Technology and Health Care 30 (2022) S143–S153
DOI 10.3233/THC-228014
IOS Press

Acne detection and severity evaluation with interpretable convolutional neural network models

Hao Wen[a,b], Wenjian Yu[a,*], Yuanqing Wu[b,*], Jun Zhao[b], Xiaolong Liu[b], Zhexiang Kuang[b] and Rong Fan[b]
[a]*Department of Computer Science and Technology, BNRist, Tsinghua University, Beijing, China*
[b]*Beijing Jingdong Shangke Information Technology Co., Ltd, Beijing, China*

Abstract.
BACKGROUND: Acne vulgaris is one of the most prevalent skin conditions, which harms not only the patients' physiological conditions, but also their mental health. Early diagnosis and accurate continuous self-monitoring could help control and alleviate their discomfort.
OBJECTIVE: We focus on the development and comparison of deep learning models for locating acne lesions on facial images, thus making estimations on the acne severity on faces via medical criterion.
METHODS: Different from most existing literature on facial acne analysis, the considered models in this study are object detection models with convolutional neural network (CNN) as backbone and has better interpretability. Thus, they produce more credible results of acne detection and facial acne severity evaluation.
RESULTS: Experiments with real data validate the effectiveness of these models. The highest mean average precision (mAP) is 0.536 on an open source dataset. Corresponding error of acne lesion counting can be as low as 0.43 ± 6.65 on this dataset.
CONCLUSIONS: The presented models have been released to public via deployed as a freely accessible WeChat applet service, which provides continuous out-of-hospital self-monitoring to patients. This also aids the dermatologists to track the progress of this disease and to assess the effectiveness of treatment.

Keywords: Facial acne, convolutional neural network, object detection, interpretability, self-monitoring

1. Introduction

Medical image processing has long been an active research field. The resources for the research of medical image processing are rich, including MRI, X-ray, CT, dermoscopy images, etc. Among these resources, images taken by smartphones, or other portable cameras, are the most accessible and the most convenient for the purpose of out-of-hospital monitoring [1]. Accessibility and the ability to conduct out-of-hospital monitoring are often crucial to whether or not the research results could be widely accepted and put into practical use, especially as APPs for smartphones. More and more medical research work, for example [2–4], are being carried out with images taken by portable devices as their

*Corresponding authors: Wenjian Yu, Department of Computer Science and Technology, BNRist, Tsinghua University, Beijing 100084, China. E-mail: yu-wj@tsinghua.edu.cn. Yuanqing Wu, Beijing Jingdong Shangke Information Technology Co., Ltd, Beijing 101111, China. E-mail: wuyuanqing@jd.com.

main resource of study. The company SkinVision published a series of papers [5–8] concerning the usage of smartphones for analyzing skin conditions. Research work, harnessing the powerfulness of the convolutional neural networks (CNNs), on skin diseases inspired us most and also directly motivated our studies. In [9], researchers made a comparative study on performances of 5 CNN models, namely ResNet-50, Inception-v3, DenseNet121, Xception and Inception-ResNet-v2, on the classification of 6 common facial skin diseases. They pretrained these models using other body part images and performed transfer learning using facial images. The best performance was achieved by Inception-ResNet-v2 with average precision 63.7% and average recall 67.2%. More importantly, a new dataset XiangyaDerm was proposed in this paper. This dataset, as was later described in more details in [10], is much larger in its number of images than other existing open-access datasets. Most recently, a group from Google Health published a study [11] about deep learning systems for the detection of a broader range of skin diseases. Their system achieved 71% top-1 and 93% top-3 accuracies respectively, outperforming the dermatologists.

Acne vulgaris, usually referred to as acne, is one of the most prevalent skin conditions. Although typically seen among adolescents, it can persist into adulthood. Acne harms not only the patients' physiological conditions, but also their mental health, by laying detrimental effects on self-esteem [12]. Early diagnosis and accurate continuous self-monitoring could help control and alleviate discomfort brought by acne. Self-monitoring here includes a broad range of concepts, including self-diagnosis, self-evaluation of the effect of treatment or therapy provided by the dermatologists, etc., hence helps both the patients and the dermatologists. Hayashi et al. [13] gave a comprehensive study on statistical relations between acne severity and the number of acne lesions. Particularly, they gave easy-to-use rules for estimating acne severity from lesion counting (of papules and pustules, which are two subtypes of acne) per half face. However, interpretability, which makes predictions more convincing, is an important issue for the existing automated diagnosis systems. It is actually a common weakness of those end-to-end deep learning models. "end-to-end" here refers to producing solely the number of acne lesions or the level of acne severity. Therefore in this paper, we focus on the application of different object detection CNN models on the problem of locating acne lesions in facial images. Then, acne severity could automatically be produced via numerical grading schemes, thus serving our initial purpose of self-monitoring.

