Skin Lesion Classification Model Enhanced by Optimizing RGB Weights

Zhixiang Yao[1], *BS*, Ramana Balla[1], *BS*, Linke Li[2], *BS*

1. Department of Biomedical Engineering, Duke University, Durham 27708, USA; {zy154; vrb12}@duke.edu
2. Department of Electrical and Computer Engineering, Duke University, Durham, 27708, USA; ll442@duke.edu

Abstract

Pigmented skin lesions identification plays an important role in accurately detecting harmful skin diseases, while image detection and classification techniques are now widely applied to improve the overall accuracy of skin cancer detection. Our project proposed an InceptionV3-based model enhanced by optimizing RGB weights (with WeightedRGBLayer) to classify the skin lesion image dataset HAM10000. Given the imbalanced distribution of the dataset, 100 images were sample from all seven categories, and augmented to construct a new set with 2800 images. A baseline was setup with a batch normalization layer and the InceptionV3 architecture resulting a 92% accuracy. As the WeightedRGBLayer being implemented prior to the InceptionV3 module, algorithm's variance was reduced. Several operations on the images have been applied to the dataset to simulate different imaging conditions, including Gaussian blur and lens distortion. Our RGB-optimizing layer enhanced model performance, indicating better classification ability compared to those without the WeightedRGBLayer.

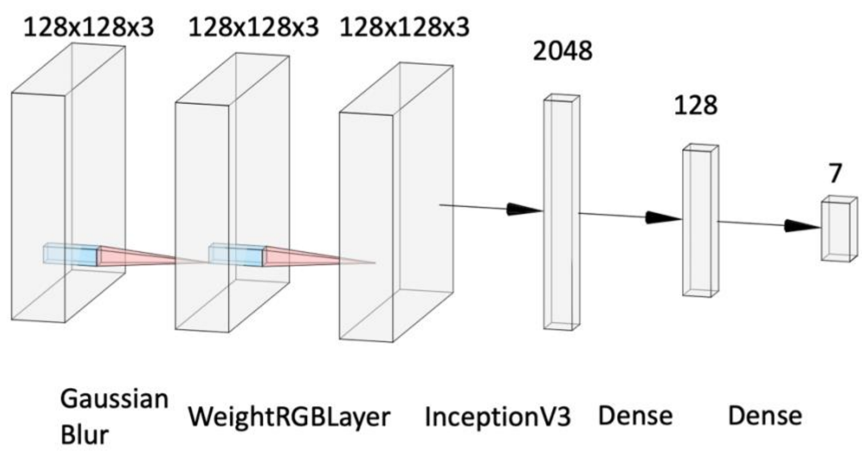


Figure: A schematic of the model. Image inputs are 128*128*3, RGB colored images; the gaussian blur layer simulates the lens blurry; the WeightedRGBLayer reassigns the RGB channel weights; InceptionV3 is selected as base module; final output is a one-hot encoded classification matrix.

Skin Lesion Classification Model Enhanced by Optimizing RGB Weights

1. Introduction

Skin lesion is an abnormal growth, sore or rash that develops on the skin, which can be caused by factors including infections, allergies and exposure to carcinogen. As one of the most widely used skin lesion datasets, HAM10000 dataset covered more than 95% of all pigmented lesion examined in biomedical practice. While most of them are harmless to human health, it is necessary to accurately identify them as early as possible.

Recently, with the increasing use of machine learning and deep learning techniques in biomedical diagnosis, the efficiency of illness prediction increases proportionately. [1] In the proposed work, traditional convolutional neural network (CNN) has been used for feature extraction and image classification [2], [3], while Deep Convolutional Neural Network (DCNN) also helps to classify lesion into dermoscopic images, and thus used for efficient malignant skin lesions detection. Among the DCNN models, Inception-V3 effectively captures both low-level and high-level features of input image, thereby improving its ability to classify objects accurately. [4]

However, traditional deep learning models suffer from high variance due to its non-linearity, so they are sensitive to initial conditions of neural network, such as initial weight and dataset size. [4] Since color is an important factor of image representation, color information extraction during weight initialization helps to improve the overall efficiency of vision task. [5]

Additionally, to fully utilize the details of the dataset, lens distortion can also be introduced as a data augmentation technique to simulate real-world scenarios. [6] By artificially applying distortions into the training data, the model is exposed to a wider range of image variations and can learn to recognize objects more accurately.

Our project proposed a Skin Lesion Classification Model Enhanced by Optimizing RGB Weights and distortion lens to improve the overall image classification capability. Specifically, we first introduced our weighted RGB-layer in front of the inception V3 model to reduce the high variance of the algorithm. Additionally, lens distortion was applied to our model for more robust training. When tested with images under different blur kernels, it indicates better classification result compared with the baseline method with only one batch normalization layer before the InceptionV3 layer.

Related work

Many state-of-the-art algorithms have been proposed for efficient image classification. An approach based on the concept of "divide and conquer" has been proposed by Hameed et al. [7] Pereira and colleagues proposed a method to enhance the performance of skin lesion classification algorithms by integrating gradients with the local binary patterns (LBP) technique. [8] Their approach was designed to utilize the border-line properties of the lesion segmentation mask. VGGNet [9] yielded high performance in the 2014 ILSVRC classification challenge. The Inception architecture of GoogLeNet [10] was also designed to perform well even under strict constraints on memory and computational budget.

Since many deep learning models suffer from high variance and are sensitive to initial condition of neural networks, appropriate weight initialization helps the neural network to sustain stable learning and shorten time of convergence. Zero initialization makes every layer the same in back propagation which is not viable. [4] Deng et al. applied RGB color mode to modify random weights at the beginning of the model, which improves the overall efficiency of training. [11]

Dataset

The HAM10000 dataset [12] is a large set of images of pigmented skin lesions that have been captured using dermoscopic imaging techniques. Skin lesions recorded in this dataset are commonly reported in clinics, and the images were obtained from various sources. HAM10000 is composed of 10015 images with seven categories, including Actinic keratoses (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv), and vascular lesions (vasc). However, this dataset is significantly imbalanced (**Figrue 1**). 66.9% of images are categorized as nv, while df and vasc classes only contributes about 1% each. Thus, it is necessary to proceed the dataset with augmentation strategies in data preprocessing.

