uc3m | Universidad Carlos III de Madrid

Master Degree in Information Health Engineering
Academic Year 2020-2021

Master Thesis

“Segmentation of local structures of pigmented skin lesions in dermoscopy images”

María Beatriz Loureiro Casalderrey

Fernando Díaz de María

Madrid, 12/09/2021

Segmentation of local structures of pigmented skin lesions in dermoscopy images

María B. Loureiro Casalderrey[1] and Fernando Díaz de María[2]

[1,2] Department of Signal Theory and Communications, University Carlos III, Madrid, Spain

E-mail[1]: 100443638@alumnos.uc3m.es
E-mail[2]: fdiaz@ing.uc3m.es

Abstract

The incidence of melanoma keeps rising every year. This type of cancer, when not diagnosed on time, is often lethal. For this reason, there is a big interest in moving its diagnosis toward early stages. The usual procedure to detect melanoma is by visual inspection of a highly trained specialist. However, it is subjective, time-consuming, and susceptible to errors. A solution to avoid these problems is the use of automated, computer-aided diagnosis. Here, we evaluate the performance of five proposed models for the automatic segmentation of five clinically meaningful visual skin lesion patterns associated with melanoma in dermoscopic images. The models are the result of fine-tuning the DeepLabV3 architecture using 5 different methods. The models were trained on the database from the 2018 Lesion Attribute Detection challenge from the International Skin Imaging Collaboration (ISIC). Our results indicate that the use of weights to give more importance to the less represented classes improves the performance of the model on their segmentation. Furthermore, reducing the dilation rate of the convolution layers from the Atrous Spatial Pyramid Pooling (ASPP) module of the network improves the performance of the model due to the small size of the skin lesion attributes to be segmented. Also, we identified the regions of the images that led to a bad performance of the model and analyzed the results in terms of soft labeling. Lastly, we proposed some ideas for the improvement of the automatic segmentation of skin lesion attributes in future works.

Keywords: semantic segmentation, dermoscopy, skin cancer, melanoma, DeepLabV3

1. Introduction

Melanoma has one of the fastest growing incidence rates of any cancer worldwide. Although it represents a small percentage of skin cancer, it is the most aggressive and responsible for the majority of skin cancer deaths (Dinnes, 2018). Its care becomes increasingly complex and expensive in advanced stages as it is highly metastatic, i.e. it has the capacity to invade other tissues and organs thus incrementing the caliber of malignancy. When diagnosed early, before the metastasis stage, melanoma can be curable, but when it is not detected in time it is often incurable and leads to the death of the patient. For these reasons, research efforts are being undertaken worldwide to move its diagnosis toward earlier stages (Forsea, 2020) (Khan, 2021) .

Dermoscopy, also known as dermatoscopy, epiluminescence microscopy, incident light microscopy, or skin surface microscopy, is a non-invasive, in vivo technique used for the examination of pigmented skin lesions. This technique enables the magnification of skin lesions and the visualization of subsurface skin structures that are usually not visible to the naked eye (Goyal, 2019) (Sonthalia, 2021). Dermoscopy is one of the most widely used imaging techniques in dermatology and since its incorporation into the common medical practice, the diagnosis rate has been improved (Kittler, 2002).

In a clinical examination, the usual procedure to differentiate melanoma from other kinds of skin lesions is visual inspection of images following several methods, such as ‘ABCD rules’, ‘Menzies method’, and ‘7-point checklist’ (Khan, 2021). However, this process requires the work of highly trained experts and is subjective, time-consuming, and not free from errors. A solution to these problems is the use of computer-aided diagnostic (CAD) systems, which consume less time and are objective, more robust, and more accurate than the previously mentioned techniques (He, 2019).

In the last two decades, CAD systems have been gradually incorporated into the clinical practice supporting medical experts in their decisions. One possible utility of CAD systems in the problem of melanoma diagnosis is the direct classification of images between melanocytic and non-melanocytic lesions. Nevertheless, dermatologists complain that these kinds of algorithms are not designed to work as a support tool as they do not provide any comprehensive information to justify the diagnosis. Another possibility is the development of systems that focus on the identification of specific information of special interest for dermatologists. They provide the doctor with an analysis of the different structures found on the lesion and, with this information, he or she can improve his or her decisions (López-Labraca, 2018). This work lies within this second approach.

As stated on its website, “The International Skin Imaging Collaboration (ISIC) is an academia and industry partnership designed to facilitate the application of digital skin imaging to help reduce melanoma mortality” (isic-archive.com, n.d.). One of its main objectives is to boost the development of diagnostic artificial intelligence algorithms. To that end, ISIC has created and is continuously improving a large, open-source, pubic access database of skin images. Additionally, ISIC organizes challenges that motivate researchers to develop and improve such algorithms with the result of hundreds of publications in the medical and scientific fields every year (isic-archive.com, n.d.).

In this article, we propose an alternative solution to common segmentation networks for the particular problem stated in one of these challenges, ‘Lesion Attribute Detection’, from 2018 (ISIC 2018 Challenge - Task 2: Lesion Attribute Detection , n.d.). The goal is to develop an algorithm that automatically predicts the location of five relevant types of dermoscopic attributes, established as clinically-meaningful visual skin lesion patterns, in dermoscopic images.

2. Related Work

2.1 Semantic segmentation

Semantic segmentation is the process of assigning a class or a label to each pixel of an image clustering together regions that belong to the same object class (Seyedhosseini, 2015) (Thoma, 2016). In the past, classical machine learning techniques like SVM, Random Forest, or K-means Clustering were used to solve image segmentation problems (A. Murugan, 2019). However, deep learning has proven to work better on this kind of task and convolutional neural networks (CNNs) have gradually replaced the classical methods (Seo, 2020).

A CNN can be adapted to perform semantic segmentation by replacing the last layer of a classification network, which is generally a softmax layer, with a convolutional layer. As CNNs use downsampling layers, semantic segmentation networks include upsampling layers which are used to upscale lower resolution pixel-wise predictions to match the original resolution of the images. For this reason, most of the semantic segmentation architectures, such as FCN-ResNet101, FastFCN, or Unet consist of an encoder followed by a decoder. Other architectures have been designed in a way that spatial resolution is preserved so there is no need for a decoder.