Our contributions in this paper could be summarized as follows. First of all, we explore the potentials of state-of-the-art (SOTA) object detection models on the problem of facial acne analysis. The best object detection CNN model achieves the mean absolute error (MAE) of 3.49 on the filtered ACNE04 dataset proposed in [14]. For the precise meaning of "filtered", refer to the Section 3.1. The object detection models own the advantage of high interpretability that those "end-to-end" methods lack. Second, we have developed a framework of continuous self-monitoring of facial acne severity using the presented object detection models, which is already in use. Lastly, the trained models are released to public domain via GitHub.[1] The structure of the rest of the paper is sketched as follows. In Section 2, historical methods on the problem of facial skin conditions analysis are briefly reviewed. Sections 3 and 4 are devoted to the demonstration of our novel methods and framework and their performances on facial acne analysis and monitoring. Discussions on the application and limitations of our new models follow in Section 5, where several future work directions are also presented.

2. Related work

There is substantial interest among researchers dealing with medical image processing in the design

[1] https://github.com/wenh06/acne_detection/ and https://github.com/wenh06/yolov4_acne_torch.

of automated acne analysis system. The first attempt specialized in this area can be dated back to [15], although its resources of study were multi-spectral images. Methods previously explored in this research field are mainly in two categories: traditional computer vision (CV) segmentation methods [2,3,16,17], and deep learning methods (in particular CNN models) [4,14].

2.1. Traditional CV methods

Before the breakthrough and the thus following wide adoption of CNN methods, traditional CV methods had long been the mainstream. Roughly speaking, traditional CV methods generally first extract candidates of regions of interest that differ typically in color or texture with normal part of skin. Such regions may include acne, moles, stains, etc. Afterwards, selections are performed among region candidates via manually extracted features.

Wong et al. [2] developed an iterative stochastic region-merging to segment skin lesions, which outperformed the SOTA methods, SLS [18], MGVF [19] and LSAC [20], by then. SVM is used in [16] for the extraction of acne away from other skin pigmentation. Until recently, there are still research work following such paradigms. [17] used entropy-based filtering and thresholding to determine the region-of-interest and used discrete wavelet frames to extract acne lesion features for classifying different types of acne. The possibility of hard-coding the dermatologist criteria for the establishment of a diagnosis system was explored [3].

2.2. Deep learning regression

In [4], researchers from Microsoft implemented a regression model via performing transfer learning on a pretrained ResNet152 model. Their model was capable of producing acne severity assessment as score ranging from 0 to 5 from a selfie image. In order to achieve higher robustness, Zhao et al. [4] took advantage of some trick of image augmentation, named skin patch rolling. Another group of researchers from Nankai University developed a hybrid model in [14], which was capable of producing predictions of both (the distribution of) the number of acne lesions and the corresponding severity at the same time. The main contribution of [14] was the combination of the task of Label Distribution Learning (LDL) with the normal classification task. This multitask structure improved performances both on the acne severity classification problem and on the acne counting problem, as claimed by Wu et al. [14]. The output of LDL could provide higher credibility than a single end-to-end classification output as well.

2.3. Saliency: An alternative

Saliency analysis is another possibility, although not specially designed for locating lesion regions when using end-to-end CNN classifiers as those tested in [9,11]. Researchers of [11] performed saliency analysis for the model they developed via integrated gradients, in which case lesion regions, or rather the spatial regions of feature maps that their model paid the most attention to when making specific predictions, were highlighted. Up to our knowledge, Grad-CAM [21], and its variants like Grad-CAM++ [22], are widely accepted as the SOTA methods for performing saliency analysis for CNN models. Implementations of Grad-CAM and Grad-CAM++ are straightforward, especially for those sequential models. The drawbacks of such "attention" analysis are apparent as well. When making wrong predictions, the models may focus on regions that are not where the lesions lie. Moreover, the "target" layers to draw such "attention" maps are hard to determine. All these reasons make it an unstable method of saliency analysis for the task of skin lesion localization.