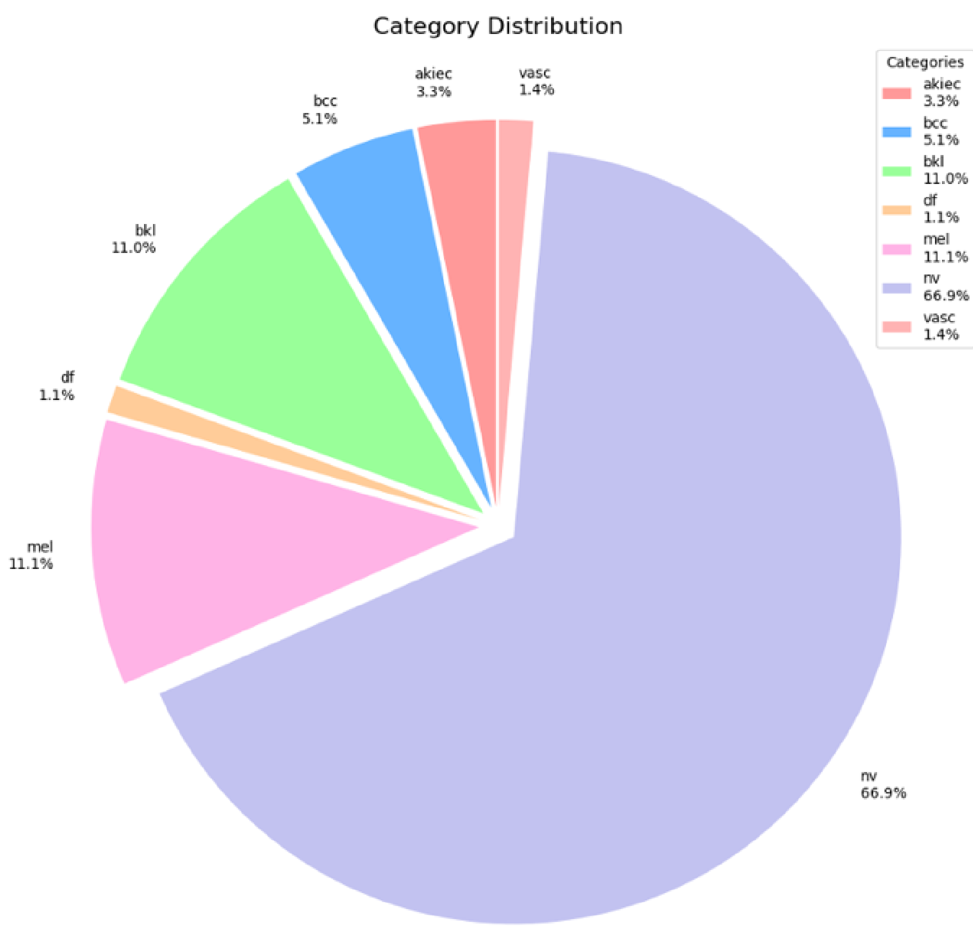


Figure 1: Distribution of different lesion classes in HAM10000

Data preprocessing

Images were resized from 450*600 to 128*128 and 100 images from each category was sampled to form a new dataset. Given a vertical flip, a horizontal flip and a 180-degree rotation on the images, the new dataset is augmented to 2,800 images with 400 images in each category. (An example augmentation output of one image is given in **Appendix 1, Figure 1**)

2. Model Architecture

Our model is essentially based on the InceptionV3 module. To simulate light signals traveling from patient's skin to the lens of the camera, a Gaussian blur layer is added following the input. So as to reweight the color channels, the WeightedRGBLayer was inserted between the Gaussian blur layer and the inceptionV3 module. In addition, the input of the final model as well as the output of Gaussian blur layer and the WeightedRGBLayer were batch normalized to reduce the impact of extreme values generated by the layer. Two softmax functions based dense layers were applied to yield a 7-class final output.

Data splitting

In this project, the 100 images from the 7 categories and their augmented outputs were the main subject of training and final evaluation. The 2,800 images were first split into train and test sets in an 80:20 ratio. A validation set was generated by sampling 10% from the train set. The new train set, without the 10% sampled out for validation, was used for model training. Thus, our test set has 560 images, the validation set has 224 images, and training set has 2016 images.

InceptionV3

InceptionV3 (**Figure 2**) [4], [10] architecture is a simplified version of the GoogLeNet, aiming to simplify the model processing and improve the performance. By stacking several 1 × 1, 3 × 3, 5 × 5 convolution layers within a single module, the architecture makes it possible to process colored 3-channel image dataset efficiently.

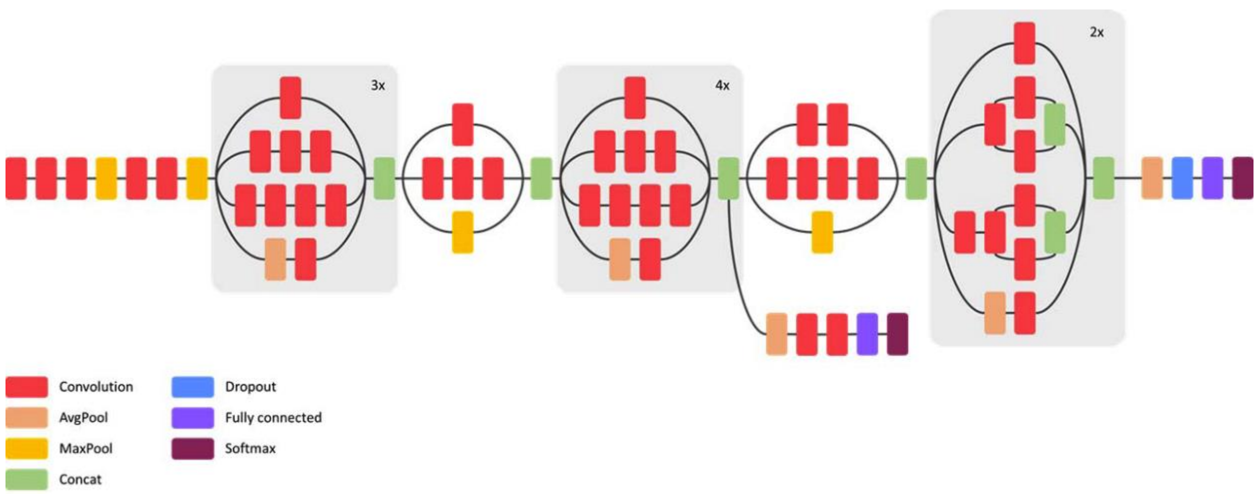


Figure 2: Structure of the InceptionV3 Architecture (Figure adopted from Chang et al. [13])

WeightedRGBLayer

The WeightedRGBLayer first splits the input image into red, green, and blue channels as three matrices. With randomly generated weight number assigned to each layer, three new layers are merged together as a new image to be sent into the InceptionV3 base model. (An example image output is given in **Appendix 1, Figure 2**)

Lens Distortion

A custom layer was used for simulating lens distortion by running images through it before being fed into the classification model. The lens distortion is achieved by constructing a camera matrix, consisting of 5 definable parameters/coefficients, k1, k2, k3 (which are manipulated below to create the two distortions), f (camera focal length, set to 45 for both distortions), cx and cy (defines the optical center, both are set to 64 to mark the center of the 128x128 image). The layer achieves distortion by using the distortion matrix to perform distortion correction, but since the images are undistorted to begin with, the resulting correction makes the images distorted opposite to the distortion matrix.

