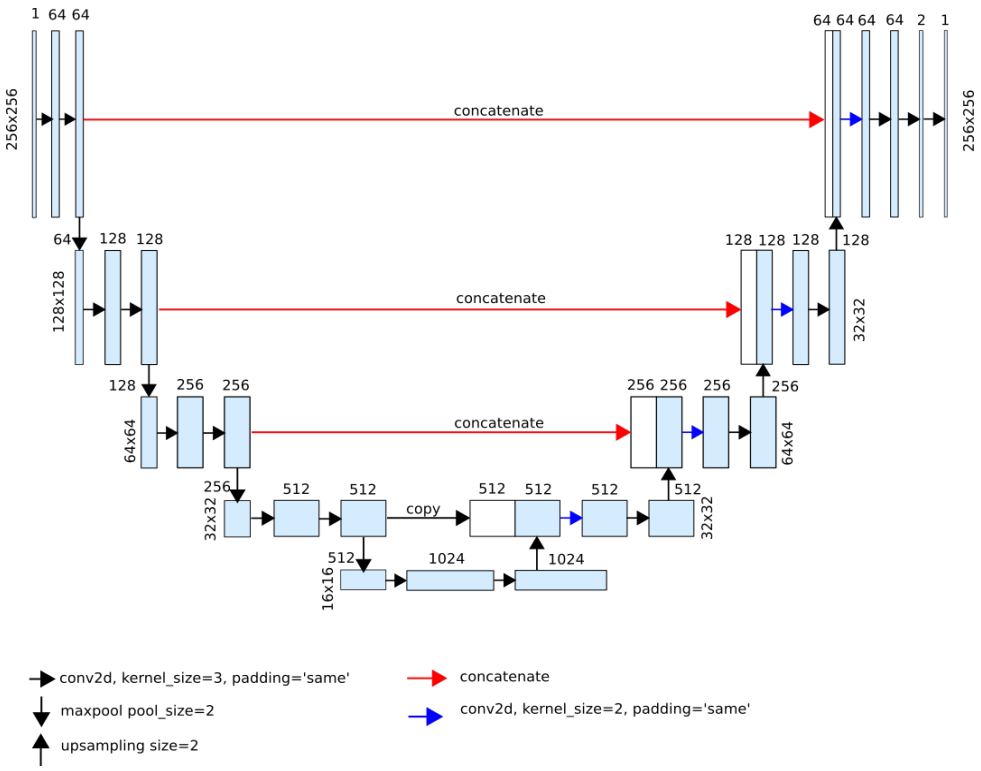


Figure 7: U-Net Architecture

At each block except the final output block, ReLU activation function is used. At the output of the network, sigmoid activation is applied. The ReLU is defined in the equation 5.

$$RELU(X) = max(0, x) \tag{5}$$

This is one of the most used activation function in convolutional neural network. As illustrated in figure 8 the function is half rectified from the bottom and is equal to X when it is above zero. That is derivative of the function is zero when the input to ReLU is negative. Whereas derivative is one when input to activation function is positive. When we are training with a reasonably sized batch, in this case 16, there will be some data points giving positive values at any given node. This avoids vanishing gradients problem faced in other activation functions and allows model to converge.

Having ReLU in combination with sigmoid certainly helps us in this project where, in the binary mask, we have only two values 0's and 1's. Meaning, sigmoid's final outcome is in terms of probabilities. That is whether a particular pixel belongs to a lesion or not. However, it is worth trying Leaky ReLU if ReLU does not produce satisfactory results. In this case its not necessary since results yielded by ReLU are satisfactory.

In the U-Net proposed by Pascal Kaiser et al. [23] authors have used softmax as a final activation function. In this case however, sigmoid proves to be more satisfactory. The curve of the sigmoid function is more like s-shaped line as shown in the figure 10. This shows us that, in case the input to function is either very large positive number or very large negative number, output will always be between zero and one. In the context of this project this allows us to quantify pixels. That is, if output of the function is above certain threshold, we take it as a white pixel in the segmented mask. Another reason to use non-linear function such as sigmoid is that it performs well on complex

