

# Skin Lesion Detection in Dermatological Images using Deep Learning

---

**Jose Carlos Moreno-Tagle\***, Jimena Olveres, Boris Escalante-Ramirez

*School of Engineering, National Autonomous University of Mexico (UNAM), Mexico*

*E-mail: [astroboy3.1415@gmail.com](mailto:astroboy3.1415@gmail.com)*

This paper demonstrates that it is possible to approach the skin lesion classification problem as a detection problem, a much more complex and interesting problem, by training a deep neural network based detection architecture and applying image processing techniques to a dermatology dataset as part of the data augmentation strategy with satisfactory and promising results. The image dataset used in the experiments comes from the ISIC Dermoscopic Archive, an open-access dermatology repository. In particular, the ISIC 2017 dataset, a subset of the ISIC archive, released for the annual ISIC challenge was used. We show that it is possible to adapt a high quality imaging dataset to the requirements demanded by a deep learning detection architecture such as YOLOv3. In conjunction with image processing techniques as a previous step, the deep neural network was successfully trained to identify and locate three different types of skin lesions in real-time.

*Artificial Intelligence for Science, Industry and Society, AISIS2019  
October 21-25, 2019  
Universidad Nacional Autónoma de México, Mexico City, México*

---

\*Speaker.

## 1. Introduction

Skin is the human body's largest organ [1]. Skin performs many vital functions, it provides protection from the outside world by guarding the body against damaging sunlight, extreme temperatures, as well as germs and harmful substances [2]. Skin is also responsible, among many other things, for sensation, heat regulation and vitamin D production [3].

Skin cancer is the most common form of cancer in the United States [4]. According to the American Academy of Dermatology, it is estimated that one in five Americans will develop skin cancer in their lifetime. As with any other type of cancer, the sooner it is spotted, the better the prognosis becomes. When skin cancer is caught early, is highly treatable. This type of cancer is mainly diagnosed through a visual inspection. The detection process begins with an initial clinical screening [5]. It may be followed, depending on the screening results, by a dermoscopic analysis, a biopsy and histopathological examination. Melanoma is a rare form of skin cancer that develops from skin cells called melanocytes which are located in the layer of basal cells at the deepest part of the epidermis, they are also found in the iris, inner ear, nervous system, heart and hair follicles among other tissues. Melanoma represents fewer than 5% of all skin cancers in the United States. However, melanoma is also the deadliest form of skin cancer, it is responsible for about 75% of all skin cancer related deaths [6]. Some studies have suggested that melanoma annual incidence rate has increased among the caucasian population from 4 to 8% [7]. According to the American Cancer Society in its Cancer Statistics 2019 paper, it is estimated that the number of new melanoma cases diagnosed in 2019 will increase by 7.7 %.

Dermatologists use different visual inspection methods to determine if a skin lesion under analysis may be melanoma (malignant). Some of these methods are the ABCDE rule, 7 points checklist and the Menzies method just to name a few. All these methods are based on the visual features that make a malignant melanoma lesion distinguishable.

In recent years, there has been a rising interest in applying deep learning techniques to medical problems. Deep neural networks have shown amazing results in skin lesion classification tasks, some outstanding examples of these can be found in [8, 5, 9]. Dermatology is a medical specialty which can greatly benefit from the powerful feature extraction capabilities of deep neural networks.

### 1.1 Deep Neural Networks for Detection Tasks

You only look once or YOLOv3 is a state of the art real-time object detection system, the first version was developed by Redmon et al. (2016) [10]. YOLOv3 is extremely fast and accurate [11] and offers an excellent trade-off between accuracy and speed. This state of the art detection system distinguishes itself by applying a single neural network to the full image [10] and is capable of performing inference on a  $320 \times 320$  image in 22 [ms] on a Titan X GPU. YOLO divides the image into regions and predicts bounding boxes and probabilities for each region, the bounding boxes are weighted by the predicted probabilities [11]. The detections are thresholded by a confidence value of 0.25 as default, so any object which reaches a confidence value of 0.25 or higher will be detected. In this way just the detections with the highest confidence remain.

YOLOv3 uses a new deep neural network architecture called Darknet-53. Darknet-53 has 53 convolutional layers and mainly uses  $3 \times 3$  and  $1 \times 1$  filters as well as shortcut connections [11]. Darknet-53 employs leaky ReLU as the activation function for the convolutional layers.