2.2 DeepLabV3

DeepLabv3 is an open-sourced state-of-the-art semantic segmentation model designed by Google (Chen, 2019). The peculiarity of this CNN is that, instead of using the combination of max-pooling and striding at consecutive layers to broaden the receptive field, as in other semantic segmentation networks, it adds dilated convolutions, also know as atrous convolutions or convolutions with holes. It mostly preserves the spatial dimensions and avoids the need

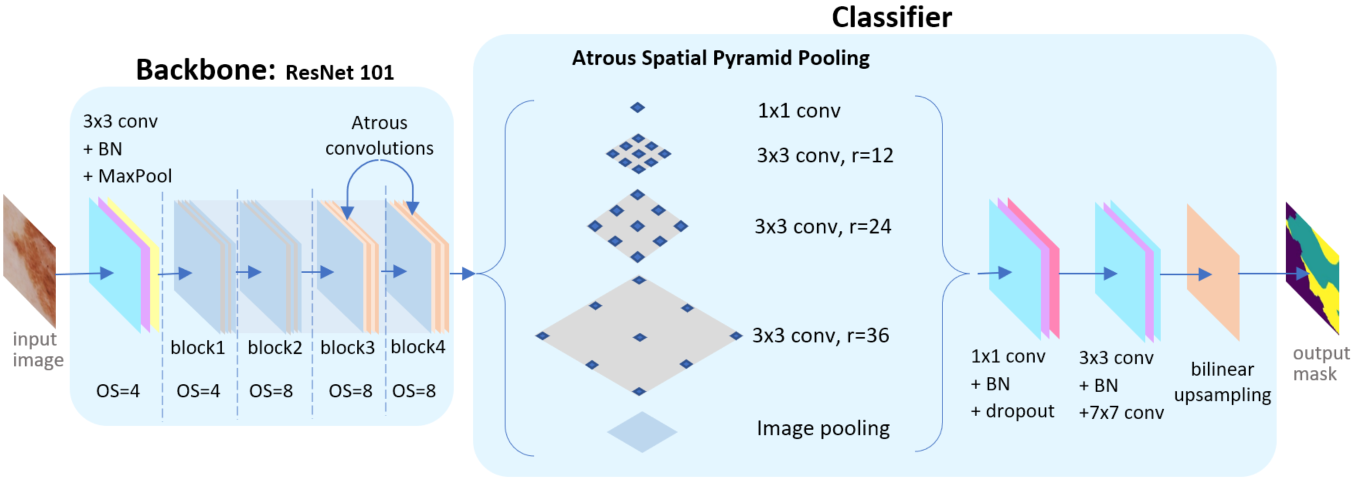

Figure 1. DeeplabV3_ResNet101 architecture. On the left: backbone formed by ResNet 101 modified with atrous convolutions in blocks 3 and 4. The output stride (OS) grows up to 4 after the first group of layers and increases up to 8 after the second block of the backbone. On the right, the classifier containing atrous spatial pyramid Pooling.

for posterior upsampling layers. The architecture of this neural network is illustrated in figure 1.

Figure 2 illustrates dilated convolutions. They introduce a new parameter to convolutional layers, the dilation rate, which specifies the spacing between the values in a kernel. In a 2-D dilated convolution, if the rate is one, a regular convolution is performed. If the rate is greater than one, a convolution with holes, sampling the input every rate pixels in the two dimensions takes place. These layers are located in the last two blocks of the backbone, which operates as principal feature extractor, with various dilation rates. We have used ResNet101 as a backbone.

Furthermore, Atrous Spatial Pyramid Pooling (ASPP) is used so that the network can recognize objects at different scales. The idea of these layers is to apply multiple atrous convolutions with different dilation rates to the feature map and to fuse them together. Each convolution captures objects and useful image context at different scales.

In the backbone, there are a couple of layers with stride 2 which decrease the resolution of the feature maps. The authors of the article 'Rethinking Atrous Convolution for Semantic Image Segmentation'(Chen, 2019), denote *output_stride* the ratio of the spatial resolution of the input image and the final output resolution of the feature maps from the backbone. In this architecture in particular there are three layers with stride 2. This produces an *output_stride* of 8, i.e. the output feature maps are 8 times smaller than the input image.

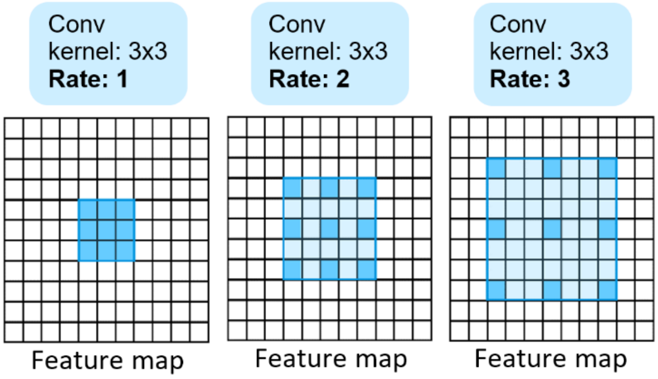

Figure 2. Atrous convolution with kernel size 3 × 3 and different dilation rates. Atrous convolution with rate = 1 corresponds to a standard convolution. The use of large values of atrous rate enables a larger field-of-view for the model.

3. Experiments

3.1 Database

The database consists of 2750 dermatoscopic images with different shapes of common pigmented skin lesions in JPEG format and their correspondent masks. Images and masks are organized in three datasets: train, validation, and test. Masks are composed of pixel-level annotations. Each pixel presents an index from 0 to 6 representing the classes listed below (ISIC Challenge Datasets, n.d.).

0 - Healthy skin. Normal skin that surrounds the lesion.

1 - Others: Skin that is part of the lesion but does not lie within any of the five relevant dermoscopic attributes.

2 - Milia-like cysts: Round, well circumscribed, whitish or yellowish structures (Minagawa, 2017).

3 - Negative pigment network. Areas relatively lighter that draw an apparent grid of a network next to darker areas filling the apparent 'holes' of the network (Pizzichetta, et al., 2013).

4 - Pigment network: Intersecting brown lines forming a reticular pattern that looks like a grid (Kittler, et al., 2016).

5 - Streaks: Lineal pigmented projections at the perimeter of a melanocytic lesion composed of radial streaming (lineal streaks) and pseudopods (bulbous projections) (Kittler, 2016).

6 - Globules: Black, brown, round to oval well-demarcated structures larger than 0.1 mm. They can be regularly or irregularly distributed, aggregated, or located along the periphery of a melanocytic lesion (Kittler, 2016).

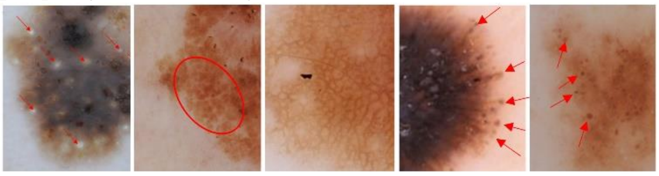

Figure 3. Relevant dermoscopic attributes. From left to right: milia-like cysts, negative pigment network, pigment network, streaks, and globules.

For this study, class 0 was ignored in order to focus on the identification of attributes inside the lesion.

All images and masks were cropped in order to remove as many healthy skin pixels as possible and to maintain all of the pixels belonging to the lesion. Besides, as illustrated in figure 4, all images were resampled reducing the shortest dimension to 256 pixels and the largest dimension to 256 multiplied by the result of dividing the number of pixels from the largest dimension by the number of pixels from the shortest dimension.