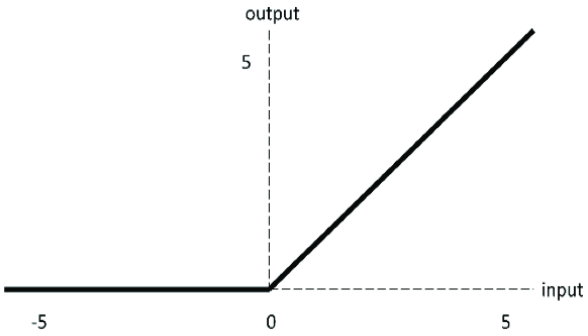


Figure 8: ReLU Activation

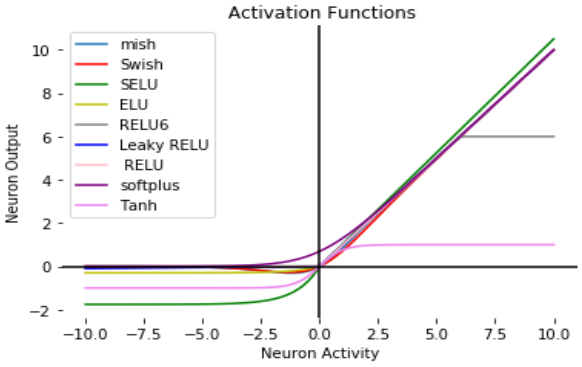


Figure 9: Activation Functions

decision boundaries. For example in terms of skin lesions, we always have unpredictable boundaries and linear function such as softmax will not be able to catch these well. Whereas if we use sigmoid function, because of its nature, the neural network will be able to find non-linearly separable boundaries. The sigmoid function is defined in the equation 6.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{6}$$

The optimizer for the model is "adam". The main intuition behind using adam is that it enables us to slow down the process of gradient descent in order to avoid finding local minima always. One of the drawback of the adam is that it is computationally heavy. Since we have enough resources available in terms of cloud storage and GPU it is possible to use adam optimizer. And since its a binary segmentation problem final loss function is binary cross entropy.

3.3.2 FCN-8

Fully Connected Convolutional network was proposed by Piramanayagam et al.[24] in order to work on aerial or satellite image classification. Image classification here focuses on pixel wise classification which makes FCN-8 network also suitable for segmentation tasks. It is one of the very few networks which is suitable for Object detection, classification and image segmentation tasks. Figure 11 illustrates network architecture which was proposed by proposed by Piramanayagam et al.[24]. For the simplification lets consider left block as encoder block and right one as a decoder block of the network. In encoder block, for each layer we have two convolution operations with the filter of

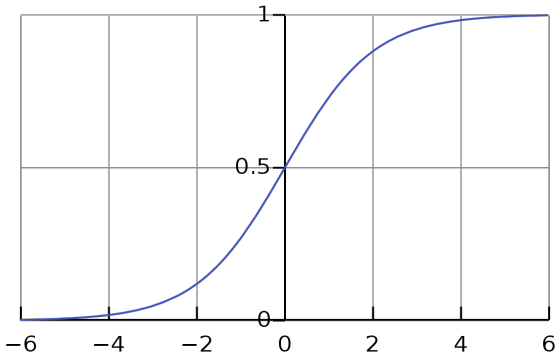


Figure 10: Sigmoid Function

size (3 * 3) followed by MaxPooling operation with the filter of size (2 * 2). The encoder block, more or less resembles VGG16 convolution network. Conventionally, when it comes to classification, each image is downsized and passed through series of convolution and fully connected layers. But when we are discussing its application in the segmentation domain these convolutional layers are converted into (1 * 1) layers in the decoder block. Application of the MaxPooling in the encoder block allows output not to be a label but a similar dimensional image as input. This shows that fully connected network can be trained on pixel-to-pixel for semantic segmentation tasks[25].

Exponential Linear Unit also known as elu activation is used for each block of the network except for the final one. Elu is used in order to fix some of the issues regarding ReLU activation function and it is sometimes used as an alternative to ReLU.

$$ELU(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha(e^x - 1) & \text{if } x < 0 \end{cases} \tag{7}$$

Equation 7 refers to the ELU activation, here αgenerally has a value between 0.1 and 0.3. If the input to the activation function is greater than zero, it is the same as ReLU. Meaning, result will be y-value corresponding to its x counterpart. But when input to the activation function is less than zero, y-value gets dependent on x and alpha. We can tweak this parameter but generally the recommended value is 0.1. However, this makes Elu a littler more computationally heavy compared to ReLU.

Deep learning networks are more prone to overfitting. For this project we are using annotated patches this increases the possibility that model might overfit. Chances of overfitting are also higher on the larger networks, with the possibility of learning every smaller detail of the weights and model noise. This simply happens because model learns statistical noise which results in great performance on the training data. When we test the model on test set we realize that there is difference in the performance. Implementation of dropout at each layer solves the problem of overfitting. In the U-Net discussed in section 3.3.1, each layer has a dropout of 0.1. In the FCN-8 dropout with 0.7 is implemented in the last block of encoder network.