Two types of lens distortions were simulated (**Appendix 1, Figure 3**). The first distortion (Distortion 1) was simulated by using a lens distortion layer with k1=-0.01, k2=-0.05, k3=0.01. The second distortion, barrel distortion, was attempted to recreate a common lens distortion encountered by wide-angle lenses. To this effect, a lens distortion layer was created with k1=0.01, k2=0.05, k3=0.01.

Two sets of datasets were created by simulating all of the training and validation dataset with each distortion. A total of 4 models were trained with the distorted datasets, two for each dataset. One was with a model without the weighted RGB layer, and the other model with the weighted RGB layer. For all of the models, no gaussian blur was utilized since the lens distortion itself functioned as the point spread function of our simulated camera lens.

Evaluation Methods

All models in this project were trained with Adam optimizer with 0.0001 learning rate. The loss function is categorical cross-entropy, and batch size is 20. The accuracy of trained model over test dataset was evaluated after 35 epochs.

3. Results

After 35 epochs of training of our models, the learning curve appeared to reach a plateau in accuracy and loss curves.

Base Model Performance

Applying only one batch normalization layer before sending the image to the InceptionV3 module, the model has a 92% accuracy over our dataset (**Figure 3**). The precision, recall, and f1 score are 0.92, 0.92, and 0.92 respectively (**Appendix 2, Table 1**). In addition, the model's prediction over the vasc and df category is significantly higher than the other types of skin lesion.

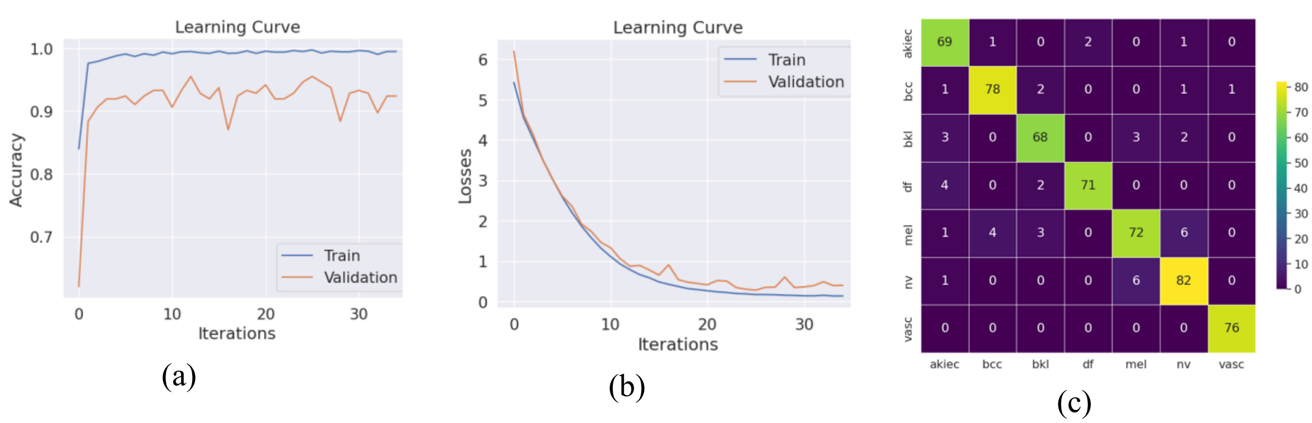


Figure 3: Learning curve of training with base model: (a) accuracy, (b) loss; and (c) confusion matrix of the model prediction on test set.

Gaussian Blur

By applying the Gaussian blur prior to the base model, the accuracy of the model is significantly reduced. In addition, by changing the standard deviation (σ), the blurring of image varies as well, and so does the accuracy of our model. In a general trend, the larger the σ, the greater the blur, and the poorer the model performance (**Appendix 1, Figures 5-8**; **Appendix 2, Table 2**).

WeightedRGBLayer

At different level of blurriness of the image, the performance of the model is improved for most cases (except for $\sigma = 0.5$) by applying the WeightedRGBLayer (**Appendix 2, Table 2**). In addition, by assigning different initial values to the model, the performance of the model was different as well (**Appendix 2, Table 3**). In addition, since the weights were assigned randomly, it was found that the final weights after training are different from one model fit to another. Moreover, the optimized weights for color channels after each training varied. For most cases, at least one channel has 0 weight if the initiator was set as “Uniform”, and green channel was less likely to be given the largest weight as observed from our trainings under different Gaussian blur conditions (**Appendix, Table 4**).

Lens Distortions

The metrics show that both the models trained on images simulated with Distortion 1 performed worse than the base model (**Appendix 2, Table 5**). Having said that, the weighted RGB model performed relatively better as anticipated. However, the models trained on images simulated with barrel distortion managed to outperform even the base model. In addition, the weighted RGB model was outperformed by the model not utilizing weighted RGB layer.

4. <u>Discussions</u>

The InceptionV3 architecture is indeed performing well as expected with a 92% high accuracy. In addition, it is found that the Gaussian blur influences the precision in predicting df, mel, nv, and vasc most significantly. It is also noticeable that the prediction appears to be at a high level for akiec, bcc, and bkl even with a very strong σ applied with the Gaussian blur layer. However, the recall and f1-score for the prediction outcome of these categories are very low, indicating that the model might be missing important features over the database. This is also indicating that the lens property is essential in diagnostic of skin lesion through images acquired.

The WeightedRGBLayer is proved to be improving the performance of our model with a low-σ Gaussian blur input. With a stronger blurry operation on the image, the model with WeightedRGBLayer performs similarly to the model without the layer. However, we only tried testing the model setting the initializer as “uniform” in our project at every σ-level. We examined the WeightedRGBLayer with different initializers with $\sigma = 0.5$ for Gaussian blur layer. Surprisingly, the performance of the model with WeightedRGBLayer by setting initializer as “uniform” performs the worst, compared to initializers including “RandomNormal”, “ones”, and “initializers.Constant(Value=0.5)”. For future steps, we will examine the model with several different initial values of the WeightedRGBLayer’s weight parameters.

The WeightedRGBLayer is proved to be improving the performance of our model with a low-σ Gaussian blur input. With a stronger blurry operation on the image, the model with WeightedRGBLayer performs similarly to the model without the layer. However, we only tried testing the model setting the initializer as “uniform” in our project at every σ-level. We examined the WeightedRGBLayer with different

initializers with $\sigma = 0.5$ for Gaussian blur layer. Surprisingly, the performance of the model with WeightedRGBLayer by setting initializer as "uniform" performs the worst, compared to initializers including "RandomNormal", "ones", and "initializers.Constant(Value=0.5)". For future steps, we will examine the model with several different initial values of the WeightedRGBLayer's weight parameters.