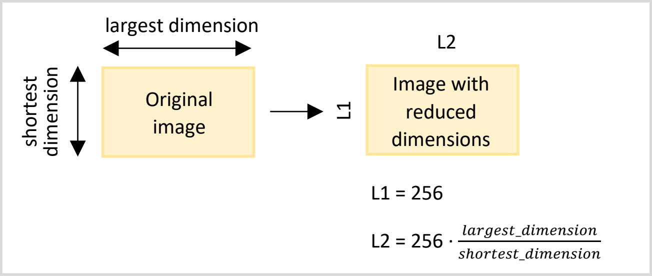
largest dimension
shortest dimension
Original image
L1
L2
Image with reduced dimensions
L1 = 256
L2 = 256 · largest_dimension / shortest_dimension

Figure 4. Resampling of the images.

On training images, a second, random crop of 128 by 128 pixels was performed. On validation images, this second crop was done on the center of the images.

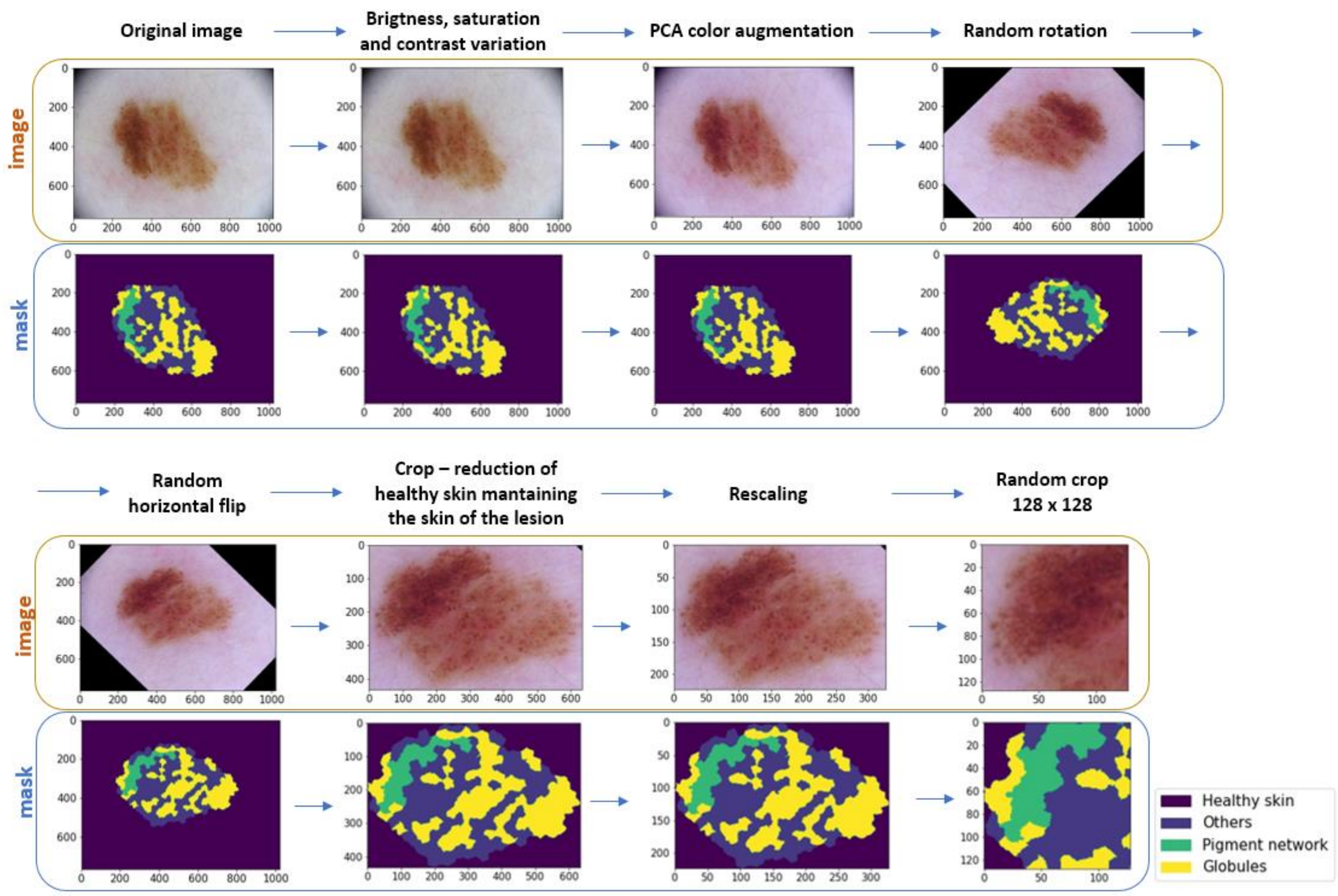

Figure 5. Pipeline for the augmentation and preparation of the training data.

3.2 Data augmentation

Figure 5 illustrates the data augmentation steps applied over the training dataset. Data augmentation was performed in order to acquire data with more variability, reduce the probability of overfitting, and increase the generalization ability of the model.

- *Brightness, saturation, and contrast variation*. The images in the dataset were acquired with a variety of dermatoscopy types from several institutions. We added a variation of brightness, saturation, and contrast to achieve a better generalization from the network for different types of images and make it more robust to images slightly different from the ones found in the training set.

- *PCA Colour Augmentation*. It consists of shifting the intensities of the RGB channels based on which are the most relevant values in the image, which is denoted by the principal components of the colors of the pixels (Bargoti, 2016).

  To achieve this, Principal Components Analysis (PCA) is performed on the set of pixel values of the three color channels of an image. Then, new images are created by adding multiples of the calculated principal components, with magnitudes proportional to the corresponding eigenvalues multiplied by a random variable drawn from a normal distribution $\mathcal{N}(0,0.1)$. Thus, the three channel values from each pixel from a new image are calculated as follows:

$$I_{XY\_new} = [\, I_{XY}^{R}, I_{XY}^{G}, I_{XY}^{B}] + [\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3},][\alpha_1\lambda_1, \alpha_2\lambda_2, \alpha_3\lambda_3]^{T} \quad (1)$$

where I_{XY} represents the pixel in the (x,y) position, p_i and λ_i are the ith eigenvector and eigenvalue of the 3x3 covariance matrix of RGB pixel values respectively, and α_1 is the random variable drawn from the normal distribution (Krizhevsky, 2012).

In this way, as shown in figure 6, image color can be altered in a natural way creating realistic images with colors that could be found in other datasets.

- *Random rotation*. A random rotation between -180 and +180 degrees was implemented, to make the network generalize for attributes located at different locations and with different dispositions

- *Random flip*. A random horizontal flip was introduced for the same reason explained in the previous step.