The FCN-8 network has the same optimizer as U-Net 3.3.1. This allows us to slow down the process of gradient descent. Since it is a binary segmentation, the model loss is binary cross entropy

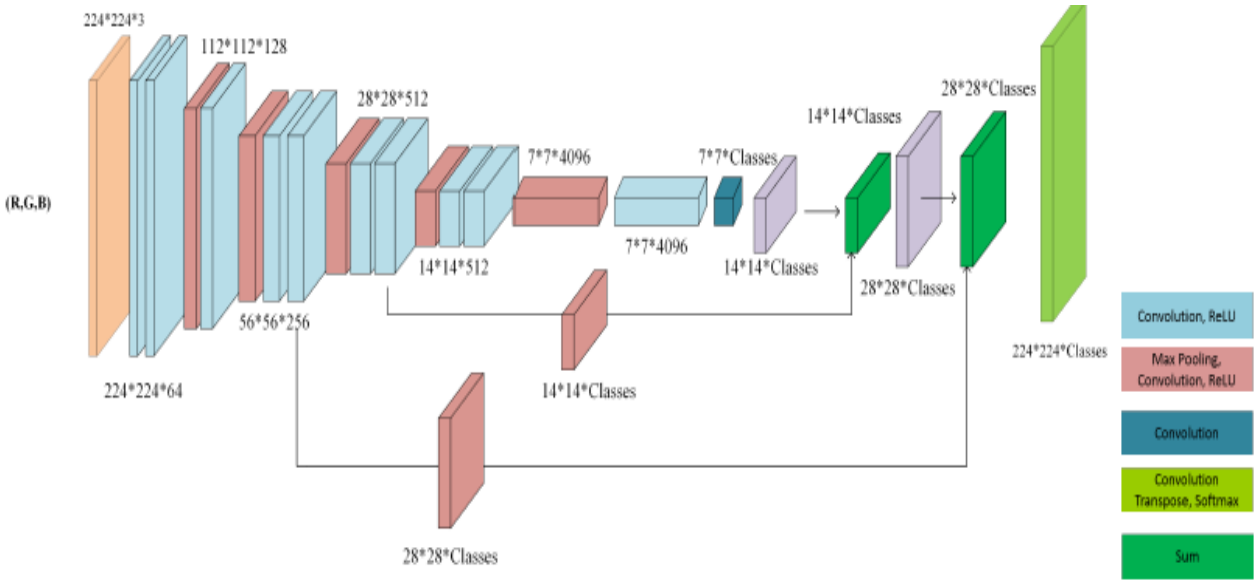


Figure 11: FCN-8 Architecture

the same as U-Net.

3.3.3 Model Training and Evaluation

Strategy for training both FCN8 and U-Net models is to divide entire dataset in training and testing split. Model is trained on the training set. Training and testing split is of 9:1 ratio. In the training set there are total 1440 images and in test set there are 160 images. Traditionally for the deep learning tasks, measures such as model accuracy and model loss provide enough information about the trained model. When it comes to image segmentation along with model accuracy and loss, we are also interested in the position of the pixels classified as lesion. To measure this, we use Intersection over Union or Jaccard index[26].

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{8}$$

This simply describes the extent of overlap between two regions. Ideal IOU score of the model would be 1. Equation 8 defines how we can calculate this index. J is the Jaccard distance. A and B are two sets. In our case they are ground truth and predicted masks. Model is evaluated on test set based on it's IOU. Final image and mask array is a 4-dimensional stacked array of dimension (images, width, height, channels). Once we have split data into training and testing set, it is fundamental to make sure that ground truth masks are loaded with their corresponding images. For this, we performed some sanity checks and plotted random images with its masks. Figure 12 demonstrates one such attempt.

Estimation of the size of lesion is dependent upon the size of tag. It is the only physical entity of which dimensions are clear. That is diameter of the tag is 2.5cm. This thesis includes two deep learning models.

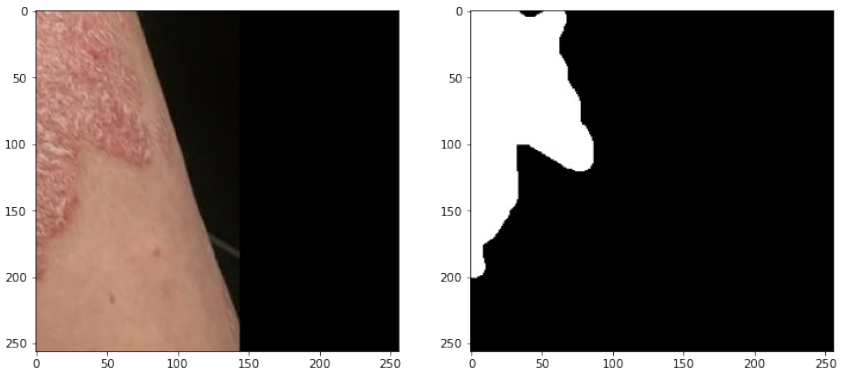


Figure 12: Sanity check before Model training

1. One which extracts the lesion from an image
2. Second which extracts the tag from same image.

Figure13 illustrates the training procedure followed up until now. This involves annotation, creating patches and experimentation of U-Net and FCN-8 for lesion and tags. Extraction of tags is necessary as based on it, the size of lesion is computed. The procedure to train U-Net for the tag is similar to what was followed for lesion segmentation. Except for the input, that is annotated tag masks are used here instead of lesions.