Indeed, saliency analysis was the very first trial we made after developing several skin disease classification models (5-classes, 17-classes, 22-classes) in an attempt to reproduce and expand results in [9,11]. We selected images from Dermnet (http://www.dermnet.com/) and the dataset SD260 proposed in [23] and performed transfer learning on a Xception model pretrained on ImageNet [24] provided in keras.[2] When applying Grad-CAM or Grad-CAM++ for saliency analysis for our models, we made experiment using "add_12" and "conv2d_4" as target layers (corr. to feature maps the last block with skip connections of Xception), but their performances were unstable. As this problem is not the main issue we study in this paper, we shall not go any further on this topic.

3. Materials and methods

Object detection (also known as object localization) is the most straightforward and perhaps the most convincing method for acne analysis on facial images. Once the acne lesions are located in the face, numerical scoring scheme could be applied to make assessments of acne severity. Object detection has achieved great success in many application areas, for instance in autonomous driving, but its potential in facial acne analysis seems to have not been fully exploited. Actually, Wu et al. [14] did conduct experiments on this topic as a comparison, but did not provide much detailed information. This might owe to the fact that their "joint learning" method rather than object detecting was their choice of method for facial acne analysis.

The drawback of object detection methods lies in their time efficiency, especially those with heavy backbone CNN, e.g. those which used ResNet-101 as backbone as listed in Table 9 of [25]. This could be a serious issue for autonomous driving, but would be of less sensitivity in our scenario.

As it is rare for open-access skin disease image datasets to have bounding box annotations locating the lesions, which usually only have disease names as annotations, it is particularly remarkable that the authors of the dataset ACNE04 provided such annotations for each images of their dataset. This greatly facilitated our study, and we adopted this dataset as the base material for developing and validating our models.

3.1. Dataset and data cleaning

As is described in the article by Wu et al. [14], the dataset ACNE04 contains 1457 images (indeed 1513 in total in the zip file provided by its authors), along with annotations of bounding boxes and the number of lesions and severity for each image. The dataset can be found in the GitHub repository https://github.com/xpwu95/LDL. However, despite the images collected following the requirements (mainly concerning "half face") of the Hayashi grading criterion [13], the dataset ACNE04 was contaminated by a number of images with very low quality. A small part of the images were taken when the camera was placed too far away from the volunteers. Facial skin makes up too small areas in these images, which makes them not suitable for developing our model. Therefore, we excluded all these images mentioned above, leaving the 1222 high quality images for our training and validation. In order for more precise localizations of the lesions, we also shrank the original bounding boxes provided by Wu et al. [14], so that each bounding box enclosed the lesion almost precisely. By doing this, the negative impact of normal skin (texture, color, etc.) to the detectors were reduced to the least when performing training. In the rest of this paper, when referring to ACNE04, we would mean this cleaned dataset.

[2] F. Chollet et al., "Keras", https://keras.io, 2015.

3.2. Selection of models of object detection

We performed tests using a range of the existing SOTA object detection methods, mainly general purposed ones rather than those specially tailored for small object detection models, in order to pick out one or more usable models. Two families of detectors were selected as candidates: the one-stage single shot family in consideration of time efficiency and the two-stage Region-based CNN (R-CNN) family in consideration of the performance of the task of acne detection itself.

Single Shot Multibox Detector (SSD) [26] with MobileNet v1 as backbone (the backbone was VGG-16 in the original paper [26]) was the first detector we tried. As its name implies, SSD uses a single neural network which produces a collection of bounding boxes at multiple scales and corresponding scores for the presence of certain objects. Yolov4 [25], the SOTA one-stage detector, was also tested. Yolov4 gathers many of the recently developed techniques, including newly proposed "mosaic" image augmentation method, hence is expected to be more powerful. We also tried faster R-CNN [27] with ResNet as backbone and with Inception-v2 as backbone. Faster R-CNN is an evolved member of the R-CNN family. Detectors that belong to the R-CNN family have an extra stage of region proposal through Selective Search or through Region Proposal Networks (RPNs), hence are more time consuming, especially those with heavy backbones like Inception-ResNet-v2. However, in general they have better performances on the object detection task itself.

It should be emphasized that since acne lesions are small objects, anchors should better be adjusted, for example using kmeans proposed in [28].