- *Random crop*. A random crop of 128 by 128 pixels was done once the image had been rescaled to a 256 by a porportional magnitude pixel image.

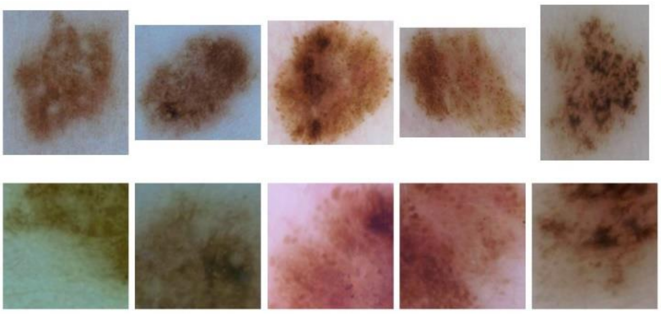

Figure 6. Data augmentation. At the top, five original images. Below, images generated by data augmentation from the original images.

3.3 Performance metric – Jaccard Score

The Jaccard Index is a measure of similarity commonly used in semantic segmentation. The Jaccard Index between two masks A and B is calculated as follows:

$$J(A,B) = \frac{|\mathrm{A} \cap \mathrm{B}|}{|\mathrm{A} \cup \mathrm{B}|} \qquad (2)$$

In the case of multi-class segmentation, the mean Jaccard Score is calculated by taking the Jaccard score of each class and averaging them.

$$J(A,B)_{total} = \frac{\sum_{i=1}^{n} J(A,B)_i}{n} \qquad (3)$$

where *n* is the number of classes (Csurka, 2013) .

3.4 Obstacles

We identified three main obstacles that we tried to circumvent during the training of the models:

- **Class unbalance**. As shown in table 1, there is a big unbalance between classes, which difficults the training of the network.

Class	Others	Milia-like cysts	Negative pigment network	Pigment network	Streaks	Globules
Number of pixels	**1,76x10^9**	5,23x10^7	2,68x10^7	**4,99x10^8**	8,93x10^6	4,65x10^7

Table 1. Number of pixels of each class in the training set. Class 'others' contains the highest number of pixels, followed by class 'Pigment network'.

- **Small structures.** In contrast to the objects of the database on which DeepLabV3 was first trained, the structures to be segmented in this case are very small, as shown in figure 7. Thus, the original architecture of the network might not be optimal for this particular problem. A network that focuses on local structures rather than in context information located very far from the object could suit better the segmentation of attributes of skin lesions.

- **Low resolution of the images**, aggravated by the downsampling layers of the backbone. The initial reduction of the image dimensions to 256 by a proportional number of pixels causes the loss of a big amount of information. The information lost in some of the images is critical for the correct identification of the attributes present in the lesions. As shown in figure 7, some of the structures, well delineated and clearly distinguishable in the original images, become blurred spots that are not differentiable from other kinds of structures.

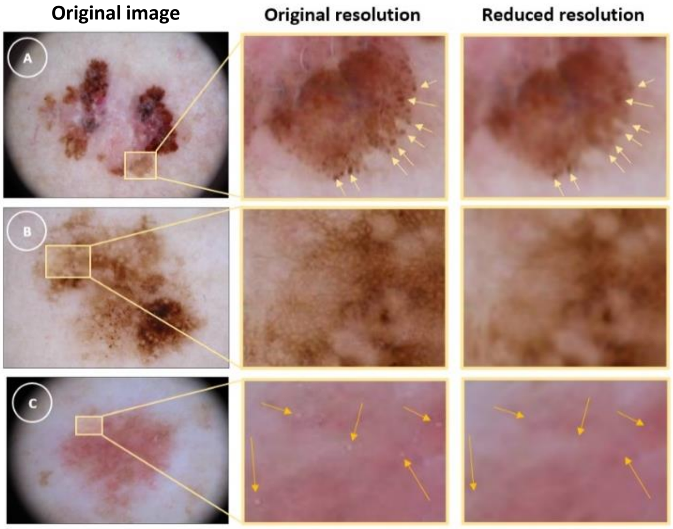

Figure 7. Comparison of different structures visualized with the original resolution and after resizing the images to a lower resolution. A) Globules: they become blurred spots melted with the rest of the lesion B) Pigment network: it loses its characteristic reticular pattern becoming indistinguishable from the rest of the lesion C) Milia-like cysts: they become lighter areas without the strong contrast with respect to its surroundings that characterizes them.

3.5 Models

In our attempt to deal with the problems described in the previous section, we explored different training methods as well as modifications in the architecture of the network. Here we are presenting five of these models.

All models were the result of fine-tuning the DeepLabV3_ResNet101 network pre-trained on the COCO train2017 dataset, on the 20 categories that are present in the Pascal VOC dataset. The backbone and classifier learning rates used were 1x10^{-5} and 1x10^{-3} respectively. As an optimization objective, we used Cross-entropy loss, and the minimization was performed with Adam optimizer. Early stopping was applied by saving the weights of the network from the iteration with the highest validation Jaccard score. A batch size of 16 images was used, and the models were trained for 120 epochs. The models are presented hereunder.

- **Model 1 – 'Without weights'**. We trained the network without weights for the 6 classes.
- **Model 2 – 'With weights'**. We trained the network using weights, giving more importance to the less represented classes. The formula used to calculate the weights of the classes was:

$$Weight_{class\_i} = \frac{n_{majority\_class}}{n_{class\_i}} \quad (4)$$

where *n* is the amount of pixels of class *i* and the majority class is the one with the highest number of pixels.

The subsequent models were trained using weights.

- **Model 3 – 'Leaving out class 1'**. The class 'others' is not well defined because it encompasses everything that is not considered to belong to any of the remaining classes. To reduce the effect of this class during training, we first fine-tuned the network without considering the class 'others' and then, starting from the resulted weights, we performed a fine-tuning including this class and reducing the learning rate.

- **Model 4 – 'Reducing dilation rates of ASPP layer to 6, 12 and 24'**. As the structures to be segmented are small and local, we hypothesized that reducing the dilation rates of the atrous spatial pyramid pooling layer could result in a better performance of the network for this particular problem. We changed the dilation rates of the second, third, and fourth layers of the ASPP module to 6, 12, and 24.

- **Model 5 – 'Reducing dilation rates of ASPP layer to 4, 8 and 12'**. As the previous model raised good results we tried reducing the dilation rates more. We used the dilation rates 4, 8, and 12.

4. Results and discusion

4.1 Analysis and comparison of the models

4.1.1 Comparison of models 1 and 2

In figures 10 and 11, the confusion matrices of the first two models, trained with and without weights, are shown. It is observed that in the first model, trained without applying weights, the majority classes, 'others' and 'pigment network', are over-classified due to its increased prior probability. As a result, pixels from the rest of the classes are misclassified and mainly fall in the class 'others'. This is explained by the fact that the network is trained with many more instances of these two classes than the others and therefore exhibits a bias towards them and practically ignores the minority classes.