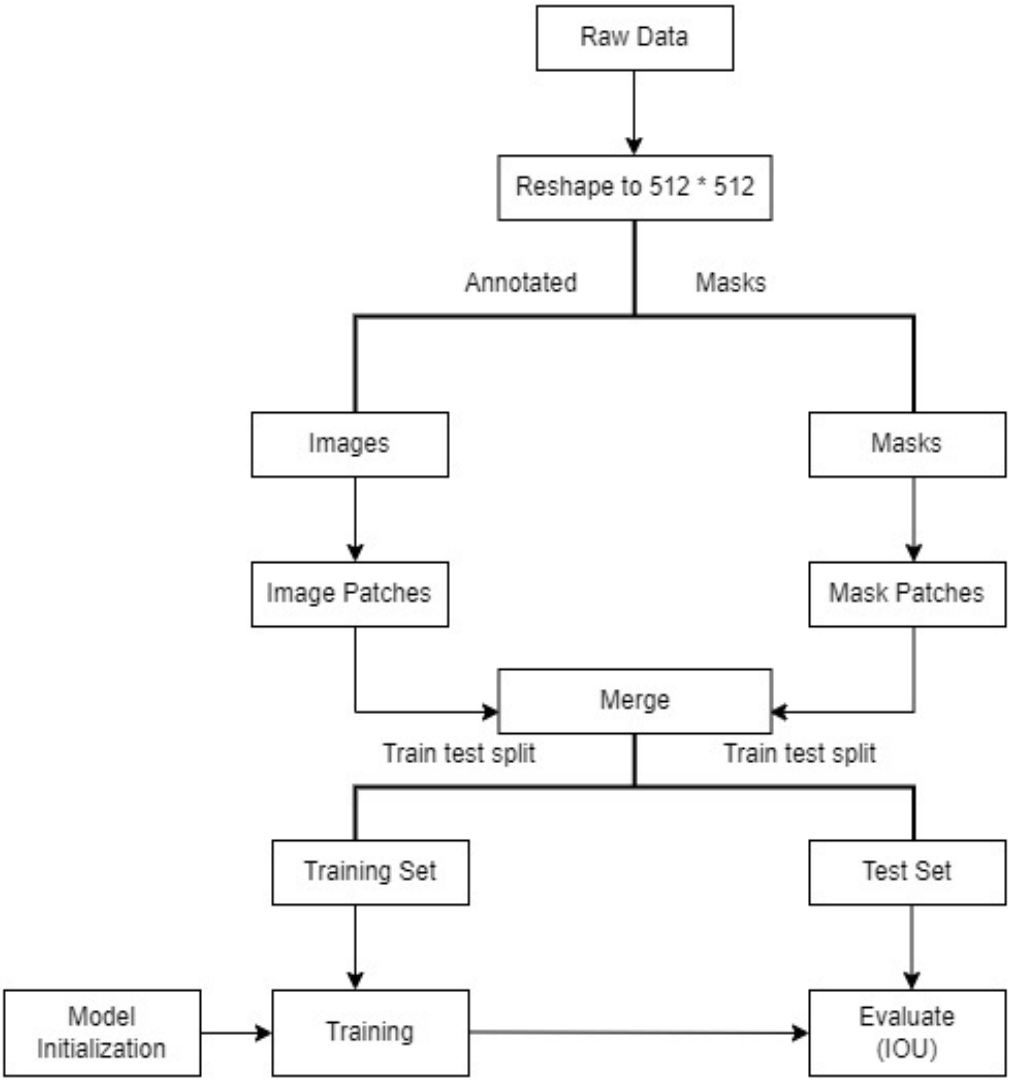


Figure 13: Model Experimentation(U-Net and FCN8)

Model	IOU
FCN8(50 Epochs)	62.5
U-Net(50 Epochs)	84.5
FCN8(100 Epochs)	77.0
U-Net(100 Epochs)	**87.2**

Table 2: Model Evaluation with IOU Score

4 Results

Table 3 describes all crucial parameters used for training FCN8 and U-net networks. Creating patches of images turns out to be handy in reducing required computational power and runs model faster. All of these experiments are performed on Microsoft Azure Databricks(runtime 10.4) platform. From figure 15a and figure 16a we can notice that both models perform significantly well if we consider accuracy of the network. Since we are using elu activation in all blocks, FCN8 network takes significantly more time to execute. In terms of accuracy U-Net is a little better than FCN8. As we can notice since both models are somewhat equal in performance. It is important that we focus on IOU for final performance evaluation. Table 2 describes IOU scores produced by all models on the test set. U-Net is much better when it comes to IOU. This is clearly because U-Net is designed in such a way that it performs well when it comes to predicting on edges and curves. Figure 18a and figure 18b justifies the better IOU score of U-Net. IOU scored achieved by FCN8 network still does not tells us the entire story of why FCN8 is not preferred for this task. Visualization and evaluation of the model on the test set turns out to be a key. Implementation of the concatenation for each block and mirror replicated decoder block is vital in U-Net. We can notice that FCN8 architecture in figure 11 does not have robust decoder block as U-Net and also there is no concatenation performed. This serves as a bridge between encode and decoder.