Every convolutional layer in YOLOv3 makes use of batch normalization. The normalization of the extracted features is accomplished by adjusting and scaling the activations of the input layer.

YOLOv3 predicts  $t_x, t_y, t_w, t_h$  which are the predicted coordinates of the bounding box. The first pair of values correspond to the coordinates of the center of the detected object and the last pair correspond to the width and height of the bounding box. For the bounding box prediction task, the sum of squared errors is used as loss function. YOLOv3 also predicts an objectness score for each bounding box using logistic regression. The objectness score is 1 if the bounding box overlaps a ground truth object by more than any other bounding box prior. For class prediction YOLOv3 does not use a softmax function. Instead, independent logistic classifiers are used in conjunction with binary cross-entropy as loss function. YOLOv3 performs image analysis at three different scales to predict boxes. Several convolutional layers are added to the base feature extractor Darknet-53, the last of these layers predicts the bounding box, objectness and class predictions.

It is a common scenario in which YOLOv3 may detect the same object multiple times where the bounding boxes differ subtly in size and center location. To address this issue, YOLOv3 applies non-max suppression to remove all redundant overlapping bounding boxes with a lower confidence score.

## 2. State of the Art

One of the most relevant articles about the applications and the potential that deep learning can offer in the field of dermatology was presented by [5]. In [5], they make the following proposals: use the Inception v3 neural network architecture, collect and make use of the largest number of dermatological images up to that moment from different databases, create a taxonomy of skin diseases and an algorithm to map those diseases into training classes, make use of transfer learning with fine tuning in all layers to retrain the network with the dermatological images.

The performance of this network was tested against 21 certified dermatologists in two critical cases: carcinomas vs seborrheic keratosis (identification of the most common cancer vs benign lesion) and melanomas vs nevus (identification of the most deadly cancer vs visually similar benign lesion). The experiments performed by [5] showed that the general accuracy of the proposed convolutional neural network is  $72.1 \pm 0.9\%$  (mean  $\pm$  sd). In contrast, two dermatologists achieve an accuracy of 65.56% and 66.0% in a subset of the validation set. During the validation of the algorithm with a partition of nine classes of diseases, the accuracy of the CNN is  $55.4 \pm 1.7\%$  while two dermatologists get an accuracy of 53.3% and 55.0%.

From the results shown by this last study, a great interest was born from different research groups to exploit the deep learning tools in order to overcome the results previously achieved. Since then, many different approaches have been proposed to address this problem and mitigate the lack of a database as rich and extensive as the one used in [5]. Some of the most recent works with important results are those presented by [8, 9, 12] and [13], they all focus strictly on solving the classification problem.

## 3. Database

The International Skin Imaging Collaboration archive (ISIC archive) is a high quality dermo-

Class	Number of images in each partition		
	Training set	Validation set	Test set
Melanoma	748	83	233
Nevi	748	83	233
Keratosis	748	83	233

Table 1: Balanced dataset by applying over-sampling/under-sampling techniques. Segmented images were added as part of the data augmentation.

scopic image database, the images were collected from leading clinical centers internationally and acquired from a variety of devices within each center [14]. The images that can be obtained directly from the ISIC archive provide relevant clinical information such as the age and sex of the patient, localization of the lesion, diagnosis and confirmation method of such diagnosis. At the time of preparing this work, the dataset consisted of 19,330 benign lesions and 2,286 malignant lesions.

In particular, we performed our experiments with the ISIC 2017 dataset, the images from this dataset are a selected subset of images from the ISIC archive, intended to be used in its annual challenge. The dataset consists of images from three classes: 521 melanoma, 386 seborrheic keratosis and 1,843 nevus.

The ISIC 2017 dataset is partitioned into training, validation and testing data. Deep learning models learn the most when they are trained with the largest possible amount of images. Based on this idea, once the dataset was balanced by applying under-sampling and over-sampling methods to the appropriate classes and adding segmented images as new data, we decided to merge the datasets into one and divide it into 90% for training and 10% for validation.