In addition, since all trainings were done with only 2,800 images, we are expecting to train this model with a larger, and more representative dataset. We've also found that duplicated images are also provided in HAM10000 dataset, which may lead to problems including test set not being not completely independent from the training set, overfitting, and the model being not generalized for new data. Jain et al. has reorganized the dataset with 5514 unique images from HAM10000 [1]. However, the new dataset is still strongly imbalanced. Thus, proper oversampling or down sampling strategies should be applied to the dataset.

The results of the models trained with simulations of distortion 1 were in accordance with expectations. Since the images were visibly degraded after being distorted, all the metrics for both the models were less impressive than the baseline model. But, the RGB model having the better performance further validated the effectiveness of an illumination layer to enhance classification.

However, the base model trained on the barrel distorted dataset managed to slightly outperform even the base model. This maybe because, as a result of the barrel distortion, the center of images appeared to be much more enhanced even if the overall quality image does undergo some degree of degradation in quality, and since all of the images have the lesion at the center, we suspect it was easier for the model to learn the underlying relationship of each lesion and hence perform a better classification. Since the RGB model did not outperform its counterpart this time, we suspect this may be due to the fact that the illumination also managed to enhance the degradation presented in the distorted images.

5. Conclusions

In this project, we used InceptionV3 module as our base model architecture to train a sampled and augmented dataset from HAM10000 dataset, which performs well in predicting skin lesion types from image inputs. With Gaussian blur and lens distortion by distortion 1, image quality is reduced, resulting in a drop in precision of prediction as well as the recall and f1-score of the model prediction. Applying WeightedRGBLayer before sending the image into InceptionV3 architecture allows the model to recover the performance. However, the parameters including the initializer is not yet optimized. Moreover, given the small scale of the training and test dataset, a larger, more generalized, and more representative dataset needs to be developed. In addition to this, the impressive performance of barrel distorted images gives another potential mode of enhancement of images. However, this may not be as effective as weighted RGB as barrel distortion only enhances the center of the image and thus might need to be used in specific scenarios where the lesions are present in the center.

Reference

[1] S. Jain, U. Singhania, B. Tripathy, E. A. Nasr, M. K. Aboudaif, and A. K. Kamrani, "Deep Learning-Based Transfer Learning for Classification of Skin Cancer," Sensors, vol. 21, no. 23, p. 8142, Dec. 2021, doi: 10.3390/s21238142.

[2] R. Marks, "The Epidemiology of Non-Melanoma Skin Cancer: Who, Why and What Can We Do about It," J. Dermatol., vol. 22, no. 11, pp. 853–857, Nov. 1995, doi: 10.1111/j.1346-8138.1995.tb03935.x.

[3] A. Farooq, X. Jia, J. Hu, and J. Zhou, "Transferable Convolutional Neural Network for Weed Mapping With Multisensor Imagery," IEEE Trans. Geosci. Remote Sens., vol. 60, 2022, doi: 10.1109/TGRS.2021.3102243.

[4] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-December, pp. 2818–2826, 2016, doi: 10.1109/CVPR.2016.308.

[5] P. Schwedhelm, D. Baldauf, and S. Treue, "The lateral prefrontal cortex of primates encodes stimulus colors and their behavioral relevance during a match-to-sample task," Sci. Rep., vol. 10, no. 1, pp. 1–12, 2020, doi: 10.1038/s41598-020-61171-3.

[6] J. Rong, S. Huang, Z. Shang, and X. Ying, "Radial Lens Distortion Correction Using Convolutional Neural Networks Trained with Synthesized Images," in Computer Vision -- ACCV 2016, vol. 1, no. 2, 2017, pp. 35–49.

[7] N. Hameed, A. M. Shabut, M. K. Ghosh, and M. A. Hossain, "Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques," Expert Syst. Appl., vol. 141, p. 112961, 2020, doi: 10.1016/j.eswa.2019.112961.

[8] P. M. M. Pereira et al., "Skin lesion classification enhancement using border-line features – The melanoma vs nevus problem," Biomed. Signal Process. Control, vol. 57, p. 101765, 2020, doi: 10.1016/j.bspc.2019.101765.

[9] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–14, Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.1556.

[10] C. Szegedy et al., "Going deeper with convolutions," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 07-12-June-2015, pp. 1–9, 2015, doi: 10.1109/CVPR.2015.7298594.

[11] Z. Deng, Y. Cao, X. Zhou, Y. Yi, Y. Jiang, and I. You, "Toward efficient image recognition in sensor-based IoT: A weight initialization optimizing method for CNN based on RGB influence proportion," Sensors (Switzerland), vol. 20, no. 10, 2020, doi: 10.3390/s20102866.

[12] P. Tschandl, C. Rosendahl, and H. Kittler, "Data descriptor: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," Sci. Data, vol. 5, pp. 1–9, 2018, doi: 10.1038/sdata.2018.161.

[13] W. J. Chang, L. B. Chen, C. H. Hsu, C. P. Lin, and T. C. Yang, "A Deep Learning-Based Intelligent Medicine Recognition System for Chronic Patients," IEEE Access, vol. 7, no. April, pp. 44441–44458, 2019, doi: 10.1109/ACCESS.2019.2908843.