Each time when the trained model is called for the prediction, each single image is broken into four patches as demonstrated in figure 6a. Model predicts on each of these patches and these patches are merged together into original shape of (512 * 512). This is the final output representing lesion.

Tag is one the entity which is constant in all images and comparatively easier to retrieve. Figure 17a and figure 17b demonstrates U-Net model performance on the tag data. We trained this network for only 20 epochs because it converged earlier as extraction of tags turns out to be a easier. We know the size of this tag which is 2.5 cm in diameter. First we calculate the pixel ratio of extracted lesion and tag. Consider that if white pixel percentage in the lesion is 23% and white pixel percentage of the tag is 11% then this implies that tag to lesion ratio is (2.2:1). If we multiply actual size of the tag that is area of circle with diameter 2.5cm by 2.1, we will get the size of lesion in centimeter square.

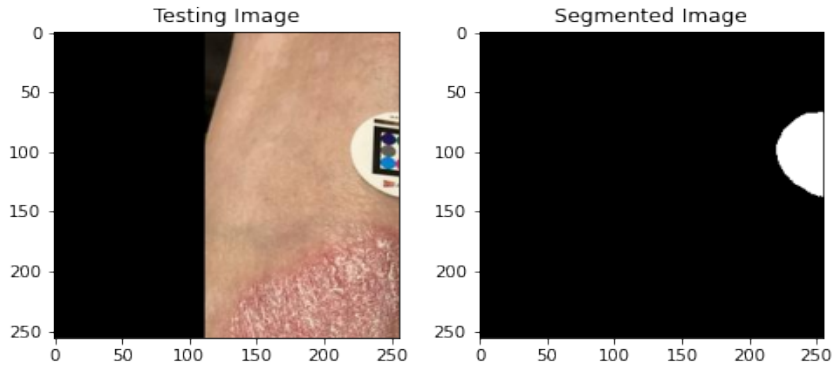


Figure 14: U-Net Prediction on Tag

Model	Epochs	Activation	Dropout	Loss	Accuracy	IOU	Time(hrs)
U-Net	50	ReLU	0.1	3	96.4	84.5	3.2
FCN8	50	ELU	0.7	4.5	98.02	62.5	6.8
U-Net	100	ReLU	0.2	0.2	98.00	87.2	6.5
FCN8	100	ELU	0.7	0.3	98.02	77.00	12.5

Table 3: U-Net and FCN8 Model Performance with important parameters of the network

5 Conclusion and Future work

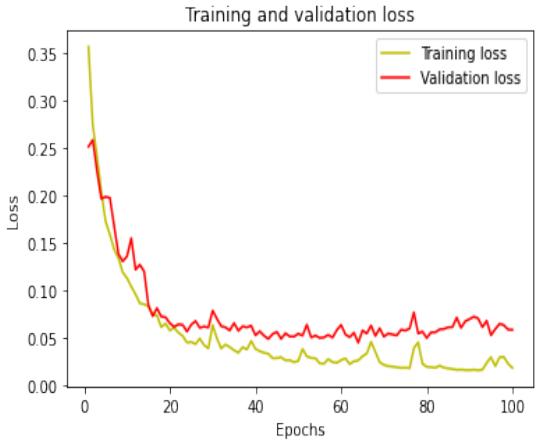
Precise segmentation of psoriasis for accurate identification of the skin lesion is of utmost significance. Extracted lesion and its size highly affects further diagnosis involved in psoriasis. Segmentation of the lesion enables dermatologists to have a closer look at the lesion without having to physically intervene the process. Based on results discussed in the previous section4 it is clear that for this task of segmenting psoriasis lesion, U-Net is better. We start by reshaping images sent by patients. Further, we break image into a patch and predict on these patches.

Based on extracted lesion and tag, We compute the size of lesion and tag present in the image. This includes the pipeline for the entire process, starting with reshaping of patients image until we calculate lesion size. Figure 19 illustrates the whole process of this project. Experimental results indicate that proposed method can achieve a very high accuracy of 98.00% with IOU score of 87.2%.

Extracted lesion and its size opens door for the classification problem of the psoriasis. In the clinical trials performed at CHDR, dermatologists extract lesion first. Subsequently, depending on the color features and texture features, they classify the severity of the psoriasis. Table 1 discussed earlier in the section 1.1 describes severity index. Extracted lesion size and severity index is tracked by dermatologists over the period of time. Expectation here is that both would eventually reduce to zero. For this each patient is asked to capture their lesion daily/weekly depending on the severity index.