### 3.1 Data Augmentation

As a preprocessing step before exposing the deep neural network to the original dataset images, we performed data augmentation over the training set. Adding more training data to deep learning models tends to boost its performance. In the particular case of dermatological images, we can take advantage of its nature to apply several transformations to each image. For this kind of images, it does not harm the deep neural network performance if the image is blurred, flipped horizontally, vertically or rotated a certain amount of degrees.

The selected techniques for data augmentation were horizontal flips, rotations given a random value between  $-90$  and  $90$  degrees and transformations in the colors of the original images through the saturation, exposure and hue values. In particular, the values of such parameters were set to 1.5, 1.5, and 0.1 respectively. The previous transformations are applied to each image of the training set, thus generating a new sample for the network.

## 4. Experiments and Results

We have the hypothesis that increasing the number of data can help improve performance in the detection process. Thus, two different image processing segmentation techniques were ap-

plied: segmentation in the HSV space with Grabcut [15] and edge detection. The main method for achieving image segmentation was Grabcut in the HSV color space. However, not all the results were satisfactory so experimenting with some other approaches that could deliver good results was a viable option. Figure 1 shows an example illustrating the above situation.

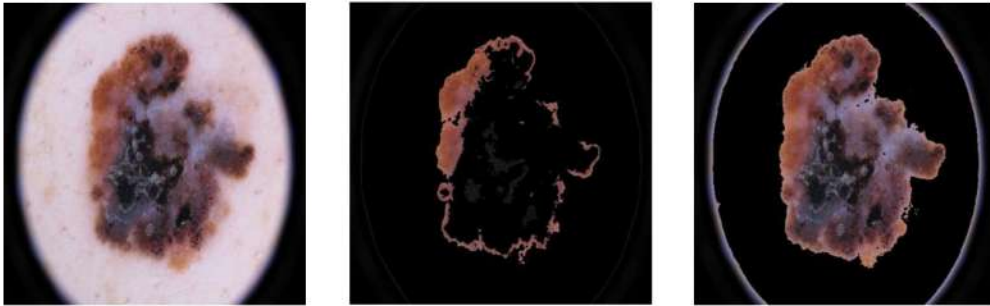


Figure 1: Original image of a melanoma (left). Image incorrectly segmented by Grabcut (center). Segmented image using thresholding (right).

For those images where the segmentation result is not acceptable either because the method completely eliminated the lesion or the amount of surrounding skin is excessive, the chosen segmentation method was edge detection through the Canny filter [16]. Figures 2 and 3 show some of the stages of the segmentation process by both methods.

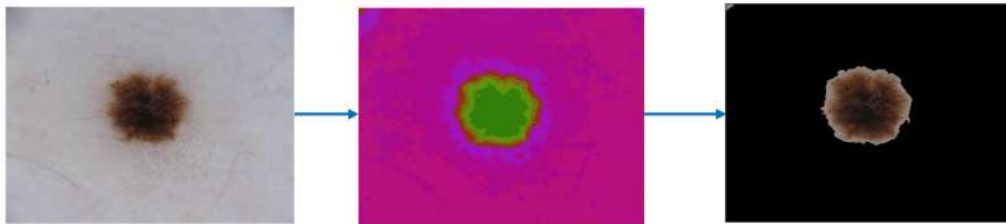


Figure 2: Grabcut-segmented skin lesion in the HSV color space



Figure 3: Segmented skin lesions through edge detection.

It can be seen that the segmentation of the lesions is not perfect, however this should not diminish the performance of the model but on the contrary, these imperfections could help the model to achieve a better distinction between classes since some of the surrounding skin is preserved. During a visual inspection when evaluating a case, Dermatologists consider the colors of a lesion as well as those presented by the surrounding skin and can then offer a diagnosis.

YOLOv3 was trained for 522 epochs and for over 200,000 iterations using stochastic gradient descent with the following hyperparameters: momentum = 0.9, learning rate decay = 0.0005 and learning rate = 0.001. All images during this stage were resized to  $608 \times 608$ .

At the end of the training stage, the performance of YOLOv3 was evaluated over the validation set. As shown in Table 2, it was observed that the best results were obtained from the weights corresponding to iteration 40,000, reaching a precision of 73% and a recall of 80%. The following figures show the evolution of both metrics over different iteration steps, i.e., different weight values resulting from the optimization process.