On the other hand, on the second model, the minority classes are better classified. In this case, more importance is given to the minority classes. The prior probability of the classes is altered so that the probability of assigning a pixel to a minority class is higher. This comes with a descend on the success at classifying pixels belonging to the majority class.

As observed in table 2, model 1 outperforms model 2 in terms of Jaccard Score, indicating that, despite the improvement in classifying pixels from the minority classes, the worsening on the classification of pixels from the majority classes diminishes the overall classification score. Despite this, as the principal interest of the study is the correct labeling of the 5 relevant attributes, we kept using weights on the training of the successive models. Besides, as shown in figures 8 and 9, the first model rapidly overfits, while the second one still has room for improvement.

4.1.2. *Discussion of the results from model 3*

The results in table 2 indicate that the performance of model 3, 'Leaving out class 1' is worse than model 2. This can be explained for two reasons:

- In the other cases, the model starts from the weights of a network that has been trained to classify 20 different and complex classes while in this case it was trained on the classification of only 5 classes. Thus, the first possible reason for the lowering in performance is that when the model is first trained with the five classes excluding class 1, it centers its attention on those 5 classes worsening the capacity of the backbone to extract features and to identify patterns that are not present on those images. Hence, making the adaptation of the network to the identification of different and newly introduced classes difficult.

- Our second guess is that as the learning rate was lowered after the introduction of pixels from class 1, it should be trained for a longer time to actually be able to compare it with the other models.

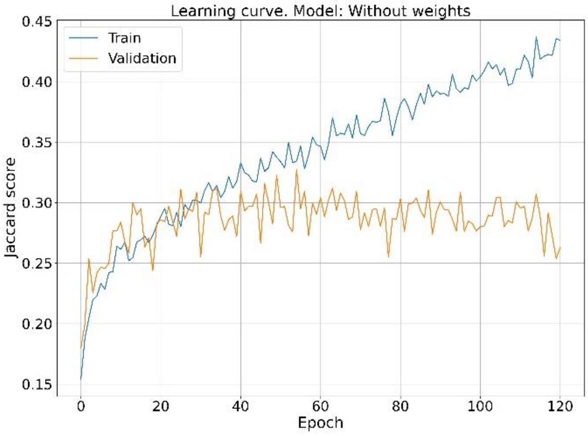

Figure 8. Learning curve of model 1 ('Without weights')

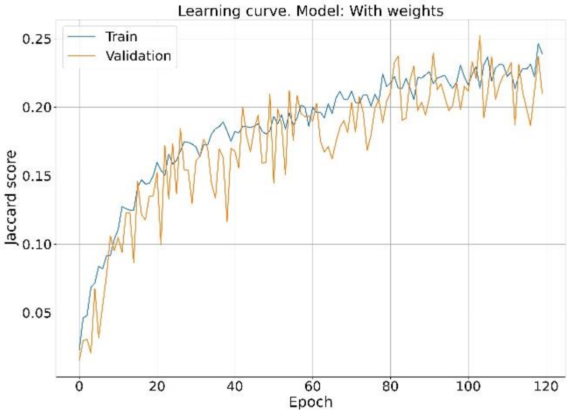

Figure 9. Learning curve of model 2 (‘With weights’)

Model	Jaccard score						
	Label 1	Label 2	Label 3	Label 4	Label 5	Label 6	Mean
Without weights	0.7653	0.0190	0.0240	0.2818	0.0283	0.0978	0.2027
With weights	0.5785	0.0363	0.0428	0.2466	0.0123	0.0734	0.1650
Leaving out class 1	0.5934	0.0233	0.0021	0.2079	0.0307	0.0593	0.1528
Reducing dilation (6,12,24)	0.5892	0.0373	0.0498	0.2378	0.0114	0.0814	0.1678
Reducing dilation (4,8,12)	0.6098	0.0386	0.0395	0.2595	0.0157	0.0771	0.1734

Table 2. Jaccard scores from all the models.

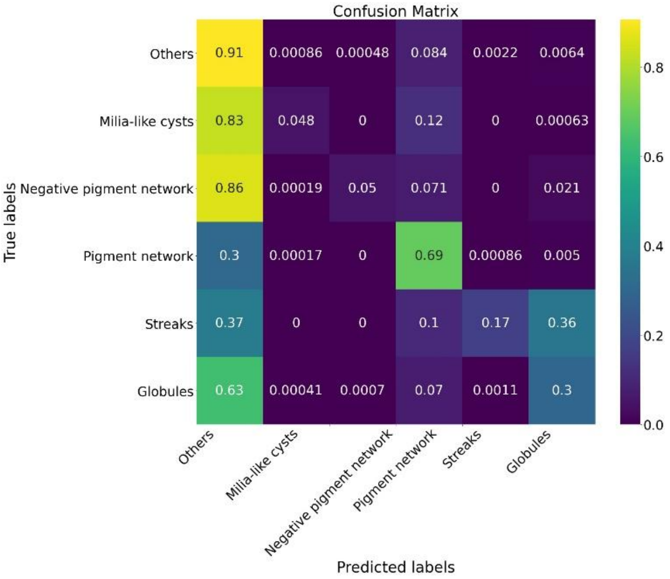

Figure 10. Confusion matrix of model 1 ('Without weights’)

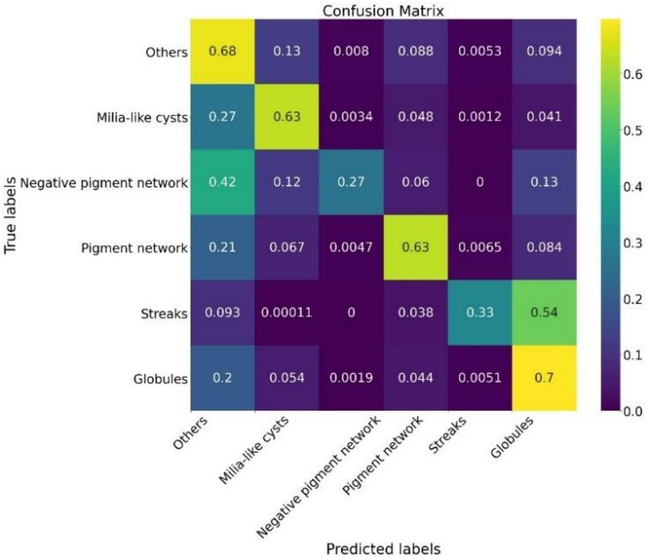

Figure 11. Confusion matrix of model 2 ('With weights’)

4.1.3. *Discussion of the results of model 4*.