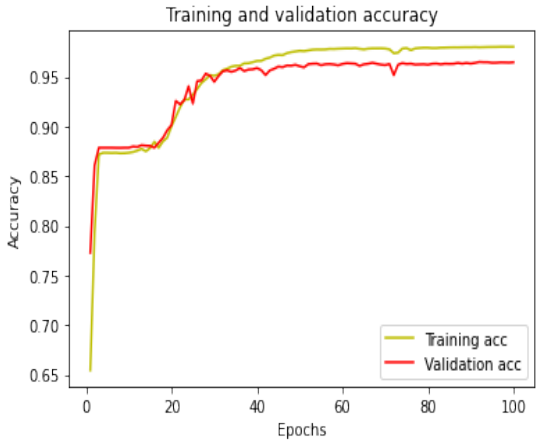
This approach of extracting lesions suggests that, when gathering trial@home data, patients should have some standard procedures to follow. For example from image 20a and image 20b it can be seen that unwanted background noise can cause disruption in size computation of the lesion. Well

(a) U-Net Lesion Model Accuracy

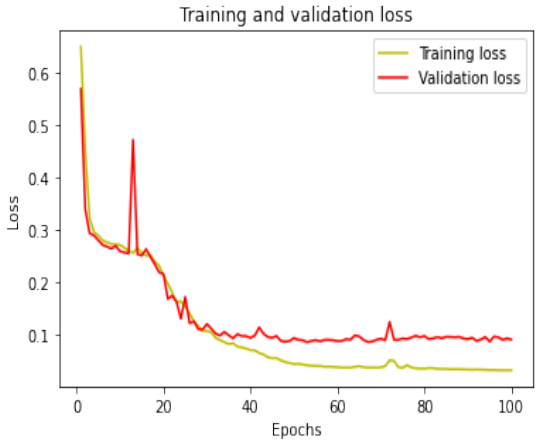


(b) U-Net Lesion Model Loss

Figure 15: U-Net Lesion Model Performance



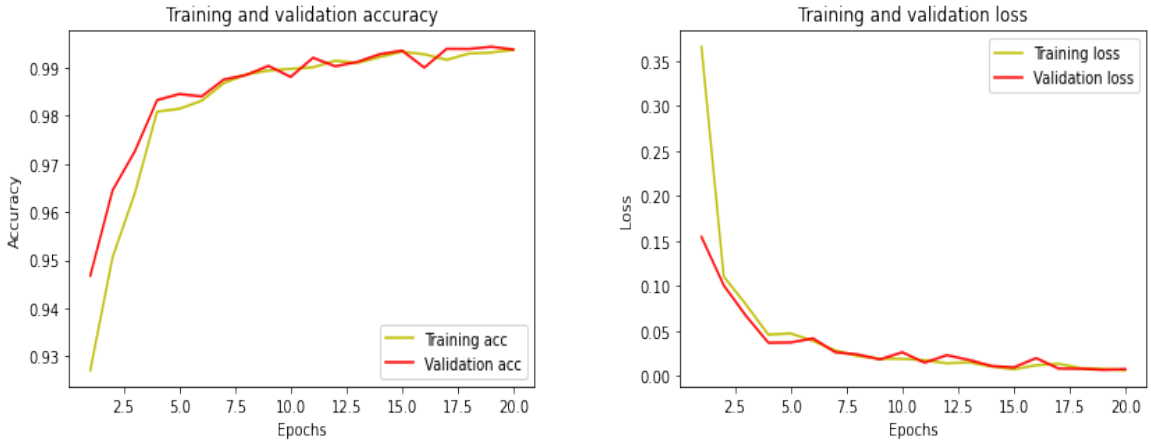
(a) FCN-8 Lesion Model Accuracy



(b) FCN-8 Lesion Model Loss

Figure 16: FCN-8 Lesion Model Performance

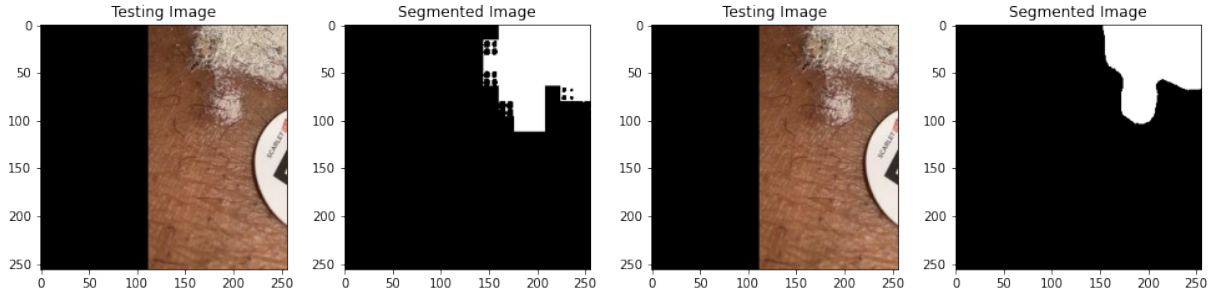
formed standard procedure for trials at home would ensure that collected data is of good merit. For example, camera angle and distance of the camera can affect the size of the lesion. Meaning we can notice that there is a difference in lesion size of the same patient in two different images. Predefined distance and placement of smartphone can simply cover this limitation. Another limitation of this approach is that all of ground truths used for training were not annotated by domain expert. In order to have more accurate, precise and robust hypothesis this should be performed by experienced clinical scientists or dermatologists.



