The AAAI-17 Joint Workshop on
Health Intelligence
WS-17-09

# A Deep Multi-Task Learning Approach to Skin Lesion Classification

**Haofu Liao, Jiebo Luo**
Department of Computer Science, University of Rochester
Rochester, New York 14627, USA
Email: {hliao6, jluo}@cs.rochester.edu

## Abstract

Skin lesion identification is a key step toward dermatological diagnosis. When describing a skin lesion, it is very important to note its body site distribution as many skin diseases commonly affect particular parts of the body. To exploit the correlation between skin lesions and their body site distributions, in this study, we investigate the possibility of improving skin lesion classification using the additional context information provided by body location. Specifically, we build a deep multi-task learning (MTL) framework to jointly optimize skin lesion classification and body location classification (the latter is used as an inductive bias). Our MTL framework uses the state-of-the-art ImageNet pretrained model with specialized loss functions for the two related tasks. Our experiments show that the proposed MTL based method performs more robustly than its standalone (single-task) counterpart.

## Introduction

Visual aspects of skin diseases, especially skin lesions, play a key role in dermatological diagnosis. A successful identification of the skin lesion allows skin disorders to be placed in certain diagnostic categories where specific diagnosis can be established (Cecil, Goldman, and Schafer 2012). However, categorization of skin lesions is a challenging process. It usually involves identifying the specific morphology, distribution, color, shape and arrangement of lesions. When these components are analyzed separately, the differentiation of skin lesions can be quite complex and requires a great deal of experience and expertise (Lawrence and Cox 2002).

Besides the high requirement of expertise, the categorization of skin lesions by human is essentially subjective and not always reproducible. To achieve a more objective and reliable lesion recognition and ease the process of dermatological diagnosis, a computer-aided skin lesion classification system should be considered. Traditional approaches to computer-aided skin lesion/disease classification usually focus on certain types of skin diseases, such as melanoma and basal cell carcinoma, where the visual aspects of skin lesions are more regular and predictable. In those cases, human-engineered feature extraction algorithms can be easily developed. However, when we extend the lesion types to a broader range where all the possible combinations of lesional characteristics need to be considered, human-engineered feature extraction becomes infeasible and the traditional approaches fail.

Deep convolutional neural networks (CNNs) have shown to be very successful in recent years. Specifically, the vision challenges from ILSVRC (Russakovsky et al. 2015) and MS COCO (Lin et al. 2014) show that contemporary CNN architectures are able to surpass human in many vision tasks. One thing behind the success of CNN is its ability to do feature engineering automatically from a large-scale dataset. It has been reported by many studies (Razavian et al. 2014; Donahue et al. 2014; Zeiler and Fergus 2014) that features extracted by contemporary CNNs yield consistent superior results compared to the highly tuned non-CNN counterparts in many tasks. Therefore, in this study, we propose to develop a skin lesion classification model based on the state-of-the-art CNN architectures.

However, instead of treating the skin lesion classification as a standalone problem and training a CNN model using skin lesion labels only, we further propose to jointly optimize the skin lesion classification with a related auxiliary task, body location classification. The motivation behind this design is to make use of the body site predilection of skin diseases (Cox and Coulson 2004) as it has long been recognized by dermatologists that many skin diseases and their corresponding skin lesions are correlated with their body site manifestation. For example, a skin lesion caused by sun exposure is only present in sun-exposed areas of the body (face, neck, hands, arms) (Cecil, Goldman, and Schafer 2012). Therefore, a CNN architecture that can exploit the domain-specific information contained in the body locations should be intuitively helpful in improving the performance of our skin lesion classification model.

In this study, we present a multi-task learning framework for universal skin lesion (all lesion types) classification using deep convolutional neural networks. In order to learn a wide variety of visual aspect of skin lesions, we first collect 21657 images from DermQuest (www.dermquest.com), a public skin disease atlas contributed by dermatologists around the world. We then formulate our model into a dual-task based learning problem with specialized loss functions for each task. Next, to boost the performance, we fit our model into the state-of-the-art deep residual network (ResNet) (He et

al. 2015) which is the winning entry of ILSVRC 2015 (Russakovsky et al. 2015) and MS COCO 2015 (Lin et al. 2014).
**Contribution:** To our best knowledge, this is the first attempt to target the universal skin lesion classification problem systematically using a deep multi-task learning framework. We show that the jointly learned representations from body locations indeed facilitate the learning for skin lesion classification. Using the state-of-the-art CNN architecture and combining the results from different models we can achieve as high as a 0.80 mean average precision (mAP) in classifying skin lesions.