The main evaluation metrics for the models tested are the commonly used metrics average precision (AP, the usually used is mean average precision (mAP), but we left out the "mean" as in this study we had only one class for the acne lesions. The same for AR) and Average Recall (AR), first formalized in the PASCAL Visual Objects Classes (VOC) challenge [29] measured at Intersection over Union (IoU) threshold of 0.5 and later extended by the MSCOCO Challenge [30].

4. Experiments

In this section, we will give more details about the experiments we have carried out and make some comparisons with the original work [14] that first proposed the dataset ACNE04. Particularly, we designed and conducted two sets of experiments. In one set of experiment, we used only a small part about 400 images, most of which are of low acne severity, to comprise the training and validation sets for model developing. We aim to check to what extent the acne detectors trained on insufficient data can be generalized. Detectors trained with this dataset will have a suffix of "small" in the rest of the paper. In another experiment, we used almost the whole dataset (excluding the several dozen images of acne severity level 3) to train our acne detector. Detectors trained with this dataset will have a suffix of "large" in the rest of the paper.

4.1. Training details

Despite the Yolov4 detector which is re-implemented in PyTorch, we base the project on top of the object detection API [31] provided by Tensorflow [32]. This module provides many implementations of the SOTA object detection methods, for instance SSD, Faster R-CNN, with various backbone CNNs, along with pretrained models. On top of these pretrained models, we perform transfer learning using our

Table 1
Results of average precision for the 6 models

Model[3]	AP-50[4]
faster_rcnn_resnet101_large	0.536
faster_rcnn_inception_v2_large	0.434
faster_rcnn_resnet101_small	0.472
faster_rcnn_inception_v2_small	0.422
ssd_mobilenet_v1_large	0.166
Yolov4_large	0.526

Table 2
Statistical results of 3 models

Model	Score threshold	MAE	RMSE	Mean	STD
faster_rcnn_resnet101	**0.25**	**3.49**	**6.79**	**−0.16**	**6.79**
	0.50	3.50	6.92	−0.37	6.91
	0.75	3.55	7.05	−0.58	7.03
faster_rcnn_inception_v2	0.25	3.57	6.69	0.04	6.70
	0.50	3.63	7.30	−1.71	7.10
	0.75	4.40	8.79	−3.59	8.02
Yolov4	0.25	3.61	6.66	0.43	6.65
	0.50	3.65	6.88	−0.02	6.88
	0.75	3.77	7.19	−0.45	7.17

dataset. For the second set of experiments, we perform transfer learning on top of the models produced in the first experiment if applicable.

Since in this study we do not distinguish different subtypes of acne, there is only one class for the bounding boxes. In each experiment, datasets are randomly split into training and validation set by a ratio of 7:3.

5. Results

Average precision of SSD is almost constantly zero during training. Hence we treat it a failure of task and exclude this model from comparison. The 6 detectors thus obtained is compared in Table 1.

The average precision of faster_rcnn_resnet101_large is the highest, which is illustrated in more details in Fig. 1 and in Table 2 (3 models are selected for comparison). Acne lesion numbers, the essential ingredient for estimating acne severity in some numerical criterion as in [13], are obtained by counting the connected components of the acne masks obtained from the predicted bounding boxes in the images, or just counting the number of boxes after non-maximum suppression. A finer treatment would be further restricting the bounding boxes to the facial mask or the human skin mask which are obtained via other models. Figure 2 provides some examples for which detected acne lesion number matches exactly with the ground truth.

In Table 2, numerical comparisons, via MAE, root mean square error (RMSE), etc., are made using 3 different bounding boxes score thresholds, 0.25, 0.5, and 0.75. One obtains a bias array by subtracting the predicted acne lesion numbers by the ground truth acne lesion numbers. “Mean” and “STD” are the mean and standard deviation of this bias array resp. “MAE” is the mean of the absolute bias array. “RMSE” is

[3]For the suffixes “mall” and “large” of the models, see the first paragraph of Section 4.
[4]Average Precision @ [IoU = 0.50 | area = all | maxDets = 100].

Fig. 1. Examples for comparison of performances of the detectors. Acne lesions are located with green boxes. Names of the detectors are under the images, together with the number of acne lesion detected. The ground truth acne lesion number is 17. The bounding boxes and acne lesion number were drawn with score threshold = 0.5.