(a) U-Net Tag Model Accuracy

(b) U-Net tag Model Loss

Figure 17: U-Net Tag Model Performance



(a) FCN8 Model

(b) U-Net Model

Figure 18: Prediction sample on Test image

References

[1] David R. Yates, Christophe Vaessen, and Morgan Roupret. From leonardo to da vinci: the history of robot-assisted surgery in urology. *BJU International*, 108(11):1708–1713, 2011.

[2] Mohammad Jafari, Nader Karimi, Ebrahim Nasr Esfahani, Shadrokh Samavi, S.M.Reza Soroushmehr, Kevin Ward, and Kayvan Najarian. Skin lesion segmentation in clinical images using deep learning. 12 2016.

[3] Li Deng and Dong Yu. Deep learning: Methods and applications. *Foundations and Trends in Signal Processing*, 7, 01 2013.

[4] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, jan 2015.

[5] P.M. McHENRY and V.R. DOHERTY. Psoriasis: an audit of patients' views on the disease and its treatment. *British Journal of Dermatology*, 127(1):13–13, 1992.

[6] Irmina Michalek, Belinda Loring, and Swen John. *WHO Global report on psoriasis*. 02 2016.

[7] Robert M Haralick and Linda G Shapiro. Image segmentation techniques. *Computer vision, graphics, and image processing*, 29(1):100–132, 1985.

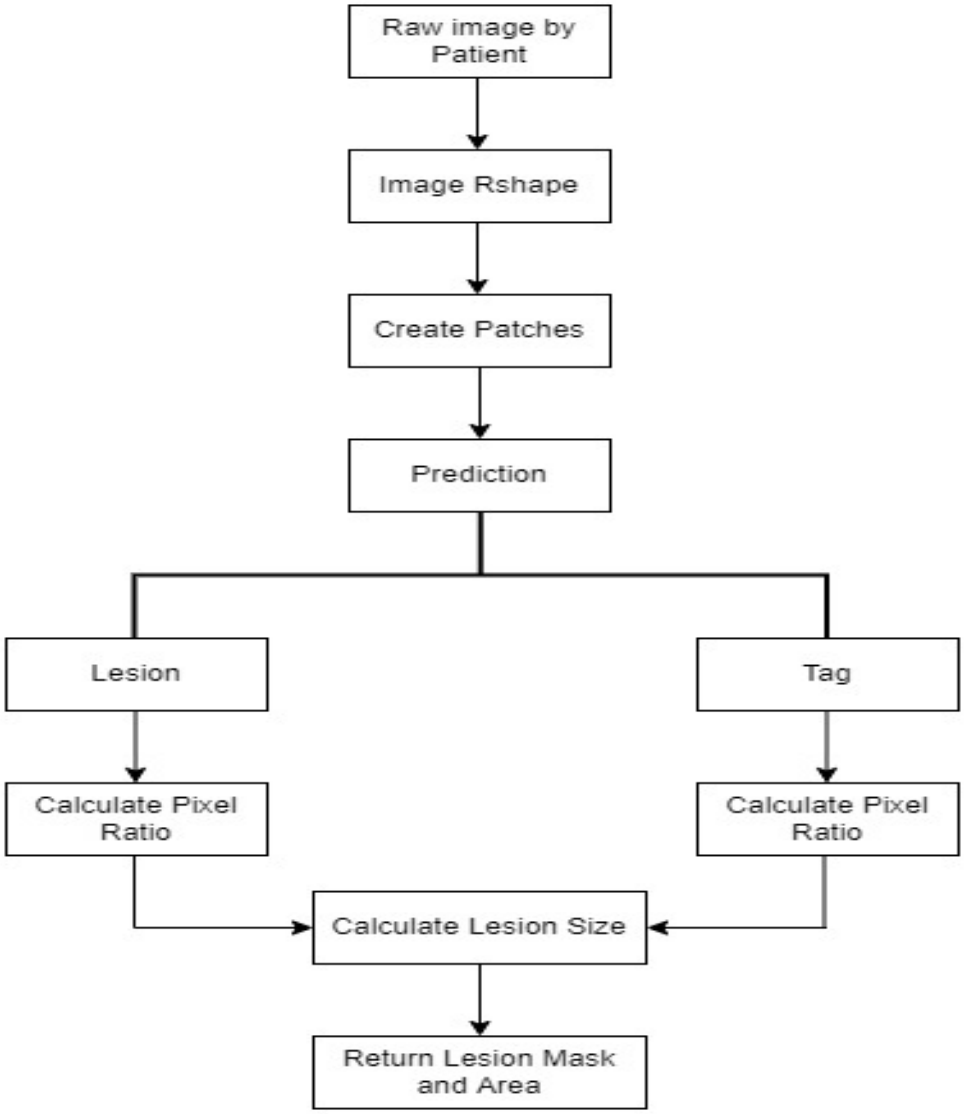
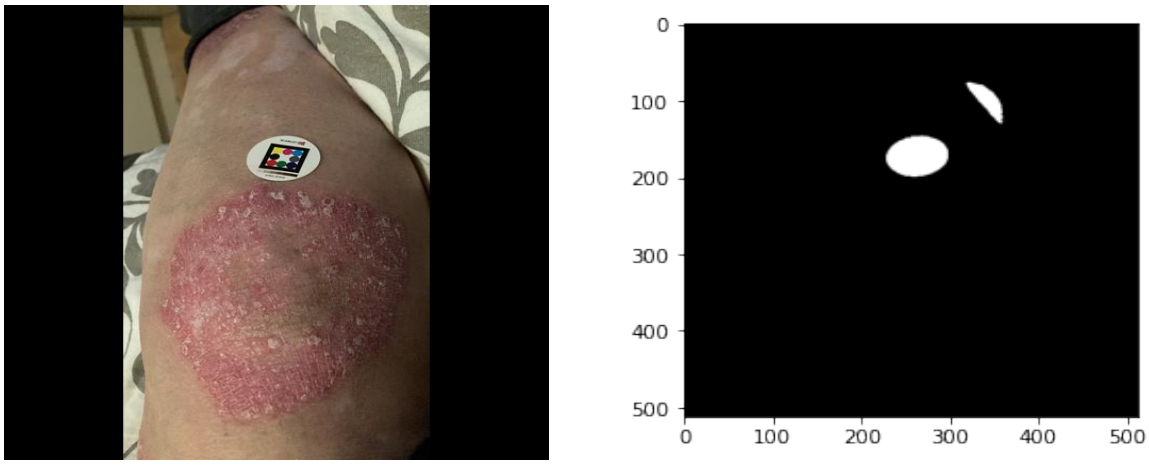