## Related Work

Most of the existing works (Arroyo and Zapirain 2014; Xie et al. 2014; Fabbrocini et al. 2014) only focus on one or a few skin disease and solve the problem using conventional machine learning approach, i.e., extracting manually engineered features from segmented lesion patches and classifying with a linear classifier such as SVM. While in our study, we target a more challenging problem where all skin diseases are considered.

Many CNN related approaches have been proposed to solve dermatology problems in recent years. Some works (Cruz-Roa et al. 2014; Wang et al. 2014; Arevalo et al. 2015) used CNNs as an unsupervised feature extractor and detect mitosis, an indicator of cancer, from histopathology images. (Esteva, Kuprel, and Thrun 2015) presented a CNN architecture for diagnosis-targeted skin disease classification. They trained their model with a contemporary CNN architecture using a large-scale dataset (23000 images). Similar to our study, they also tried to classify skin diseases in a broader range. What sets us apart from their work is instead of training with diagnosis labels and making diagnostic decision directly, our work classifies skin diseases by their lesional characteristics. According to a recent study (Liao, Li, and Luo 2016), skin lesion is proven to be a more appropriate subject for skin disease classification as many diagnoses can not be distinguished visually. Recently, (Kawahara, BenTaieb, and Hamarneh 2016) also proposed a CNN based model to classify skin lesions for non-dermoscopic images. However, they only managed to build their model on a prior art CNN architecture with a relatively small dataset (1300 images).

Multi-task learning (MTL) (Caruana 1997) is an approach to learning a main task together with other related tasks in parallel with the goal of a better generalization performance. Learning multiple tasks jointly has been proven to be very effective in many computer vision problems, such as attribute classification (Hand and Chellappa 2016), face detection (Ranjan, Patel, and Chellappa 2016), face alignment (Zhang et al. 2016) and object detection (Ren et al. 2015). However, we find no multi-task learning based algorithm has been developed for dermatology related problems.

## Dataset

All the dermatology images used in this study are collected from DermQuest. We choose DermQuest against other dermatology atlantes is because it has the most detailed annotations and descriptions for each of the dermatology image and it is the only public dermatology atlas that contains both skin lesion and body location labels. Most of the dermatology images from DermQuest are contributed by individual dermatologists. When contributing an image, they are required to input the descriptions (diagnosis, primary lesions, body location, pathophysiology, etc.) by their own. As the terminology used by dermatologists are not unified, they may use different terms and morphologies when describing a dermatology image which results in an inconsistency of DermQuest images.

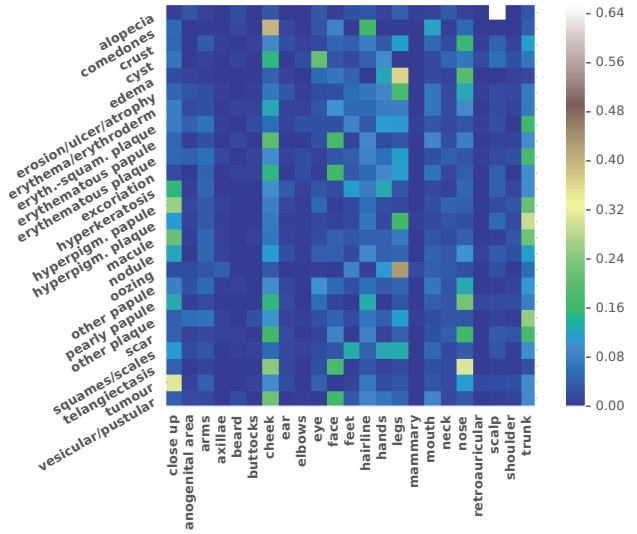


Figure 1: The correlation matrix between skin lesion and body location. Each row denotes a skin lesion and each column denotes a body location. A cell at $(i, j)$ denotes the proportion of the images with both label $i$ and label $j$ among all the $i$ images (best viewed in color).