The performance of this model on classifying pixels from the class ‘others’ is increased while the success of labeling the other classes is not compromised. Thereby, the overall results of this model surpass the ones achieved with model 2. The results suggest that the reduction of the dilation rates enables the network to focus on pixels that are close to each other and to generalize to attributes presented at small scales.

It makes sense since, to distinguish a Milia-Like cyst or a globule, one only needs to notice the intensity at the center and the contrast of this center with respect to its surroundings; or to distinguish the reticular pattern of a pigment network one only needs to identify the bright holes next to the dark stripes but it is not necessary to look at pixels located very far from the region. Furthermore, the scale on which a car is presented in an image, for example, can vary from very small if it is far from the objective to taking up most of the image if it is very close. The structures analyzed here are usually small, so there is no need for searching at large scales.

4.1.3. *Discussion of the results of model 5*.

In this case, the results outperform those of the previous model, indicating that a greater reduction in the dilation rates of the ASPP module benefits the performance of the model for this particular problem. In the next sections, we analyze the outputs generated by this model on the segmentation of the test images.

4.2 Final model error analysis

In this section, we analyze the resulting masks from the model with reduced dilation rates of 4, 8, and 12.

By visual inspection of the images and a comparison between the ground truth and the predicted labels, we could identify the regions of the image that led to a bad performance of the model and analyze the reasons.

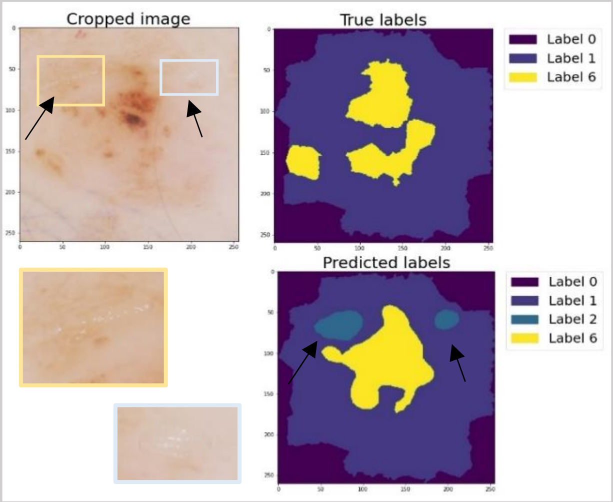

Figure 12. Example of air bubbles being mistaken for Milia-like cysts by the network.

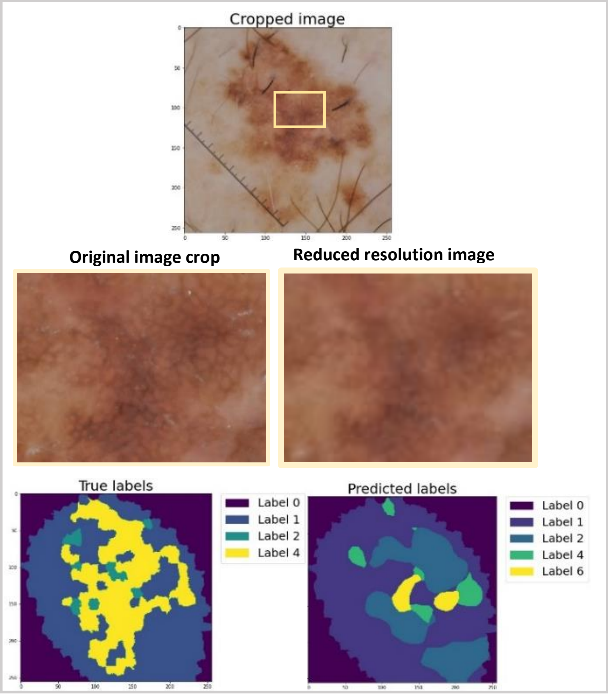

Figure 13. Example of misclassification due to resolution reduction. In this case, label 4, which corresponds to 'Pigment network' is not being correctly distinguished. It is observed that at the midel of the figure, on the left image, the reticular pattern can be easily distinguished. Nevertheless, on the right image, of worse resolution, the pattern is blurred and not easy to classify.

- **Air bubbles**. To photograph the lesions, an immersion fluid is used to cover the skin and avoid light refraction produced by the difference in the refraction index from the air. As shown in figure 12, some of the images present air bubbles that appear in the immersion liquid between the skin and the glass plate of the dermoscope. These bubbles together with the reduction of resolution problem, are constantly mistaken as Milia-like cysts.

- **Low resolution**. As explained in the previous section, for computational reasons, to train the network, all images were downsampled. This resolution reduction impedes the correct discern of the structures in the lesions even by visual inspection. As shown in figure 13, pigment network turns sometimes into blurred spots indistinguishable from other structures, and the same happens with negative pigment network. Furthermore, milia-like cysts become blurred clear spots, losing their characteristic white color and well-delineated shape and leading to mistaking any clear area for a cyst, and the same happens with globules and streaks.

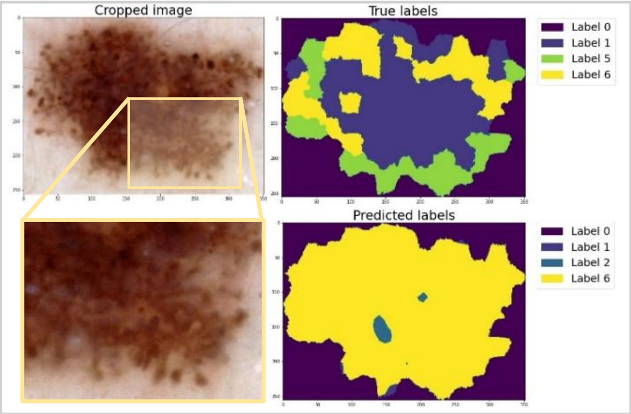

Figure 14. Example of confusion of streaks as globules. The lower part of the image, which in the ground truth labels is classified as streaks (label 5), has been assigned the class 'globules' (label 6) by the network

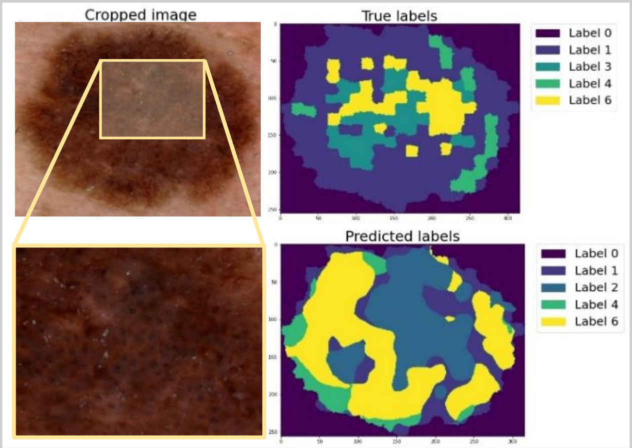

Figure 15. Example of globules (label 6) located in a dark area with low contrast being misclassified.