Appendix 1: Figures

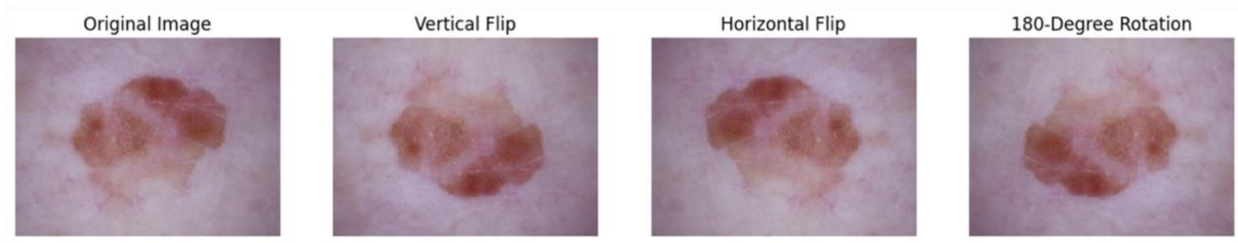
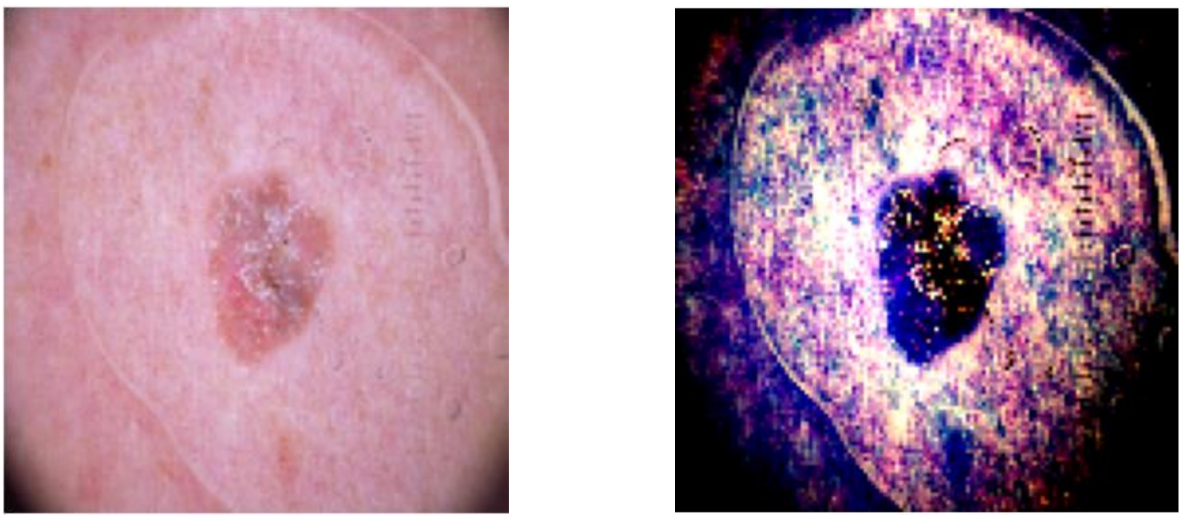
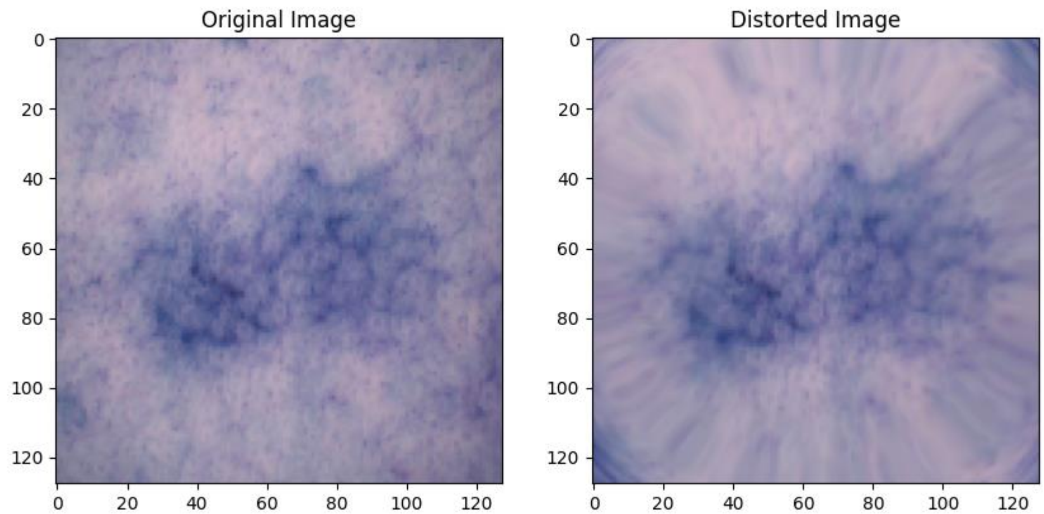


Figure 1: An example of image augmentation

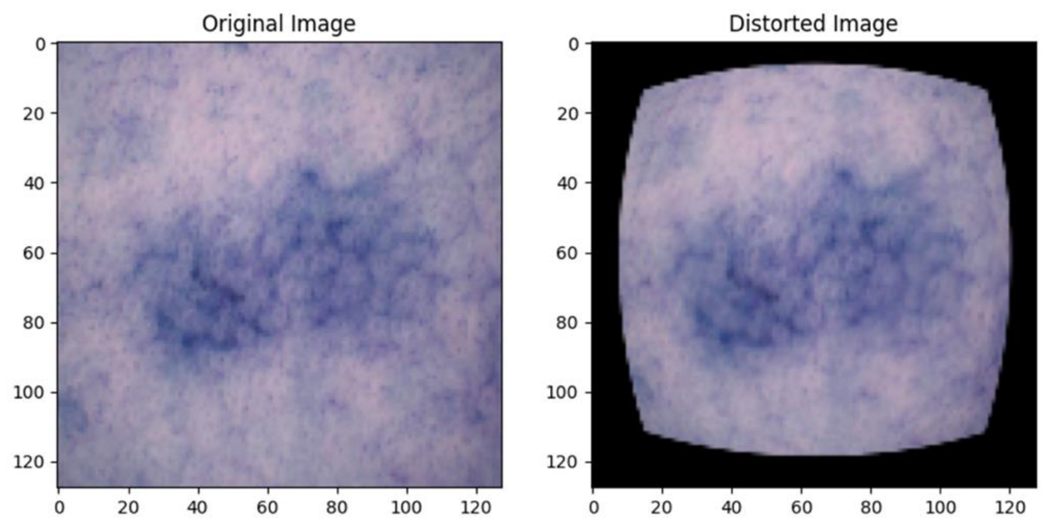


(a) (b)

Figure 2: (a) An example of original input image; (b) An example of image being reweighted by WeightedRGBLayer.