Due to the inconsistency, there are 180 lesion types in total in the DermQuest atlas , which is larger than any of the existing lesion morphology. Therefore, with the help of a dermatologist, we refined the list of lesion types to make sure they reasonably and consistently represent the lesional characteristics of the images in DermQuest. We refine and merge lesions based on the lesion morphology described in (Cox and Coulson 2004) with some modifications: 1) We removed those infrequent lesion types (less than 10 images) as they do not have enough images for our model to learn some meaningful features. 2) For some popular (greater than 1000 images) sublesion types, such as hyperpigmented papule lesion under the papule family, we do not merge them as there are enough images in the dataset so that our model can distinguish them from other sublesions under the same family. 3) Some of the lesion types have common visual characteristics, such atrophy, erosion and ulcer, we also merge them together. After the refinement, we finally come up with a lesion morphology list with 25 lesions types for the DermQuest images. Note that there might be multiple lesion labels associated with an image as a skin disease usually manifests different lesional characteristics at a time.

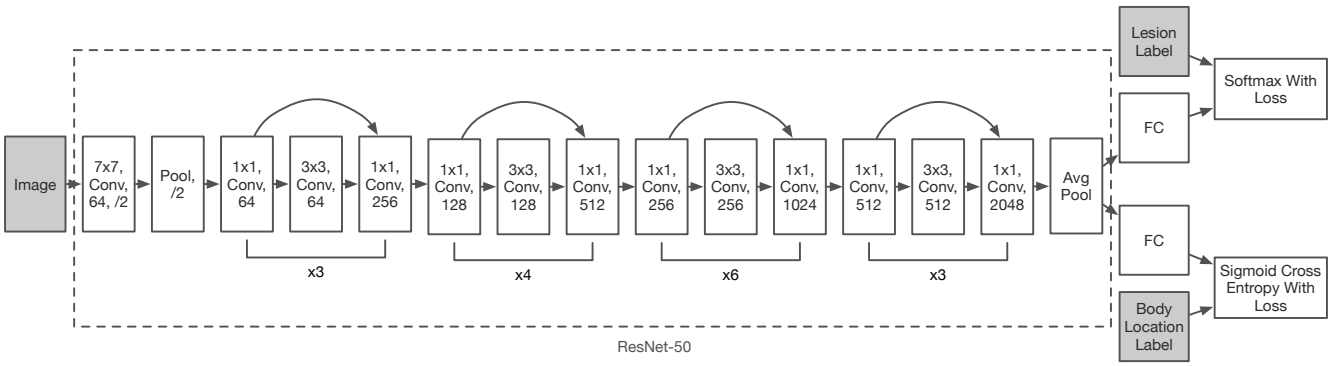


Figure 2: The network structure of the proposed method. "Conv" denotes the convolutional layer, "Pool" denotes the pooling layer and "FC" denotes the fully connected layer. The three dark blocks are the data layers for images, skin lesions, and body locations, respectively. The net architecture inside the dotted area is identical to the ResNet-50 network.

For the body location labels, the terminology used is more consistent. We do not modify too much except we removed those infrequent labels as we did for the lesions. We also merged some body locations that are too specific to not be mixed with its nearby regions in an image. For example, an image labeled with nails usually contains parts of the fingers. Thus, it is actually hard to tell whether it should be labeled with nails or fingers. Hence, we directly merge them into the "hands" category. There are 23 body locations in the final list.