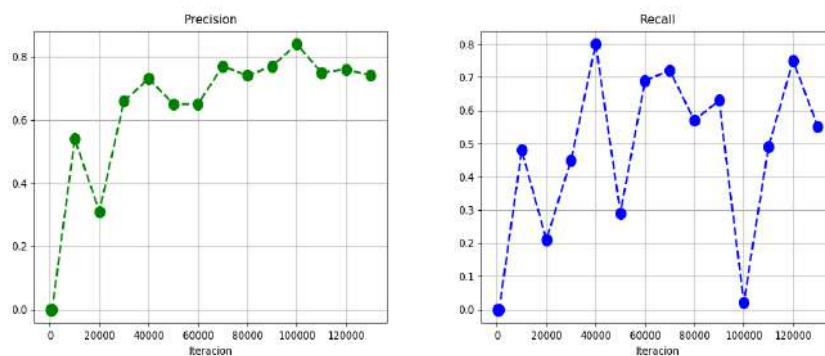


Figure 4: Performance metrics over the validation set for YOLOv3 trained on high quality dermatologic images.

By inspecting the metrics through the validation process and based on the visual predictions obtained, we can say that the deep neural network is in fact learning to identify each disease class. Attempting to correctly classify this kind of dermatological images is a challenging task for dermatologists and deep learning models as well. In [17], they report that visual inspection without help by expert dermatologists is associated with a diagnostic accuracy of around 60%.

Iteration	Precision	Recall	F1
10000	0.54	0.48	0.51
20000	0.31	0.21	0.25
30000	0.66	0.45	0.54
<b>40000</b>	<b>0.73</b>	<b>0.8</b>	<b>0.76</b>
50000	0.65	0.29	0.4
60000	0.65	0.69	0.67
70000	0.77	0.72	0.74

Table 2: Metrics evolution over the validation set.

A two minute video, composed of test images, was created to be processed in real-time by YOLOv3. Each image of the video was verified by a dermatologist to assure that the skin lesion in the image corresponded to the label attached to it. The video was processed at approximately 28 FPS on a NVIDIA GeForce GTX 1080 GPU. The following images are some of the frames from

the test video. It is important to remark that the test set has not been seen before by the deep neural network.

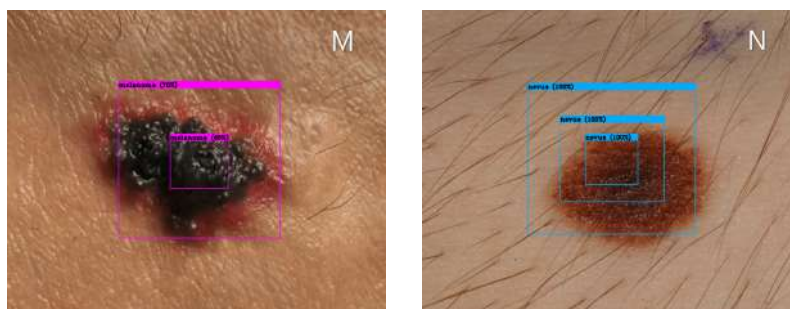


Figure 5: Predictions for test images with confidence score. The white color character indicates the ground-truth label, the bounding box and its tag is the predicted class.

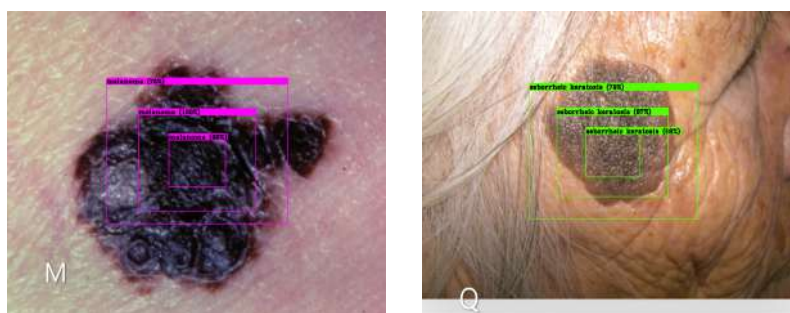


Figure 6: Predictions for test images with confidence score. The left image is correctly predicted as melanoma, the right image is labeled as seborrheic keratosis and predicted as such.