- Confusion of streaks and globules. In the confusion matrix from the second model (figure 11), it is noteworthy that a lot of pixels from the class ‘streaks’ fall into the category ‘globules’. In the very definition of the two structures, it is said that one is located along the periphery of the lesion, while the other can be located in such area too. Furthermore, the pseudopods from the streaks can easily be mistaken by globules and vice versa. This confusion was found in all the models trained with weights. As observed in figure 14, the model misclassifies these two structures frequently.

- **Globules not recognized**. In figure 15 it is observed that usually when the contrast between globules and their surroundings is not high, these structures are not correctly labeled.

- **Streaks at the periphery of the lesion**. As explained before, streaks are found at the periphery of the lesion. This bias the network increasing the probability of assigning this class to pixels found at the border of the lesion. An example is shown in figure 16.

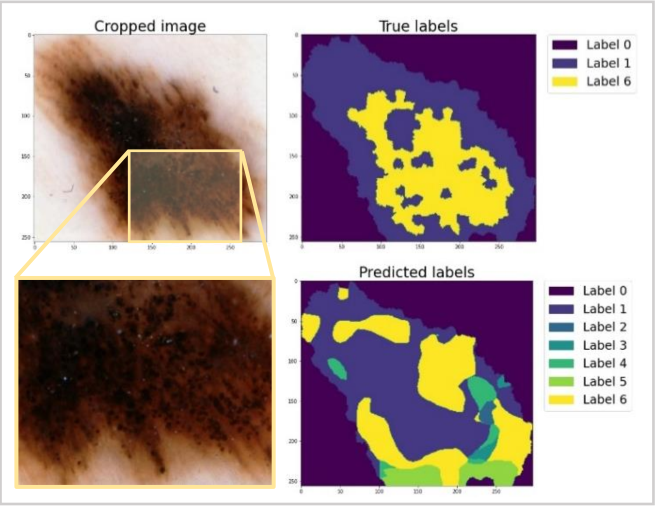

Figure 16. Example of a region situated at the border of the lesion being labeled as ‘streaks’ (label 5).

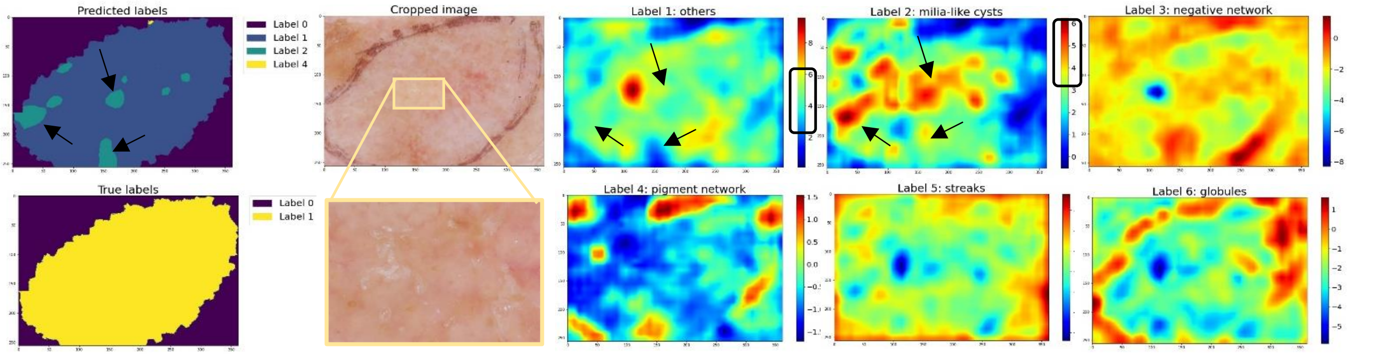

Figure 17. Analysis of the probabilities of belonging to each one of the six classes assigned to the pixels of an image. It is observed that in the mask with the true labels all the lesion is assigned label 1. In the mask with the predicted labels some regions (signaled with black arrows) are assigned the class 2 (milia-like cysts). This is due to the whitish structures observed in the zoomed image, which are air bubbles. By observing the probability maps of the classes it is clear that if those bubbles were not there, the regions would be correctly labeled as class 1.

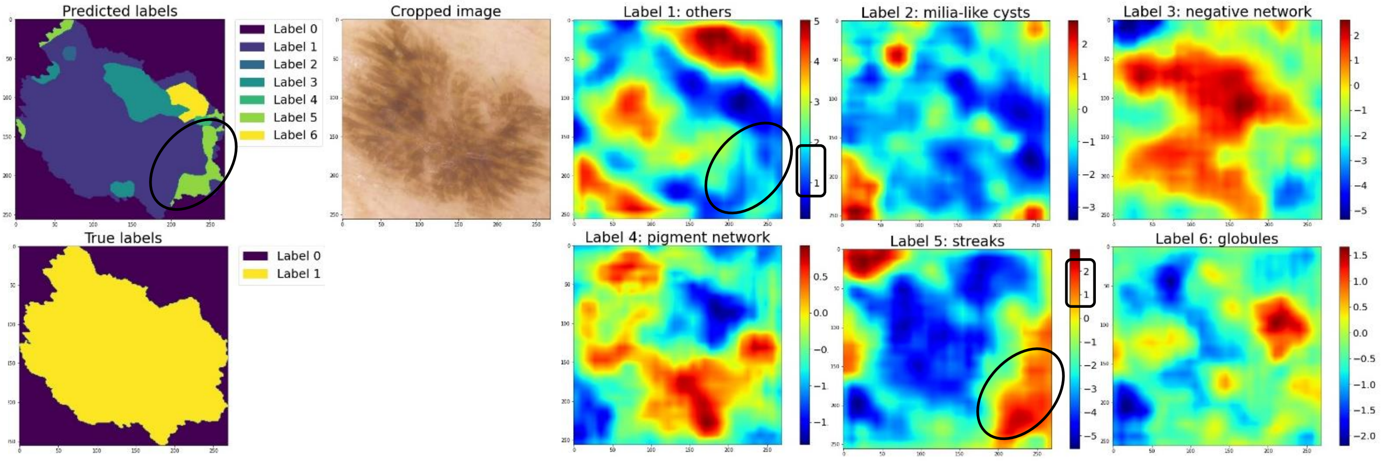

Figure 18. Analysis of the probabilities of belonging to each one of the six classes assigned to the pixels of an image. It is clear to see that, for the region coloured in green in the mask with the predicted labels, the probabilities of being assigned the classes 1 and 5 are both between 1 and 2. In this case, the assigned label was the wrong one, but the network was pretty close to correctly label the area.

4.3 Analysis of the probabilities for each class

Finally, we analyzed the probabilities of belonging to each one of the six classes assigned to the pixels.