We also investigate the correlation between skin lesions and body locations among images in DermQuest. The correlation map is shown in Figure 1. Here, each row denotes a skin lesion and each column denotes a body location. Let $N_i$ denote the total number of images in our dataset that has lesion $i$ and $M_j$ denote the total number of images that has body location $j$. Then, the cell at $(i, j)$ can be computed by

$$R_{ij} = \frac{N_i \cap M_j}{N_i} \tag{1}$$

Thus, if a skin lesion frequently appears on certain body location, we will see a high very value of $R_{ij}$. On the other hand, if a skin lesion has no specific predilection of body locations, then it will appear randomly at different body locations. Notice that we have 23 body location types. Thus, the randomness means cells in the corresponding row should have values close to $1/23$, i.e., dark blue in the color bar. For example, the cells in row "erythema/erythroderm" are almost in blue, which means "erythema/erythroderm" has little body location predilection. This is consistent with our knowledge that "erythema/erythroderm" is a very commonly seen lesion that can exists anywhere in the body. We can also see that "alopecia" is highly correlated with "scalp". It makes sense as "alopecia" is a lesion that related with hair loss.

## Methodology

### Deep Multi-task Learning

To jointly optimize the main (skin lesion classification) and auxiliary (body location classification) tasks, we formulate our problem as follows. Let $(\mathbf{X}_i, \mathbf{u}_i, v_i), i \in \{1, \dots N\}$ denotes the $i$th data in the training set, where $\mathbf{X}_i$ is the $i$th image and $\mathbf{u}_i$ and $v_i \in \{1, \dots, Q\}$ are the $i$th labels for the skin lesion and body location, respectively. As multiple lesion types may be associated with a dermatology image, the lesion label $\mathbf{u}_i = [u_1^i, u_2^i, \dots, u_P^i]$ is a binary vector with

$$u_j^i = \begin{cases} 1, & \text{if the } j\text{th lesion is associated with } \mathbf{X}_i, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Here, $P$ and $Q$ denotes the number of skin lesions and body locations in our dataset. Our goal is to minimize the objective function

$$\begin{aligned} \mathcal{L}(\mathbf{W}) = & \frac{1}{N} \sum_{i=1}^{N} \ell_{les}(\mathbf{X}_i, \mathbf{u}_i; \mathbf{W}) + \\ & \frac{1}{N} \sum_{i=1}^{N} \ell_{loc}(\mathbf{X}_i, v_i; \mathbf{W}) + \Phi(\mathbf{W}) \end{aligned} \tag{3}$$

in which $\mathbf{W}$ is the parameters of CNN, $\Phi(\cdot)$ is a regularization term, $\ell_{les}(\cdot)$ is the loss function for skin lesions and $\ell_{loc}(\cdot)$ is the loss function for body locations.

Since there might be multiple lesions associated with an input image, we use a sigmoid cross-entropy function for the skin lesion loss so that each lesion can be optimized independently. Let $s_j(\mathbf{X}_i; \mathbf{W}), j \in \{1, \dots, P\}$ denotes the $j$th output of the last fully-connected (FC) layer for the skin lesions. Then the $j$th activation of the sigmoid layer can be written as

$$a_j(\mathbf{X}_i; \mathbf{W}) = \frac{1}{1 + e^{-s_j(\mathbf{X}_i; \mathbf{W})}}. \tag{4}$$

and the corresponding cross-entropy loss is

$$\begin{aligned} \ell_{les}(\mathbf{X}_i, \mathbf{u}_i; \mathbf{W}) = & -\sum_{j=1}^{P} u_j^i \log a_j(\mathbf{X}_i; \mathbf{W}) + \\ & (1 - u_j^i) \log \left(1 - a_j(\mathbf{X}_i; \mathbf{W})\right). \end{aligned} \tag{5}$$

For the body locations, it is a many-one classification problem. Thus, we use a softmax loss function so that only

one label will be optimized each time. Let $t_j(\mathbf{X}_i; \mathbf{W}), j \in \{1, \ldots, Q\}$ denotes the $j$th output of the last FC layer for the body locations. Then the $j$th activation of the softmax layer can be written as

$$b_j(\mathbf{X}_i; \mathbf{W}) = \frac{e^{t_j(\mathbf{X}_i; \mathbf{W})}}{\sum_k e^{t_k(\mathbf{X}_i; \mathbf{W})}} \tag{6}$$

and the corresponding softmax loss is

$$\ell_{loc}(\mathbf{X}_i, v_i; \mathbf{W}) = -\log(b_{v_i}(\mathbf{X}_i; \mathbf{W})) \tag{7}$$

Finally, for the regularization term, we use the L2 norm

$$\Phi(\mathbf{W}) = \gamma \|\mathbf{W}\|_2 \tag{8}$$

where the regularization parameter $\gamma$ controls the trade off between the regularization term and the loss functions.