Figure 19: Project Pipeline



(a) Input Image (b) False Positive Tag Prediction

Figure 20: False Positive Case

[8] Shervin Minaee, Yuri Y Boykov, Fatih Porikli, Antonio J Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 2021.

[9] Samuel T. Hwang, Tamar Nijsten, and James T. Elder. Recent highlights in psoriasis research. *Journal of Investigative Dermatology*, 137(3):550–556, 2017.

[10] Swarnendu Ghosh, Nibaran Das, Ishita Das, and Ujjwal Maulik. Understanding deep learning techniques for image segmentation, 2019.

[11] Christophe Rosenberger, Sébastien Chabrier, Hélène Laurent, and Bruno Emile. Unsupervised and supervised image segmentation evaluation. *Advances in Image and Video Segmentation, IGI Global*, pages 365–393, 01 2006.

[12] Amine M. Bensaid, Lawrence O. Hall, James C. Bezdek, and Laurence P. Clarke. Partially supervised clustering for image segmentation. *Pattern Recognition*, 29(5):859–871, 1996.

[13] Robin Strudel, Ivan Laptev, and Cordelia Schmid. Weakly-supervised segmentation of referring expressions, 2022.

[14] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[15] Pedro O. Pinheiro, Tsung-Yi Lin, Ronan Collobert, and Piotr Dollàr. Learning to refine object segments, 2016.

[16] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation, 2015.

[17] R. Sumithra, Mahamad Suhil, and D.S. Guru. Segmentation and classification of skin lesions for disease diagnosis. *Procedia Computer Science*, 45:76–85, 2015.

[18] Martin Hofmann, Philipp Tiefenbacher, and Gerhard Rigoll. Background segmentation with feedback: The pixel-based adaptive segmenter. pages 38–43, 2012.

[19] M.H. Jafari, N. Karimi, E. Nasr-Esfahani, S. Samavi, S.M.R. Soroushmehr, K. Ward, and K. Najarian. Skin lesion segmentation in clinical images using deep learning. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 337–342, 2016.

[20] Antonio Torralba, Bryan Russell, and Jenny Yuen. Labelme: Online image annotation and applications. *Proceedings of the IEEE*, 98:1467 – 1484, 09 2010.

[21] Marco Stranisci, Enrico Mensa, Ousmane Diakite, Daniele Radicioni, and Rossana Damiano. Guidelines and a corpus for extracting biographical events. 06 2022.

[22] Pascal Kaiser, Jan Dirk Wegner, Aurelien Lucchi, Martin Jaggi, Thomas Hofmann, and Konrad Schindler. Learning aerial image segmentation from online maps. *IEEE Transactions on Geoscience and Remote Sensing*, 55(11):6054–6068, nov 2017.

[23] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. 2015.

[24] Sankaranarayanan Piramanayagam, Eli Saber, Wade Schwartzkopf, and Frederick Koehler. Supervised classification of multisensor remotely sensed images using a deep learning framework. *Remote Sensing*, 10:1429, 09 2018.

[25] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. 2014.

[26] John Hancock. *Jaccard Distance (Jaccard Index, Jaccard Similarity Coefficient)*. 10 2004.