As illustrated in figure 17, we found that in a lot of cases, when there was a misclassified region, the second most likely label for that region was the correct one. Moreover, we found that this was particularly the case when the areas were wrongly labeled as milia-like cysts, which was often caused by the presence of air bubbles.

Furthermore, in a lot of cases, as in the example shown in figure 18, the probability of assigning the correct class was very close to that of the finally assigned class. Thus, we can conclude that the results in terms of soft labeling are good and that the performance of the model would raise in the absence of air bubbles.

6. Conclusions.

Early detection is crucial for the successful treatment of melanoma. In this study, we explored different ways of training a model for automatic attribute segmentation of skin lesions, a tool that can help clinicians with the task of melanoma detection and diagnosis.

We found that to train the network applying weights that provide more importance to the less represented classes is a good practice to overcome the problem of class unbalance. Furthermore, a reduction in the dilation rates of the ASPP module of the network increases the performance of the model on this particular problem, where the structures to be segmented are small and local.

Moreover, we realized that training the neural network excluding the class with a higher number of instances for later training it including the majority class does not yield improved results.

Additionally, we analyzed by visual inspection the test masks predicted by the best model and compared them to the ground truth masks. We concluded that the model encounters difficulty in classifying the attributes when there is a big loss of resolution or under the presence of air bubbles. Furthermore, we found that streaks and globules are often mistaken with each other, the regions located at the periphery of the lesion are sometimes misclassified as streaks, and globules are generally not recognized by the model when the region is dark and the contrast with respect to their surroundings is low.

To end, we evaluated the separate probabilities of each class predicted by the model, from which we could derive that the model accomplishes a good performance in terms of soft labeling and that the absence of air bubbles would result in an improvement of the test score of hard labeling.

6. Future work.

During the course of this research, some ideas arose about how to obtain better results for the present problem. In this section, we introduce these ideas.

As previously discussed, the resolution reduction of the images can cause a big loss of information. This might be worsened by the reduction of the feature maps dimesions in the backbone layers. Thus, another possible approach to try to avoid information loss is to change the stride of the layers from the backbone from 2 to 1 in order to reduce the output stride value and the consequent loss of information.

As expounded in the introduction, unanimity does not exist on the task of attribute segmentation. Images are usually segmented in different ways depending on the annotator, so multiple masks from different annotators should be considered. In this way, a gold standard could be estimated to prevent the network from being trained with a bias towards the way of segmenting of a single annotator.

7. References

A. Murugan, S. H. (2019). Detection of Skin Cancer Using SVM, Random Forest and kNN Classifiers. *Journal of medical systems, 43*(8), 1-9.

Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2019). Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv 2017. arXiv preprint arXiv:1706.05587*.

Csurka, G., Larlus, D., & Perronnin, F. (2013). What is a good evaluation measure for semantic segmentation? *Bmvc*, *27*, pp. 10-5244.

Dinnes, J., J Deeks, J., Chuchu, N., Ferrante di Ruffano, L., Matin, R. N., Thomson, D. R., . . . Williams, H. C. (2018). Dermoscopy, with and without visual inspection, for diagnosing melanoma in adults. *Cochrane Database of Systematic Reviews, 12*(CD011902). doi:10.1002/14651858

Forsea, A.-M. (2020). Melanoma Epidemiology and Early Detection in Europe: Diversity and Disparities. *Dermatology Practical & Conceptual, 10*(3), e2020033. doi:10.5826/dpc.1003a33

Goyal, M., Oakley, A., Bansal, P., Dancey, D., & Yap, M. H. (2019). Automatic Lesion Boundary Segmentation in. *arXiv*(1902.00809).

H Kittler, H. P. (2002). Diagnostic accuracy of dermoscopy. *The Lancet Oncology, 3*(3), 159-165. doi:10.1016/S1470-2045(02)00679-4.

He, X., Lei, B., & Wang, T. (2019). SANet:Superpixel Attention Network for Skin Lesion Attributes Detection. *arXiv, 1910.08995*.

ISIC 2018 Challenge - Task 2: Lesion Attribute Detection . (n.d.). Retrieved 08 25, 2021, from challenge.isic-archive.com: https://challenge.isic-archive.com/

ISIC Challenge Datasets. (n.d.). Retrieved 25 08, 2021, from https://challenge.isic-archive.com/data

isic-archive.com. (n.d.). Retrieved 08 25, 2021, from https://www.isic-archive.com/#!/topWithHeader/tightContentTop/about/aboutIsicOverview

Khan, M. A., Akram, T., Zhang, Y.-D., & Sharif, M. (2021). Attributes based skin lesion detection and recognition: A mask RCNN and transfer learning-based deep learning framework. *Pattern Recognition Letters, 143*, 58-66. doi:10.1016/j.patrec.2020.12.015

Kittler, H., Marghoob, A. A., Argenziano, G., Carrera, C., Curiel-Lewandrowski, C., Hofmann-Wellenhof, R., . . . Halpern, A. (2016). Standardization of terminology in dermoscopy/dermatoscopy: Results of the third consensus conference of the International Society of Dermoscopy. *J. Am. Acad. Dermatol, 74*(6), 1093-106. doi:10.1016/j.jaad.2015.12.038

Krizhevsky, A., Sutskever, I., & E. Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in neural information processing systems, 25*, 1097-1105.

López-Labraca, J., Fernández-Torres, M. Á., González-Díaz, I., Díaz-de-María, F., & Pizarro, Á. (2018). Enriched Dermoscopic-Structure-Based CAD. *Multimedia Tools and, 77*(10), 12171-12202. doi:10.1007/s11042-017-4879-3

Minagawa, A. (2017). Dermoscopy–pathology relationship in seborrheic keratosis. *The Journal of dermatology, 44*(5), 518-524. doi:10.1111/1346-8138.13657

Pizzichetta, M. A., Talamini, R., Marghoob, A. A., Soyer, H. P., Argenziano, G., Bono, R., . . . Menzies, S. W. (2013). Negative pigment network: An additional dermoscopicfeature for the diagnosis of melanoma. *Journal of the American Academy of Dermatology, 68*(4), 552-559.

Seo, H., Khuzani, M. B., Vasudevan, V., Huang, C., Ren, H., Xiao, R., . . . Xing, L. (2020). Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications. *Medical physics, 47*(5), e148-e167. doi:10.1002/mp.13649

Seyedhosseini M, T. T. (2015). Segmentation with Contextual Hierarchical Models. *IEEE Trans Pattern Anal Mach Intell, 38*(5), 951-64. doi:10.1109/TPAMI.2015.2473846

Sonthalia, S., Yumeen, S., & Kaliyadan., F. (2021). Dermoscopy Overview and Extradiagnostic Applications. *StatPearls*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK537131/

Thoma, M. (2016). A Survey of Semantic Segmentation. *arXiv preprint arXiv:1602.06541*.