Multi-Class Lesion Diagnosis with Pixel-wise Classification Network

Manu Goyal¹, Jiahua Ng², and Moi Hoon Yap¹

¹ Visual Computing Lab, Manchester Metropolitan University, M1 5GD, UK
² University of Sheffield, S10 2TN, UK

Abstract. Lesion diagnosis of skin lesions is a very challenging task due to high inter-class similarities and intra-class variations in terms of color, size, site and appearance among different skin lesions. With the emergence of computer vision especially deep learning algorithms, lesion diagnosis is made possible using these algorithms trained on dermoscopic images. Usually, deep classification networks are used for the lesion diagnosis to determine different types of skin lesions. In this work, we used pixel-wise classification network to provide lesion diagnosis rather than classification network. We propose to use DeeplabV3+ for multi-class lesion diagnosis in dermoscopic images of Task 3 of ISIC Challenge 2018. We used the various post-processing methods with DeeplabV3+ to determine the lesion diagnosis in this challenge and submitted the test results.

1 Introduction

Automated skin lesions analysis such as skin cancer [1] and diabetic foot ulcers [2] has gain its popularity, thanks to the advancement of computer vision, machine learning and deep learning algorithms. Skin cancer is the most common cancer among all other cancers [3]. The malignant skin lesions consist of the melanocytic lesion, i.e. melanoma, and non-melanocytic lesion, i.e. basal cell carcinoma. Although melanoma is the least common type of skin cancer, it is the most aggressive and deadly cancer [4]. Hence, it is important to have early detection to save life. According to the prediction of Melanoma Foundation [5], the estimated new cases of melanoma in the United States is 87,110 (200% increased since 1973) with 9,730 predicted deaths. For International Skin Imaging Collaboration (ISIC) 2018 Challenge, we propose the use of DeeplabV3+ for Task 3: lesion diagnosis and we describe our methods in the following section.

2 Methodology

This section discusses the challenge dataset, the preparation of ground truth and proposed methods.
2.1 ISIC 2018 Challenge Dataset for Task 3

ISIC 2018 Lesion Diagnosis Challenge consists of 10015 training images, 193 validation images and 1512 testing images [6]. It consists of 7 different types of skin lesions including melanoma, nevi, basal cell carcinoma, actinic keratosis, benign keratosis, dermatofibroma, vascular lesion. Benign keratosis is super category of different types of keratosis lesions such as seborrheic keratosis, solar lentigo and lichen planus-like keratosis. The dataset used by the challenge is also known as, HAM10000 which is a collection of 11,788 dermoscopic images of different skin lesions collected from the multiple sources around the world [7]. We did not use any external data. To improve the performance and reduce the computational costs, we resized all the images to 500 × 375.

2.2 Preprocessing and Preparation of Ground Truth

We used Task 1 dataset which consists of 2594 training images and segmentation masks to train DeeplabV3+ to generate the skin lesion masks for HAMM-10000 dataset. Due to ISIC Challenge datasets comprised of dermoscopic skin lesion taken by different dermoscope and camera devices all over the world, it is important to perform pre-processing for color normalization and illumination. We pre-process the dataset with a color constancy algorithm, Shades of Gray algorithm [8], and the pre-processed results are as shown in the Fig. 1. With classification labels, we converted the skin lesion masks to represent different categories of skin lesions as shown in the Fig. 2. We used the Pascal-VOC format i.e. the input images are defined in RGB colorspace and 8-bit paletted images are used for representing the ground truth.

![Fig. 1. Examples of pre-processing by Shades of Gray algorithm [8].](image-url)
Fig. 2. Illustration of seven types of skin lesion with ground truth masks where (a) Nevi (b) Melanoma (c) Basal Cell Carcinoma (d) Actinic Keratosis (e) Benign Keratosis (f) Dermatofibroma and (g) Vascular Lesion.

2.3 DeeplabV3+

DeepLabV3 [9] is one of the best performing pixel-wise classification (semantic segmentation) networks. It assigns semantic label lesion to every pixel in a dermoscopic image. DeeplabV3+ is an encoder-decoder network which make use of CNN (Xception-65) with atrous convolution layers to get the coarse score map and then, conditional random field is used to produce final output. We trained DeeplabV3+ for 100 epochs with batch size of 4, base learning learning rate of 0.0001 with learning decay factor of 0.1, decay step of 5000 and weight decay of 0.00005. We run our experiments on a machine with the following specification: (1) Hardware: CPU - Intel i7-6700 @ 4.00Ghz, GPU - NVIDIA TITAN X 12Gb, RAM - 32GB DDR5 (2) Software: Tensor-flow.