### Implementation

The architecture of the proposed method is given in Figure 2. We build our CNN architecture on top of ResNet-50 (50 layers). Though it is possible to use a deeper ResNet to get a marginal performance gain, ResNet-50 is considered sufficient for this proof-of-concept study. To facilitate our goal in MTL, three data layers are used. One data layer is for the images and the other two data layers are for the lesion labels and body location labels, respectively. We then remove the finally FC layer in the original ResNet and add two sibling FC layers, one for the skin lesions and the other for the body locations. After the FC layers, we add a sigmoid cross entropy loss layer for the skin lesion classification and a softmax layer for the body location classification.

We use the Caffe deep learning framework (Jia et al. 2014) for all of our experiments and run the programs with a GeForce GTX 1070 GPU. As transfer learning has shown to be more effective in image classification problems (Razavian et al. 2014), instead of training from scratch, we initialize our network from the ImageNet (Deng et al. 2009) pretrained ResNet-50 model [1]. As a dermatology image may be taken from different distances, the scale of certain skin lesions may vary. Thus, following the practice in (Simonyan and Zisserman 2014), we scale each image with its shorter side length randomly selected from $[256, 480]$. This process is called scale jittering. Then we follow the ImageNet practice in which a 224 x 224 crop is randomly sampled from the mean subtracted images or their horizontal flips. In the testing phase, we perform the standard 10-crops testing using the strategy from (Krizhevsky, Sutskever, and Hinton 2012).

For the hyper-parameters, we use SGD with a mini-batch size of 20 and set the momentum to $0.9$ and the weight decay (the regularization parameter) to $0.0001$. The initial learning rate is $0.001$ and is reduced by $0.1$ when error plateaus. It is worth mentioning that the two newly added FC layers have bigger learning rate multipliers (10 for the weights and 20 for the bias) so that their effective learning rate is actually $0.01$. We use higher learning rate for these two layers is because their weights are randomly initialized. The model is trained for up to $12 \times 10^4$ iterations. Note that this is a relatively large number for fine-tuning. This is because the scale jittering greatly augmented our dataset and it takes longer time for the training to converge. During the training, we do not see any over-fitting from the validation set.

[1] We also trained the network from scratch but no performance gain was observed.

## Experimental Results

In this section, we investigate the performance of the proposed method on both the skin lesion classification and body location classification tasks. In all of our experiments, we use data collected from DermQuest. In total, there are $21657$ images that contain both the skin lesion and body location labels. To avoid overfitting, 5-folds cross-validation is used for each experiment.

### Performance of Skin Lesion Classification

For skin lesion classification, since it is a multi-label classification problem, we use mean average precision (mAP) as the evaluation metrics following the practice in VOC (Everingham et al. 2010) and ILSVRC. In this study, we use two different mAPs: 1) class-wised mAP, where we treat the sorted evaluations of all images on certain class as a ranking and compute the mAP over the classes. 2) image-wised mAP, where we treat the sorted evaluations of all classes on certain image as a ranking and compute the mAP over the images. Formally put, the two metrics can be computed using the following formulas:

$$\text{mAP-class} = \frac{1}{P} \sum_{i=1}^{P} \sum_{j=1}^{N} p_i(j) \Delta r_i(j), \tag{9}$$

$$\text{mAP-image} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{P} q_i(j) \Delta s_i(j), \tag{10}$$

Here, $N$ is the total number of images, $P$ is the total number of classes, $p_i(j)$ is the precision of the ranking for class $i$ at cut-off $j$ and $\Delta r_i(j)$ is the difference of the recall (of the ranking for class $i$) from cut-off $j-1$ to $j$. $q_i(j)$ and $\Delta s_i(j)$ can be defined similarly to $p_i(j)$ and $\Delta r_i(j)$.