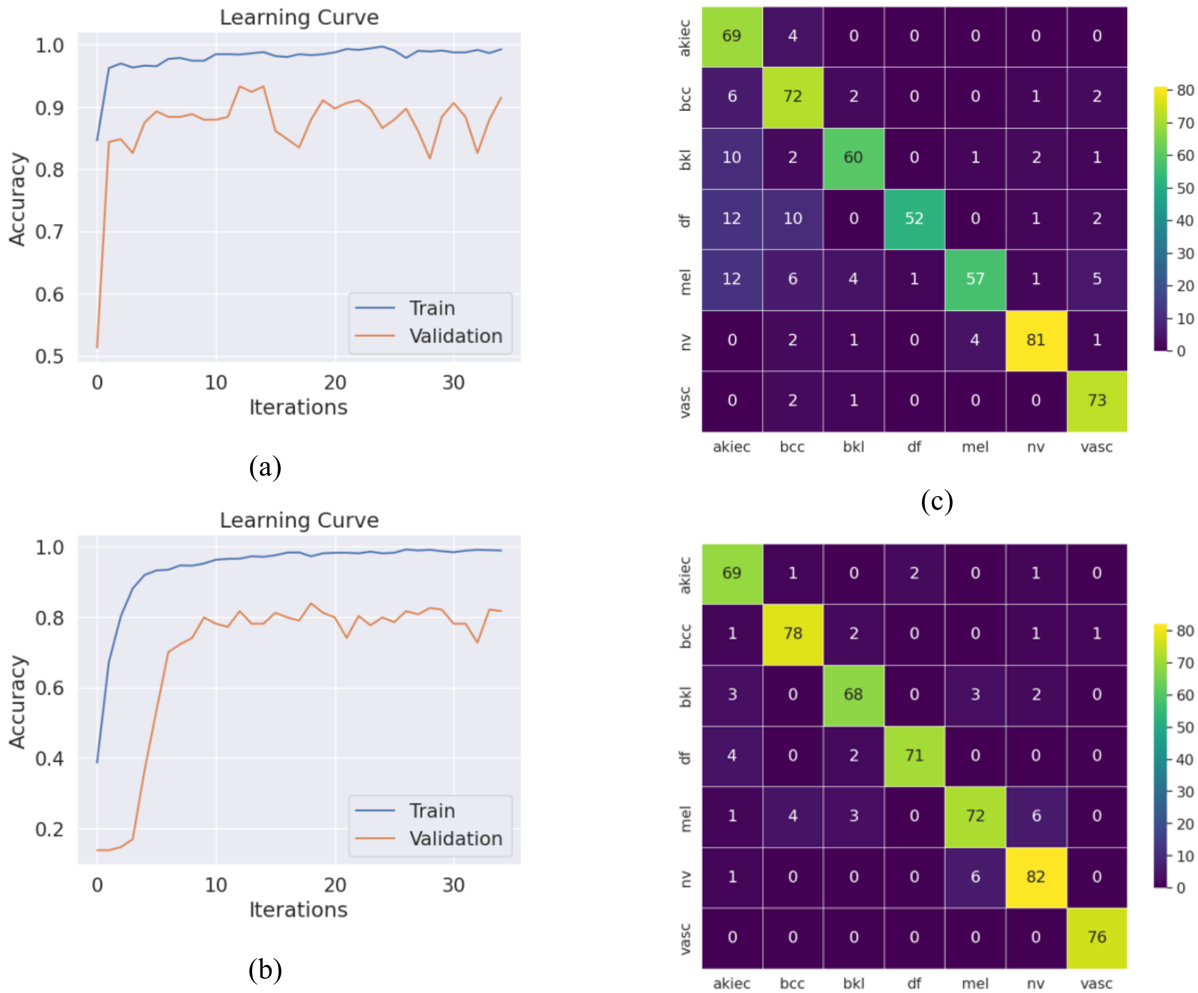


(a)



(b)

Figure 3: Sample training image undergoing lens distortions: (a) Distortion 1, (b) Barrel distortion



(a) (b) (c) (d)

<u>Figure 4: Learning curve of training with $\sigma = 0.1$: (a) without WeightedRGBLayer, (b) with WeightedRGBLayer; and confusion matrix of the model prediction on test set: (c) without WeightedRGBLayer, (d) with WeightedRGBLayer</u>

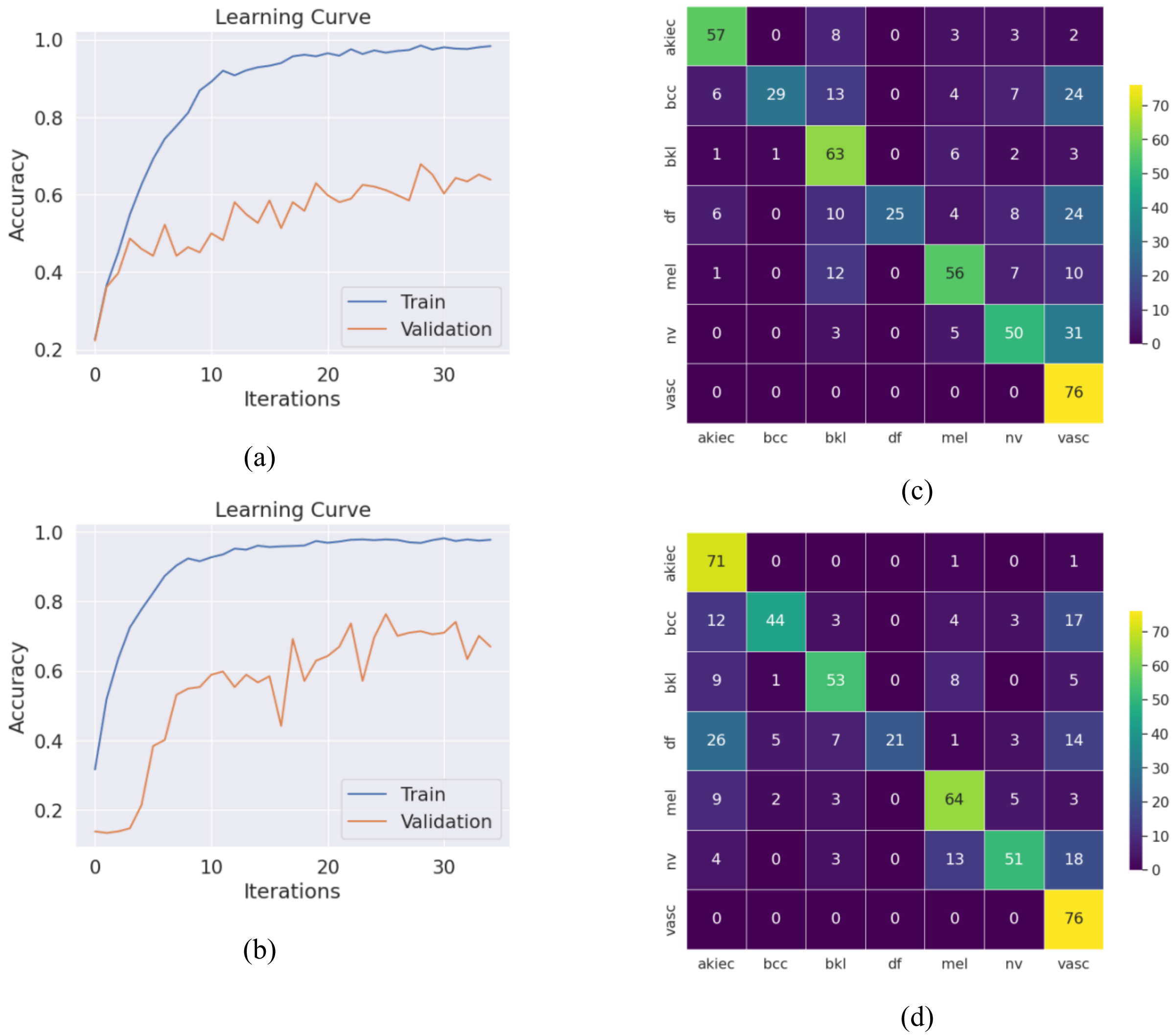


Figure 5: Learning curve of training with $\sigma = 0.3$: (a) without WeightedRGBLayer, (b) with WeightedRGBLayer; and confusion matrix of the model prediction on test set: (c) without WeightedRGBLayer, (d) with WeightedRGBLayer

