Capstone Proposal

Hsin-Wei Wang
29 June 2019

Domain Background

Skin cancer is one of the most common cancers worldwide, and it has two main types: melanoma and non-melanoma. Basal cell carcinoma and squamous cell carcinoma are the most common non-melanoma tumors accounting for 90% of skin cancers [1]. Melanoma is much more lethal than non-melanoma. It accounts for less than 5% of skin cancers in the United States but causes about 75% skin-cancer-related deaths [2]. The incidence rate of skin cancer is increasing worldwide and has become a significant economic burden for public health services [3]. From 2006 to 2015, the incidence rate of melanoma increased by 3% per year for people ages 50 and older but was stable for those younger than 50. The new cases and deaths of melanoma in the US in 2019 are estimated at 96,480 and 7,230 respectively [4]. Both melanoma and non-melanoma cancers are highly curable if the cancer is diagnosed and treated in its early stage. The estimated 5-year relative survival rate for patients whose melanoma is detected early is about 98%, but the survival rate drops to 23% if malignant cancer metastasizes to parts of the body remote from the primary tumor [4]. Therefore, a fast, automated and accurate diagnosis may help dermatologists and patients detect skin cancer earlier and lead to a high chance of survival.

To facilitate the application of skin images to help reduce skin cancer mortality, the International Skin Imaging Collaboration (ISIC) has created the ISIC Archive [5-7], the largest publicly available collection of quality controlled dermoscopic images of skin lesions as a benchmark for education and research. Since 2016, they have organized the annual "ISIC: Skin Lesion Analysis Towards Melanoma Detection" challenge including problems such as lesion segmentation, lesion attribute detection, and disease classification. These challenges have attracted global participation and encouraged novel automated algorithms.

Problem Statement

The ISIC 2019 challenge contains two tasks: 1) classify dermoscopic images without meta-data, and 2) classify images with additional meta-data [8]. In this proposal, I will only focus on Task 1. The goal of both tasks is to classify dermoscopic images into 9 different diagnostic categories as shown in Table 1. There are 25,331 images across only 8 different categories for training, but the test dataset (released on 2 August 2019) will contain an additional outlier category not represented in the training data. Therefore, the proposed algorithm should address the problem of detecting out-of-distribution images. This task is more difficult than the classification task in ISIC 2018 challenge which only has 7 possible disease categories and no need to worry about outlier class [9].

Datasets and Inputs

The dermoscopic images of ISIC Archive were compiled from a variety of devices within internationally leading clinical centers to ensure clinically representative. Each image is annotated as one of the 8 categories in the training dataset, and this dataset will be used to train deep neural networks for classifying test images in Task 1. The images are named using the

format ISIC_#.jpg where the symbol # is a 7-digit unique identifier. All dermoscopic images are in 24-bit (8 bits per channel) JPEG format and EXIF tags within the images have been removed. Dimensions of these images are inconsistent ranging from 600 × 450 to 1024 × 1024.

Table 1. The number of images for each category in the training dataset.

Diagnostic Category	Amount
Melanoma (MEL)	4522
Melanocytic nevus (NV)	12875
Basal cell carcinoma (BCC)	3323
Actinic keratosis (AK)	867
Benign keratosis (solar lentigo / seborrheic keratosis / lichen planus-like keratosis) (BKL)	2624
Dermatofibroma (DF)	239
Vascular lesion (VASC)	253
Squamous cell carcinoma (SCC)	628
None of the others (UNK)	0

Solution Statement

Previous studies have shown that skin lesion classifiers based on convolutional neural networks (CNNs) achieved performance on par with dermatologists [10]. Presented CNN methods that use transfer learning and parameters fine-tuning for skin lesion classification are the most common and best-performing approaches. Therefore, I will also adopt a similar approach using transfer learning to train a CNN with models pre-trained on the ImageNet dataset such as ResNet, Densenet and Xception, and fine-tune parameters on the training dataset provided in the ISIC 2019 challenge. During training, the input to the CNN is dermoscopic images and diagnostic category labels. The last layer of the CNN is a softmax layer so that it outputs a probability distribution over diagnostic categories.

An intuitive solution for detecting an additional outlier class on the test dataset might be to expand the training dataset by collecting or generating out-of-distribution examples. However, the scope of the outlier is not specified clearly, therefore the number of possible out-of-distribution examples can be extremely large and difficult to compile. To tackle this problem, I plan to leverage a simple and easily implementable method, ODIN (**O**ut-of-**DI**stribution detector for **N**eural networks) [11], for distinguishing in- and out-of-distribution samples. ODIN is based on two techniques, temperature scaling and input perturbation. This method does not require retraining the neural network and has been proven effective on several different network architectures. Each dermoscopic image firstly will be identified whether it is out-of-distribution by ODIN. An image is classified as the outlier class if it is regarded as out-of-distribution, otherwise the CNN mentioned above is used to classify the image as one of the 8 categories in the training dataset. There are 3 parameters of ODIN, T (temperature scaling), ε (perturbation magnitude) and δ (threshold) need tuning. I have collected an out-of-distribution dataset of 138 images from ISIC Archive to tune these parameters. These images were filtered by selecting 7 items of the lesion diagnosis element as shown in Figure 1.

The dataset will be split into 80% training and 20% validation set. In order to prevent overfitting and make the model generalize better, I'll perform image augmentation by random affine transformations and changing brightness, saturation and contrast randomly. Table 1 shows

that the number of images for each category is imbalanced. I'll use the class weighted loss function to deal with the imbalance problem.