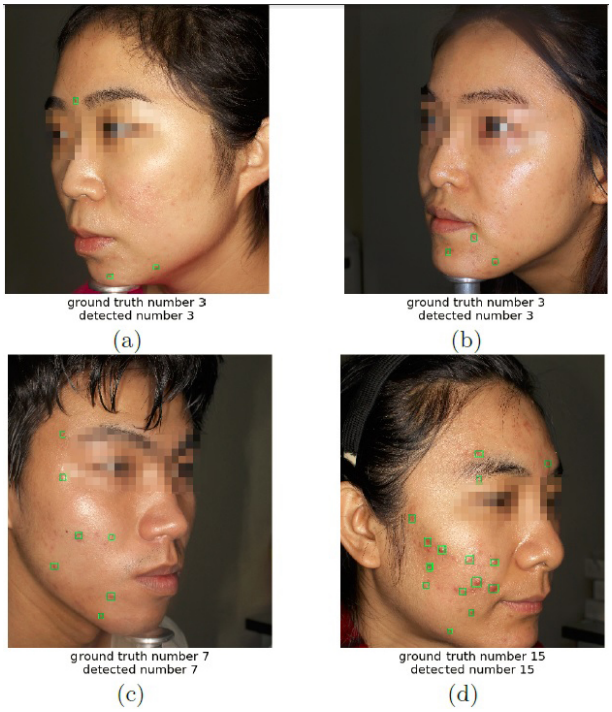

Fig. 2. Examples for which detected acne lesion number matches exactly with the ground truth. Acne lesions are located with green boxes. Detections were using faster_rcnn_resnet101_large with threshold = 0.5.

the square root the the mean of the squared bias array. Among all, faster_rcnn_resnet101_large achieves the lowest MAE of 3.49. This result is better than other object detection models tested in [14]. Although it is not as good as that obtained by the LDL model in [14], our model has better interpretability than the the LDL model.

Considering the problem of regression of acne numbers, a scatter plot of predicted acne numbers against ground truth numbers is given in Fig. 3a. Confusion matrix thus obtained, in which case we consider the problem of classification of acne severity, is shown in Fig. 3b. It can be inferred that predictions for low severity level images are considerably accurate. More detailed numerical evidence is given in Table 3,

Table 3
Comparison by groups of different severity

faster_rcnn_resnet101_large					
Severity	Score threshold	MAE	RMSE	Mean	STD
Level 0 mild 482 images	0.25	1.26	1.93	0.69	1.80
	0.50	1.22	1.88	0.63	1.77
	0.75	1.19	1.83	0.58	1.74
Level 1 moderate 578 images	0.25	3.37	4.50	1.31	4.30
	0.50	3.32	4.41	1.16	4.26
	0.75	3.30	4.36	1.02	4.23
Level 2 severe 132 images	0.25	5.76	7.21	−2.70	6.68
	0.50	5.86	7.30	−3.42	6.45
	0.75	6.32	7.65	−4.30	6.32
Level 3 very severe 30 images	0.25	31.53	34.64	−31.00	15.45
	0.50	32.90	35.84	−32.50	15.10
	0.75	33.93	36.71	−33.67	14.62

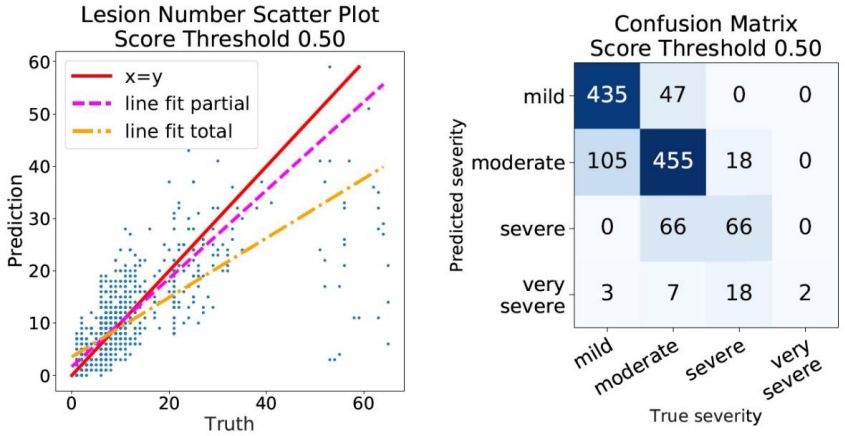

Fig. 3. The scatter plot of acne lesion number predictions (a) and the confusion matrix of acne severity prediction (b), for the model faster_rcnn_resnet101_large with score threshold = 0.5. In (a), the dashed magenta line "line fit partial" is obtained using images excluding the 30 images whose annotated acne severity are "very severe", while the dash-dot orange line "line fit total" is obtained using all images in the dataset.

where comparison via similar criteria as in Table 2 is made on groups of different acne severity, using the model faster_rcnn_resnet101_large with score threshold = 0.5. The only flaw lies in the predictions for the very severe group.