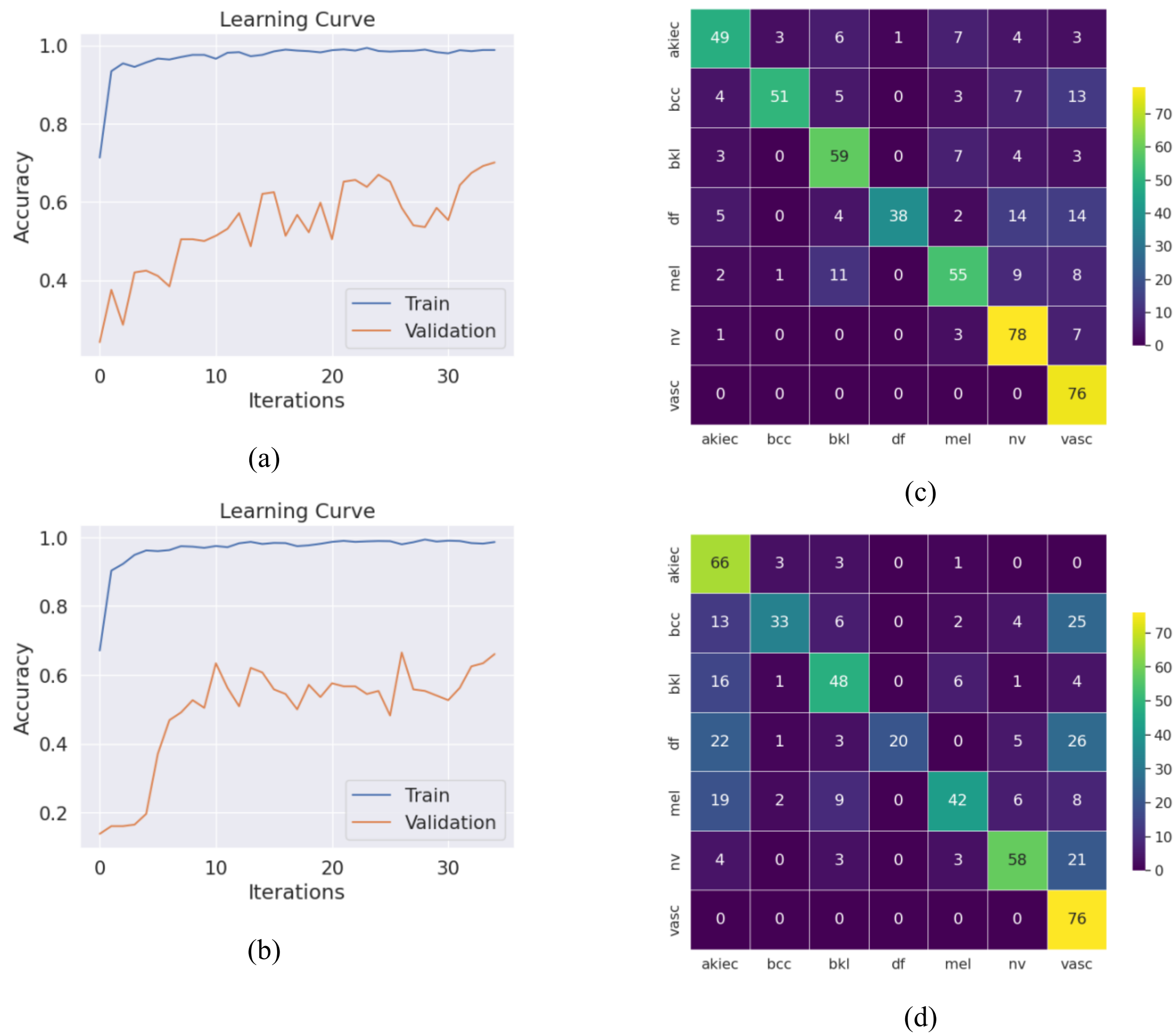
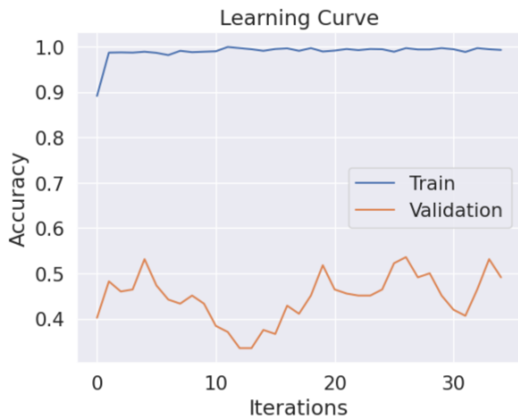
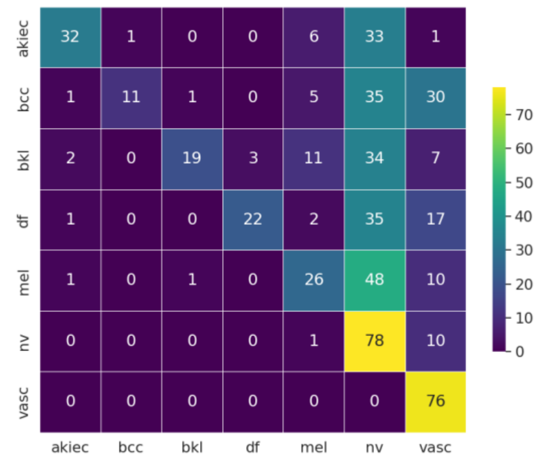


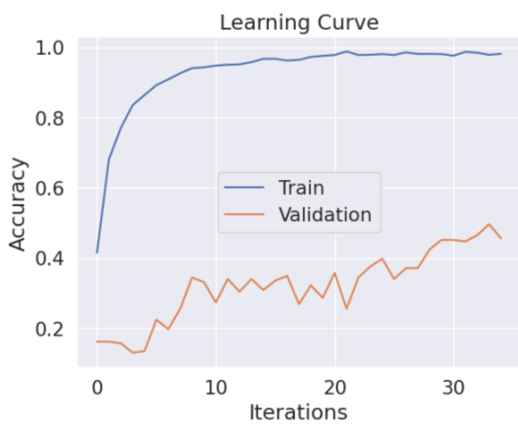
Figure 6: Learning curve of training with $\sigma = 0.5$: (a) without WeightedRGBLayer, (b) with WeightedRGBLayer; and confusion matrix of the model prediction on test set: (c) without WeightedRGBLayer, (d) with WeightedRGBLayer



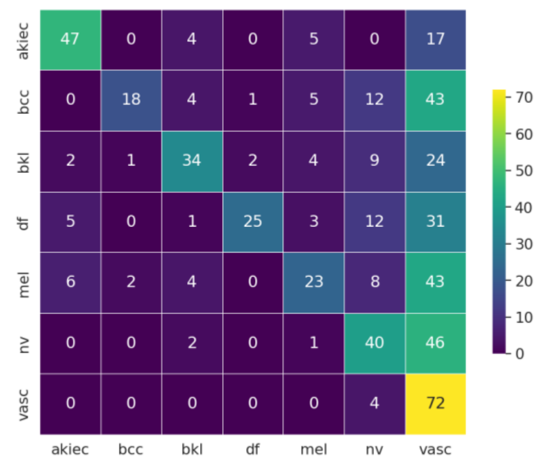
(a)



(c)