We compare our proposed method with two standalone architectures (single task) based on AlexNet and ResNet-50, respectively. For the hyper-parameters of AlexNet, we use the settings from (Krizhevsky, Sutskever, and Hinton 2012), i.e., batch size $= 256$, momentum $= 0.9$ and weight decay $= 0.0005$. For the standalone ResNet-50, we use the same hyper-parameter settings as our proposed method. Both the two architectures are fine-tuned from ImageNet pretrained models with learning rate set to $0.01$.

The classification results are shown in Table 1. Here, "AlexNet" and "ReNet" are the two standalone architectures, "MTL" is our proposed method, and "Ensemble" contains the ensemble results of "ResNet" and "MTL". First, we can see "ResNet" outperforms "AlexNet" by a big leap which shows that the use of the state-of-the-art CNN architecture helps a lot in boosting the performance. Then, we also observe a decent performance improvement against "ResNet" when using our proposed method. It means the

| Lesion Type | Average Precision | | | |
|---|---|---|---|---|
| | AlexNet | ResNet | MTL | Ensemble |
| alopecia | 0.763 | 0.845 | 0.843 | **0.855** |
| comedones | 0.687 | 0.817 | **0.861** | 0.858 |
| crust | 0.677 | 0.783 | 0.794 | **0.807** |
| cyst | 0.461 | 0.625 | 0.698 | **0.702** |
| edema | 0.633 | 0.707 | 0.751 | **0.758** |
| erosion/ulcer/atrophy | 0.774 | 0.850 | 0.867 | **0.873** |
| erythema/erythroderm | 0.742 | 0.820 | **0.844** | 0.843 |
| eryth.-squam. plaque | 0.496 | 0.658 | 0.683 | **0.690** |
| erythematous papule | 0.767 | 0.846 | 0.857 | **0.861** |
| erythematous plaque | 0.538 | 0.670 | 0.704 | **0.708** |
| excoriation | 0.467 | 0.605 | 0.635 | **0.651** |
| hyperkeratosis | 0.643 | 0.772 | 0.796 | **0.802** |
| hyperpig. papule | 0.589 | 0.690 | **0.738** | 0.730 |
| hyperpig. plaque | 0.473 | 0.637 | **0.675** | **0.675** |
| macule | 0.619 | 0.742 | **0.780** | 0.777 |
| nodule | 0.704 | 0.793 | 0.813 | **0.820** |
| oozing | 0.497 | 0.595 | **0.674** | 0.663 |
| other papule | 0.344 | 0.559 | 0.600 | **0.603** |
| pearly papule | 0.716 | 0.849 | 0.875 | **0.879** |
| other plaque | 0.331 | 0.553 | 0.549 | **0.562** |
| scar | 0.521 | 0.690 | **0.728** | 0.726 |
| squames/scales | 0.591 | 0.704 | **0.748** | 0.746 |
| telangiectasis | 0.655 | 0.821 | 0.837 | **0.848** |
| tumour | 0.598 | 0.728 | 0.768 | **0.770** |
| vesicular/pustular | 0.664 | 0.792 | 0.814 | **0.823** |
| mAP-class | 0.598 | 0.726 | 0.757 | **0.761** |
| mAP-image | 0.704 | 0.778 | 0.792 | **0.798** |

Table 1: Skin lesion classification results. "AlexNet" and "ResNet" are trained using skin lesion labels only. "MTL" is the proposed method. An ensemble of "ResNet" and "MTL" is given under "Ensemble".

joint optimization with body location classification can really benefit the learning of the lesional characteristics. Finally, we find that the highest mAP can be achieved with an ensemble of "ResNet" and "MTL", i.e., choosing the best evaluation scores of the two models for each image.

We further analyze the performance difference of each class between "ResNet" and "MTL". We find that, in general, if a skin lesion has a strong correlation with a body location, it will also have a better performance gain when using "MTL". Typical examples are "comedone", "edema", "hyperpigmented papule", "oozing", and "tumor". They all have a strong correlation with certain body locations and we see they also have at least a 4% improvement when using "MTL". However, there are some exceptions. For example, we do not see any improvement from "alopecia" even though it has a very strong correlation with "scalp". One possible reason is that the strong correlation between "alopecia" and "scalp" makes "scalp" bias too much to "alopecia" such that some variations won't be learned. We will further verify this hypothesis in the later discussion.