6. Discussion

6.1. Acne self-monitoring framework

The trained acne detectors can be, and currently have already been, deployed using tensorflow serving as a freely accessible service via WeChat[5] applet. This applet, with acne detector as one of the many

[5] One of the largest social network platforms in China.

Fig. 4. Acne detectors are able to provide numerical evidence for the lowering of facial acne severity. Acne lesions are located with green boxes.

component modules, is capable of producing a broad range of facial skin status estimate upon facial images. Once a selfie image is uploaded via this applet, it is fed into the tensorflow serving which hosts many models including the acne detection model. Different prediction tasks run parallel. The acne bounding boxes are thus obtained, along with a facial skin mask obtained by a model of facial skin segmentation based on BiSeNet. After filtering by facial skin mask to prevent misdetection in the background, final prediction is made and provided for higher level attributes calculation of the face and for front-end display.

This could help the users and perhaps their dermatologists conducting continuous out-of-hospital self-monitoring of the users' facial acne states, for example acne lesion number, severity, and moreover progression of the acne and estimation on therapies in use, etc. Example can be drawn just from the dataset ACNE04 as in Fig. 4. The 3 images come in a row with consecutive file names in this dataset. The severity is lowering, with numerical evidence being able to be drawn from the acne detector.

6.2. Limitations

As stated in [14], acne lesions are small in size, hence detectors tailored for small object detection perhaps might have better performances than general-purposed models on the problem of acne lesion locating, although shrinking the anchors is a choice but not the best. There were already research work into the topic of small objects detection, for example [33].

Another possible limitation is that we had only one class for acne. However, acne could further be classified into 6 or more subtypes[6] and even including acne scars. It should be noticed that in [13], the actual classification of acne severity was performed by counting lesions of papules and pustules, not all subtypes of acne. Confusion with other facial skin conditions, like rosacea and eczema, and skin pigmentation is another issue that is not avoidable for developing better facial acne lesion detectors, or for developing other deep learning analysis systems for facial skin conditions.

6.3. Future work

As stated in the above Subsection 5.2, a direction of our future work can be the development of a multi-class detector, which not only splits acne into 6 or more subtypes but also takes other facial skin conditions into consideration. Besides, tailoring the model network structures for more suited ones for detecting small objects is also a major concern in our future studies.

[6] Whitehead comedone, blackhead comedone, papule, pustule, nodule, cyst, etc., as in [34].

7. Conclusion

In this study, we verified the validity of several deep learning object detection models on the problem of facial acne analysis. According to the results of our experiments, we could conclude that the various detectors (Yolo, R-CNN family, but excluding the legacy SSD) are capable of relatively precisely locating out the acne lesions on facial images. Object detection is a promising method for facial acne analysis, whose potential should be more thoroughly studied.

Acknowledgments

The authors would like to thank the engineers of the innovative medical research group of the JingDong Health Corporation. Without their help in deploying the infrastructures and their helpful discussions, this research would never have had the chance to be carried out and completed. The authors would also like to thank Wu et al. [14] for sharing their research data, and the anonymous referees for their useful comments and suggestions. This work is supported by NSFC under grant No. 61872206.

Conflict of interest

None to report.

References

[1] Majumder S, Deen MJ. Smartphone sensors for health monitoring and diagnosis. Sensors. 19(9) (2019), 2164. doi: 10.3390/s19092164.

[2] Wong A, Scharcanski J, Fieguth P. Automatic skin lesion segmentation via iterative stochastic region merging. IEEE Transactions on Information Technology in Biomedicine. 15(6) (2011), 929–936. doi: 10.1109/titb.2011.2157829.

[3] Yang J, Sun X, Liang J, Rosin PL. Clinical Skin Lesion Diagnosis Using Representations Inspired by Dermatologist Criteria. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018, pp. 1258–1266. doi: 10.1109/cvpr.2018.00137.