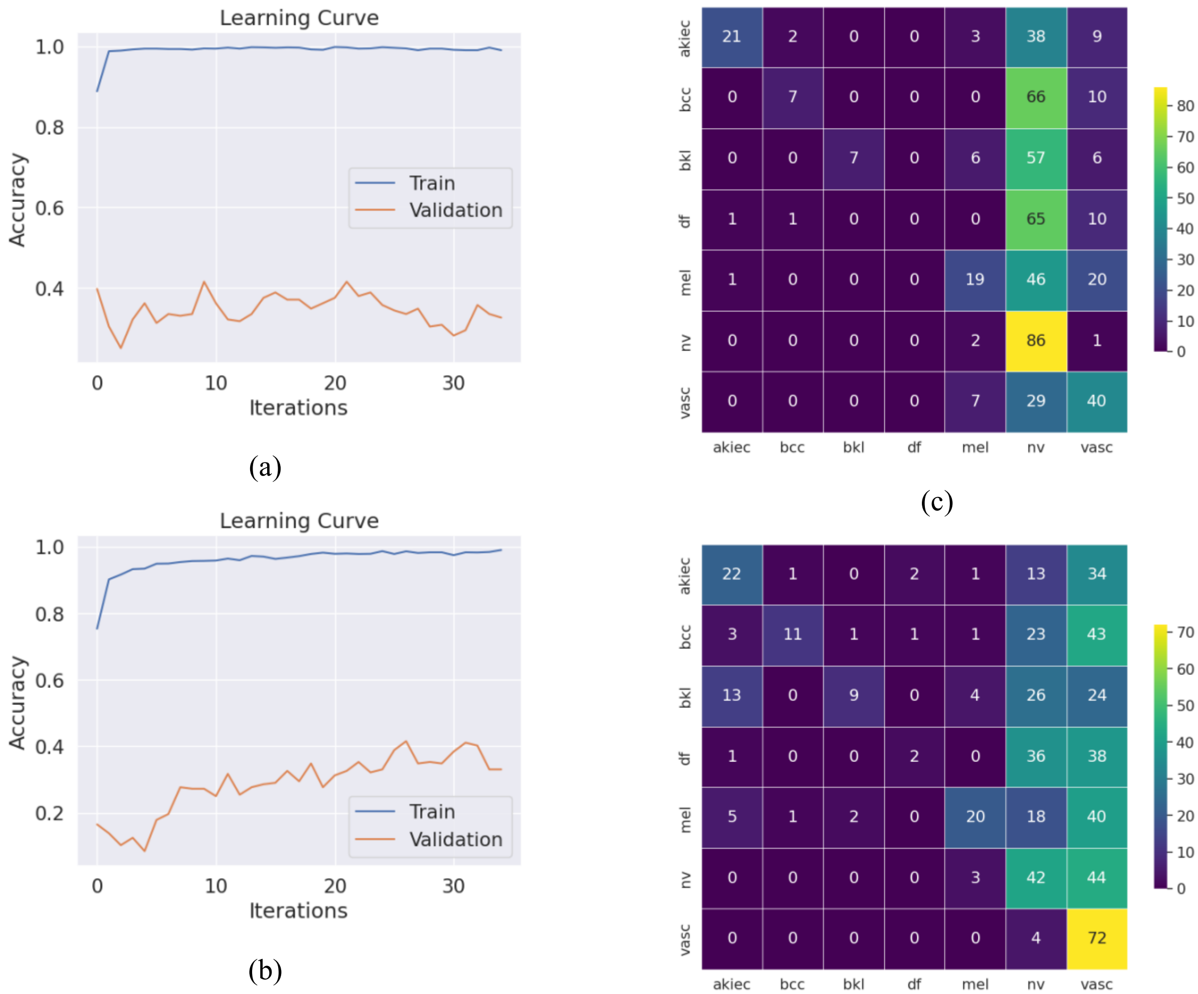


(b)



(d)

Figure 7: Learning curve of training with $\sigma = 0.7$: (a) without WeightedRGBLayer, (b) with WeightedRGBLayer; and confusion matrix of the model prediction on test set: (c) without WeightedRGBLayer, (d) with WeightedRGBLayer



(a)

(b)

(c)

(d)

<u>Figure 8: Learning curve of training with $\sigma = 1.0$: (a) without WeightedRGBLayer, (b) with WeightedRGBLayer; and confusion matrix of the model prediction on test set: (c) without WeightedRGBLayer, (d) with WeightedRGBLayer</u>

Appendix 2: Tables

Table 1: Test Report of the model

	PRECISION	RECALL	F1-SCORE	SUPPORT
AKIEC	0.87	0.95	0.91	73
BCC	0.94	0.94	0.94	83
BKL	0.91	0.89	0.90	76
DF	0.97	0.92	0.95	77
MEL	0.89	0.84	0.86	86
NV	0.89	0.92	0.91	89
VASC	0.99	1.00	0.99	76

Table 2: Test report of model with different level of Gaussian Blur

σ	WeightedRGBLayer*	Accuracy	Precision	Recall	F1-score
0.1	False	0.83	0.85	0.83	0.83
	True	0.93	0.93	0.93	0.93
0.3	False	0.64	0.74	0.64	0.62
	True	0.68	0.75	0.68	0.66
0.5	False	0.73	0.77	0.72	0.72
	True	0.61	0.72	0.61	0.60
0.7	False	0.47	0.69	0.47	0.44
	True	0.46	0.64	0.46	0.47
1.0	False	0.32	0.53	0.32	0.27
	True	0.32	0.53	0.32	0.28

* Initializer for WeightedRGBLayer is “Uniform”

Table 3: Final RGB weight outputs* with different levels of Gaussian blur

σ	R_WEIGHT	G_WEIGHT	B_WEIGHT
0.1	0.0358	0	0.0276
0.3	0.0096	0.0049	0
0.5	0.0120	0.0134	0
0.7	0	0.0079	0.0083
1.0	0	0.0206	0.0158

* Initializer for WeightedRGBLayer is "Uniform"

Table 4: Test report of model with different initializer in WeightedRGBLayer; $\sigma = 0.3$

INITIALIZER	R_WEIGHT	G_WEIGHT	B_WEIGHT	ACCURACY
UNIFORM	0.0096	0.0049	0	0.66
RANDOMNORMAL	0.0112	0	0.0298	0.79
ONES	1.0000	0.9903	1.0132	0.78
CONSTANT = 0.5	0.4965	0.4997	0.5055	0.81

Table 5: Test Metrics of lens distortion models

	WEIGHTEDRGBLAYER*	ACCURACY	PRECISION	RECALL	F1-SCORE
DISTORTION 1	False	0.69	0.74	0.68	00.71
	True	0.74	0.78	0.72	0.75
BARREL DISTORTION	False	0.89	0.91	0.88	0.89
	True	0.86	0.87	0.85	0.86

* Initializer for WeightedRGBLayer is "Uniform"