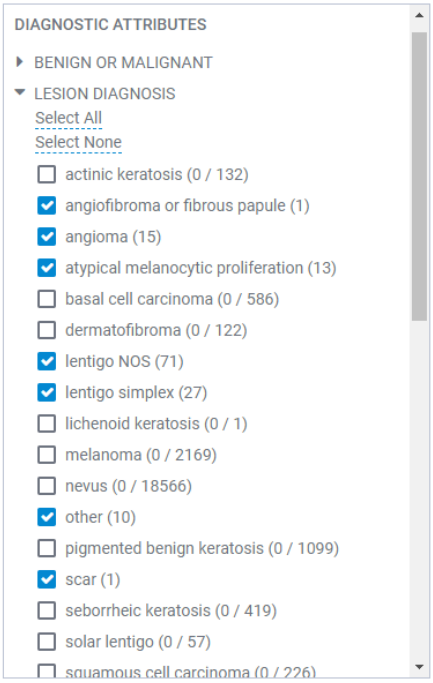

Figure 1. Selected lesion diagnosis types for the out-of-distribution dataset.

Benchmark Model

There were 139 algorithms created by 77 machine-learning labs participating in the ISIC 2018 [12]. The ISIC 2019 challenge is an open and ongoing contest. Each algorithm submitted to this challenge can be regarded as a benchmark model. According to the challenge last year, there should also be a live leaderboard this time, and the final winners will be announced on 23 August 2019. Since there won't be any published results available for comparison until test dataset released, I'll create a simple vanilla CNN as benchmark model which has 3 convolutional layers, and each is followed by a max pooling layer. A global average pooling layer is added after the last pooling layer and followed by a dense layer which has 8 nodes corresponding to the 8 categories in the training dataset. The input images will be resized to 224 × 224 pixels to be consistent with most ImageNet pre-trained models. The vanilla CNN will also integrate with the ODIN method to detect the outlier category like the solution mentioned in the Solution Statement section.

Evaluation Metrics

There is no announcement about the evaluation metric on the ISIC 2019 website [8], but the metric used for the classification task of the ISIC 2018 challenge should be a good reference [9]. In the last ISIC challenge, predicted responses were ranked based only on the normalized multiclass accuracy metric (balanced across categories). More precisely, it calculates accuracy on a per class basis and then averages those accuracies. The area under the receiver operating characteristic curve (AUC) was the second metric to break tied positions.

Project Design

High-Level solution design is mentioned in the Solution Statement section which can be roughly separated into two main steps: 1) Create a CNN to classify in-distribution samples, 2) Tune ODIN parameters to detect out-of-distribution samples. The submission deadline for the task is 9 August 2019. Since I only have very limited computational resources and there is not much time left to complete the project, I plan on using only three pre-trained models listed in Table 2 for transfer learning. The 4th-rank method [13] in the ISIC 2018 challenge have tested Adam, Nadam, and RMSprop optimizers with a batch size of 40 and found that Adam generally performed best. Therefore, each of our models will be trained by Adam optimizer with a batch size of 40. In terms of learning rate tuning, I'll follow the learning rate strategy of [13] which is a step decay schedule with a starting learning rate of 0.0005. After 50 epochs, it drops the learning rate by a factor of 0.2 and continues reducing the learning rate with the same factor every 25 epochs. Each model is trained for 125 epochs, and the model that attains the best normalized multiclass accuracy on the validation set will be saved.

Table 2. CNN architectures for transfer learning.

Architecture	Input Size	Parameters
Xception	299 × 299	22.9 M
DenseNet201	224 × 224	20.2 M
NASNet-Large	331 × 331	88.9 M

References

[1] W. C. R. F. International. "Skin cancer." https://www.wcrf.org/dietandcancer/skin-cancer (accessed 06/19, 2019).

[2] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," (in eng), *Nature,* vol. 542, no. 7639, pp. 115-118, 02 2017, doi: 10.1038/nature21056.

[3] Z. Apalla, A. Lallas, E. Sotiriou, E. Lazaridou, and D. Ioannides, "Epidemiological trends in skin cancer," (in eng), *Dermatol Pract Concept,* vol. 7, no. 2, pp. 1-6, Apr 2017, doi: 10.5826/dpc.0702a01.

[4] A. C. Society. "Cancer Facts and Figures 2019." https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2019.html (accessed 06/19, 2019).

[5] "ISIC Archive." https://isic-archive.com/ (accessed 06/20, 2019).

[6] N. C. F. Codella *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)*, 4-7 April 2018 2018, pp. 168-172, doi: 10.1109/ISBI.2018.8363547.

[7] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data,* Data Descriptor vol. 5, p. 180161, 08/14/online 2018, doi: 10.1038/sdata.2018.161.

[8] I. S. I. Collaboration. "ISIC 2019." https://challenge2019.isic-archive.com/ (accessed 06/20, 2019).

[9] I. S. I. Collaboration. "ISIC 2018." https://challenge2018.isic-archive.com/ (accessed 06/22, 2019).

[10] T. J. Brinker *et al.*, "Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review," (in eng), *J Med Internet Res,* vol. 20, no. 10, p. e11936, 10 2018, doi: 10.2196/11936.

[11] S. Liang, Y. Li, and R. Srikant, "Enhancing The Reliability of Out-of-distribution Image Detection in Neural Networks," in *ICLR*, 2018.

[12] P. Tschandl *et al.*, "Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study," (in eng), *Lancet Oncol,* Jun 2019, doi: 10.1016/S1470-2045(19)30333-X.

[13] N. Gessert *et al.*, "Skin Lesion Diagnosis using Ensembles, Unscaled Multi-Crop Evaluation and Loss Weighting," *ArXiv,* vol. abs/1808.01694, 2018.