### Performance of Body Location Classification

We also compare the performance of our method with its standalone counterpart in classifying body locations. To this end, we fine-tune another ResNet-50 model with body location labels only. For the evaluation metrics, the standard top-1 and top-3 accuracies are used as body location classification is a multi-class classification problem. The evaluation results are given in Figure 3. We can also see a performance improvement from "ResNet" to "MTL". This is somewhat counter-intuitive as the classification of a body location should have nothing to do with the skin lesions. However, as we restrict the images to be dermatological images, a slight performance gain is reasonable.

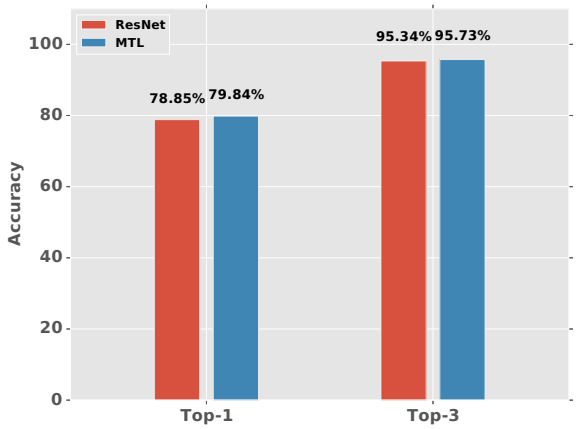


Figure 3: Body location classification results. "ResNet" is trained using body location only and "MTL" is the proposed multi-task learning method.

## Conclusions

We have developed a deep multi-task learning framework for universal skin lesion classification. The proposed method learns skin lesion classification and body location classification in parallel based on the state-of-the-art CNN architecture. To be able to learn a wide variety of lesional characteristics and classify all kinds of lesion types, we have also collected and built a large-scale skin lesion dataset using images from DermQuest. The experimental results have shown that 1) Training using the state-of-the art CNN architecture on a large scale of skin lesion dataset leads to a universal skin lesion classification system with good performance. 2) It is indeed beneficial to use the body location classification as an auxiliary task and train a deep multi-task learning based model to achieve improved skin lesion classification. 3) An ensemble of the proposed method and its standalone counterpart can achieve an image-wise mAP as high as 0.80. 4) The performance of body location classification is also improved under the deep multi-task learning framework. Our future work includes integrating the image analysis with other patient information to build an overall high-performance diagnosis system for diseases with skin lesion symptoms.

## Acknowledgments

This work was supported in part by New York State through the Goergen Institute for Data Science at the University of

Rochester. We thank VisualDX for discussions related to this work.

## References

Arevalo, J.; Cruz-Roa, A.; Arias, V.; Romero, E.; and González, F. A. 2015. An unsupervised feature learning framework for basal cell carcinoma image analysis. *Artificial intelligence in medicine*.

Arroyo, J., and Zapirain, B. 2014. Automated detection of melanoma in dermoscopic images. In Scharcanski, J., and Celebi, M. E., eds., *Computer Vision Techniques for the Diagnosis of Skin Cancer*, Series in BioEngineering. Springer Berlin Heidelberg. 139–192.

Caruana, R. 1997. Multitask learning. *Machine Learning* 28(1):41–75.

Cecil, R. L.; Goldman, L.; and Schafer, A. I. 2012. *Goldman's Cecil Medicine*. Philadephia: Elsevier/Saunders, 23th edition.

Cox, N., and Coulson, I. 2004. Diagnosis of skin disease. *Rook's Textbook of Dermatology, 7th edn. Oxford: Blackwell Science* 5.

Cruz-Roa, A.; Basavanhally, A.; González, F.; Gilmore, H.; Feldman, M.; Ganesan, S.; Shih, N.; Tomaszewski, J.; and Madabhushi, A. 2014. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In *SPIE Medical Imaging*, 904103–904103. International Society for Optics and Photonics.