## 5. Conclusions

This paper showed that a state of the art detection system like YOLOv3 can be retrained to identify skin lesions in medical images and when combined with image processing techniques to use the new segmented images as part of the data augmentation strategy, satisfactory results can be achieved. These results can be greatly improved by satisfying one of the most important requirements for using deep learning, it is essential to gather as much data as possible for training a deep neural network that can really learn to distinguish between different classes by being able to generalize after being exposed to a large number of examples. However, getting access to large datasets is still a problem when working in medical applications.

Based on the results shown, we can say that the deep neural network learned to differentiate between three types of skin lesions from the extracted characteristics. The skin lesions can be visually very similar so the problem of classification is complex for both computational models and specialists in dermatology. These results show the powerful feature extraction capacity of deep neural networks as well as their ability to adapt as solution to new problems completely different from those for which they were originally conceived.

Detection systems can provide a user or a medical specialist with useful information about the lesion under analysis in real-time, they can provide information about the location of the lesion or highlight the nature of multiple lesions by providing a diagnosis for each one of them with a certain degree of confidence. Deep learning based detection systems can be of great use for dermatologists, they can be put to work as a desktop automatic diagnostic app or can be embedded into a smartphone attached to a mobile dermatoscope.

### 5.0.1 Acknowledgements.

Jose Carlos would like to thank CONACYT (Consejo Nacional de Ciencia y Tecnologia) for the financial support (CVU: 857565) during his graduate studies. Many thanks to MD Jesus Hernandez Calderon for sharing his expertise and his willingness to continue collaborating in this project. Authors wish to thank UNAM PAPIIT grants IA103119 and IN116917.

### References

- [1] J. Kanitakis, *Anatomy, histology and immunohistochemistry of normal human skin*, *European journal of dermatology* **12** (2002) 390.
- [2] E. Proksch, J.M. Brandner and J.-M. Jensen, *The skin: an indispensable barrier*, *Experimental dermatology* **17** (2008) 1063.
- [3] C.-M. Chuong, B. Nickoloff, P. Elias, L. Goldsmith, E. Macher, P. Maderson et al., *What is the 'true' function of skin?*, *Experimental dermatology* **11** (2002) 159.
- [4] A.C.S. (ACS), "Cancer facts & figures 2019 [internet]."
- [5] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau et al., *Dermatologist-level classification of skin cancer with deep neural networks*, *Nature* **542** (2017) 115.
- [6] D. Schadendorf and A. Hauschild, *Melanoma in 2013: Melanoma—the run of success continues*, *Nature Reviews Clinical Oncology* **11** (2014) 75.
- [7] J.F. Gallegos Hernández and O.E. Nieweg, *Melanoma cutáneo (mc): diagnóstico y tratamiento actuales*, *Gaceta Médica de México* **150** (2014) 175.
- [8] N.C. Codella, D. Gutman, M.E. Celebi, B. Helba, M.A. Marchetti, S.W. Dusza et al., *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)*, in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 168–172, IEEE, 2018.
- [9] Y. Fujisawa, Y. Otomo, Y. Ogata, Y. Nakamura, R. Fujita, Y. Ishitsuka et al., *Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis*, *British Journal of Dermatology* **180** (2019) 373.
- [10] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, *You only look once: Unified, real-time object detection*, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.
- [11] J. Redmon and A. Farhadi, *Yolov3: An incremental improvement*, *arXiv preprint arXiv:1804.02767* (2018) .
- [12] B. Harangi, *Skin lesion classification with ensembles of deep convolutional neural networks*, *Journal of biomedical informatics* **86** (2018) 25.



- [13] A. Menegola, J. Tavares, M. Fornaciali, L.T. Li, S. Avila and E. Valle, *Recod titans at isic challenge 2017*, *arXiv preprint arXiv:1703.04819* (2017) .
- [14] I.S.I.C. (ISIC), “Isic 2016: Skin lesion analysis towards melanoma detection.”
- [15] F.I. Tushar, *Automatic skin lesion segmentation using grabcut in hsv colour space*, *arXiv preprint arXiv:1810.00871* (2018) .
- [16] J. Canny, *A computational approach to edge detection*, *IEEE Transactions on pattern analysis and machine intelligence* (1986) 679.
- [17] H. Kittler, H. Pehamberger, K. Wolff and M. Binder, *Diagnostic accuracy of dermoscopy*, *The lancet oncology* **3** (2002) 159.