Fig. 3. Examples of different types of semantic segmentation for validation set: (a) result with one class lesion type; (b) result with no lesion detected; and (c) result with multiple lesion types.
Post-processing Methods to determine lesion class  After training, we used the images in validation set and testing set to produce outputs. We received three types of results from the inference as shown in the Fig. 3. The number of lesion types detected is shown in Fig. 4.

For single type detection, we directly assume diagnosis confidence of 1.0 of detected lesion class. There are very few cases of no detection in both validation set and testing set, we assumed it as 1.0 for Nevi class. For multi-class detection, we adopted 2 different strategies to determine class of lesion. First is diagnosis confidence based on number of pixels of each lesion class in the detection means that lesion class which has most pixels in the detection is considered. Second strategy is priority based on the number of images of different type of skin lesion in the dataset as shown in the Table 1. Third method is again a priority based strategy with preference of the malignant lesions over the benign and number of images in the training set according to the Table 2.

3 Validation and Conclusion

For lesion diagnosis challenge, evaluation metric called normalized multi-class accuracy is used to determine the winner of the competition. We did not report Validation scores due to the rules of competition. We will report our complete analysis and final test results in a later stage as it will be announced later this month. We are submitting these three methods for this year’s competition.
## Table 1. Priority strategy based on distribution of skin lesions in HAMM training set

| Priority | Lesion Class          | Number of Images in training set |
|----------|-----------------------|----------------------------------|
| 1        | Dermatofibroma        | 115                              |
| 2        | Vascular Lesion       | 142                              |
| 3        | Actinic Keratosis     | 327                              |
| 4        | Basal Cell Carcinoma  | 514                              |
| 5        | Benign Keratosis      | 1099                             |
| 6        | Melanoma              | 1113                             |
| 7        | Nevi                  | 6705                             |

## Table 2. Priority strategy based on benign or malignant and number of images in HAMM training set

| Priority | Lesion Class          | Benign or Malignant              | Number of Images in training set |
|----------|-----------------------|----------------------------------|----------------------------------|
| 1        | Dermatofibroma        | Malignant                        | 115                              |
| 2        | Basal Cell Carcinoma  | Malignant                        | 514                              |
| 3        | Melanoma              | Malignant                        | 1113                             |
| 4        | Vascular Lesion       | May be Malignant or Benign       | 142                              |
| 5        | Actinic Keratosis     | Malignant potential              | 327                              |
| 5        | Benign Keratosis      | Benign                           | 1099                             |
| 7        | Nevi                  | Benign                           | 6705                             |
References

1. Manu Goyal and Moi Hoon Yap. Multi-class semantic segmentation of skin lesions via fully convolutional networks. *arXiv preprint arXiv:1711.10449*, 2017.

2. Manu Goyal, Moi Hoon Yap, Neil D Reeves, Satyan Rajbhandari, and Jennifer Spragg. Fully convolutional networks for diabetic foot ulcer segmentation. In *Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on*, pages 618–623. IEEE, 2017.

3. Sameena Pathan, K Gopalakrishna Prabhu, and PC Siddalingawamy. Techniques and algorithms for computer aided diagnosis of pigmented skin lesion—a review. *Biomedical Signal Processing and Control*, 39:237–262, 2018.

4. National Cancer Institute. Cancer stat facts: Melanoma of the skin, 2017. last access: 26/10/17.

5. Melanoma Foundation (AIM). Melanoma stats, facts and figures, 2017. last access: 27/10/2017.

6. Noel CF Codella, David Gutman, M Emre Celebi, Brian Helba, Michael A Marchetti, Stephen W Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, 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). *arXiv preprint arXiv:1710.05006*, 2017.

7. Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions. *arXiv preprint arXiv:1803.10417*, 2018.

8. Graham D Finlayson and Elisabetta Trezzi. Shades of gray and colour constancy. In *Color and Imaging Conference*, volume 2004, pages 37–41. Society for Imaging Science and Technology, 2004.

9. Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2018.