Deng, J.; Dong, W.; Socher, R.; Li, L.; Li, K.; and Li, F. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA*, 248–255.

Donahue, J.; Jia, Y.; Vinyals, O.; Hoffman, J.; Zhang, N.; Tzeng, E.; and Darrell, T. 2014. Decaf: A deep convolutional activation feature for generic visual recognition. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, 647–655.

Esteva, A.; Kuprel, B.; and Thrun, S. 2015. Deep networks for early stage skin disease and skin cancer classification.

Everingham, M.; Gool, L. J. V.; Williams, C. K. I.; Winn, J. M.; and Zisserman, A. 2010. The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision* 88(2):303–338.

Fabbrocini, G.; Vita, V.; Cacciapuoti, S.; Leo, G.; Liguori, C.; Paolillo, A.; Pietrosanto, A.; and Sommella, P. 2014. Automatic diagnosis of melanoma based on the 7-point checklist. In Scharcanski, J., and Celebi, M. E., eds., *Computer Vision Techniques for the Diagnosis of Skin Cancer*, Series in BioEngineering. Springer Berlin Heidelberg. 71–107.

Hand, E. M., and Chellappa, R. 2016. Attributes for improved attributes: A multi-task network for attribute classification. *CoRR* abs/1604.07360.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2015. Deep residual learning for image recognition. *CoRR* abs/1512.03385.

Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R. B.; Guadarrama, S.; and Darrell, T. 2014. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the ACM International Conference on Multimedia, MM '14, Orlando, FL, USA, November 03 - 07, 2014*, 675–678.

Kawahara, J.; BenTaieb, A.; and Hamarneh, G. 2016. Deep features to classify skin lesions. In *13th IEEE International Symposium on Biomedical Imaging, ISBI 2016, Prague, Czech Republic, April 13-16, 2016*, 1397–1400.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States.*, 1106–1114.

Lawrence, C. M., and Cox, N. H. 2002. *Physical Signs in Dermatology*. London: Mosby, 2nd edition.

Liao, H.; Li, Y.; and Luo, J. 2016. Skin disease classification versus skin lesion characterization: Achieving robust diagnosis using multi-label deep neural networks. In *International Conference on Pattern Recognition (ICPR)*.

Lin, T.; Maire, M.; Belongie, S. J.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft COCO: common objects in context. In *Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V*, 740–755.

Ranjan, R.; Patel, V. M.; and Chellappa, R. 2016. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *CoRR* abs/1603.01249.

Razavian, A. S.; Azizpour, H.; Sullivan, J.; and Carlsson, S. 2014. CNN features off-the-shelf: An astounding baseline for recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2014, Columbus, OH, USA, June 23-28, 2014*, 512–519.

Ren, S.; He, K.; Girshick, R. B.; and Sun, J. 2015. Faster R-CNN: towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, 91–99.

Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M. S.; Berg, A. C.; and Li, F. 2015. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision* 115(3):211–252.

Simonyan, K., and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *CoRR* abs/1409.1556.

Wang, H.; Cruz-Roa, A.; Basavanhally, A.; Gilmore, H.; Shih, N.; Feldman, M.; Tomaszewski, J.; Gonzalez, F.; and Madabhushi, A. 2014. Cascaded ensemble of convolutional neural networks and handcrafted features for mitosis detec-

tion. In *SPIE Medical Imaging*, 90410B–90410B. International Society for Optics and Photonics.

Xie, F.; Wu, Y.; Jiang, Z.; and Meng, R. 2014. Dermoscopy image processing for chinese. In Scharcanski, J., and Celebi, M. E., eds., *Computer Vision Techniques for the Diagnosis of Skin Cancer*, Series in BioEngineering. Springer Berlin Heidelberg. 109–137.

Zeiler, M. D., and Fergus, R. 2014. Visualizing and understanding convolutional networks. In *Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I*, 818–833.

Zhang, Z.; Luo, P.; Loy, C. C.; and Tang, X. 2016. Learning deep representation for face alignment with auxiliary attributes. *IEEE Trans. Pattern Anal. Mach. Intell.* 38(5):918–930.