[4] Zhao T, Zhang H, Spoelstra J. A Computer Vision Application for Assessing Facial Acne Severity from Selfie Images. arXiv preprint arXiv:1907.07901. 2019.

[5] Maier T, Kulichova D, Schotten K, Astrid R, Ruzicka T, Berking C, Udrea A. Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result. Journal of the European Academy of Dermatology and Venereology. 29(4) (2015), 663–667. doi: 10.1111/jdv.12648.

[6] Thissen M, Udrea A, Hacking M, Von Braunmuehl T, Ruzicka T. mHealth App for Risk Assessment of Pigmented and Nonpigmented Skin Lesions – A Study on Sensitivity and Specificity in Detecting Malignancy. Telemedicine and e-Health. 23(12) (2017), 948–954. doi: 10.1089/tmj.2016.0259.

[7] De Carvalho TM, Noels E, Wakkee M, Udrea A, Nijsten T. Development of Smartphone Apps for Skin Cancer Risk Assessment: Progress and Promise. JMIR Dermatology. 2(1) (2019), e13376. doi: 10.2196/13376.

[8] Udrea A, Mitra G, Costea D, Noels E, Wakkee M, Siegel D, De Carvalho T, Nijsten T. Accuracy of a smartphone application for triage of skin lesions based on machine learning algorithms. Journal of the European Academy of Dermatology and Venereology. 34(3) (Oct. 2019), 648–655. doi: 10.1111/jdv.15935.

[9] Wu Z, Zhao S, Peng Y, He X, Zhao X, Huang K, Wu X, Fan W, L. F, Chen M, Li J, Huang W, Chen X, Li Y. Studies on Different CNN Algorithms for Face Skin Disease Classification Based on Clinical Images. IEEE Access. 7 (2019), 66505–66511. doi: 10.1109/access.2019.2918221.

[10] Xie B, He X, Zhao S, Li Y, Su J, Zhao X, Kuang Y, Wang Y, Chen X. XiangyaDerm: A Clinical Image Dataset of Asian Race for Skin Disease Aided Diagnosis. Large-Scale Annotation of Biomedical Data and Expert Label Synthesis and Hardware Aware Learning for Medical Imaging and Computer Assisted Intervention. 2019, 22–31. doi: 10.1007/978-3-030-33642-4_3.

[11] Liu Y, et al. A deep learning system for differential diagnosis of skin diseases. Nature Medicine. 26(6) (May 2020), 900–908. doi: 10.1038/s41591-020-0842-3.
[12] Williams HC, Dellavalle RP, Garner S. Acne vulgaris. The Lancet. 379(9813) (2012), 361–372. doi: 10.1016/S0140-6736(11)60321-8.
[13] Hayashi N, Akamatsu H, Kawashima M, Group AS. Establishment of grading criteria for acne severity. The Journal of Dermatology. 35(5) (2008), 255–260. doi: 10.1111/j.1346-8138.2007.00403.x-i1.
[14] Wu X, Wen N, Liang J, Lai YK, She D, Cheng MM, Yang J. Joint Acne Image Grading and Counting via Label Distribution Learning. Proceedings of the IEEE International Conference on Computer Vision. 2019, pp. 10642–10651. doi: 10.1109/iccv.2019.01074.
[15] Fujii H, Yanagisawa T, Mitsui M, Murakami Y, Yamaguchi M, Ohyama N, Abe T, Yokoi I, Matsuoka Y, Kubota Y. Extraction of Acne Lesion in Acne Patients from Multispectral Images. 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2008, pp. 4078–4081. doi: 10.1109/iembs.2008.4650105.
[16] Chang C, Liao H. Automatic facial spots and acnes detection system. Journal of Cosmetics, Dermatological Sciences and Applications. 3(1) (2013), 28–35. doi: 10.4236/jcdsa.2013.31a006.
[17] Abas FS, Kaffenberger B, Bikowski J, Gurcan MN. Acne Image Analysis: Lesion Localization and Classification. Medical Imaging 2016: Computer-Aided Diagnosis. 9785 (2016), 97850. doi: 10.1117/12.2216444.
[18] Xu L, Jackowski M, Goshtasby A, Roseman D, Bines S, Yu C, Dhawan A, Huntley A. Segmentation of skin cancer images. Image and Vision Computing. 17(1) (1999), 65–74. doi: 10.1016/S0262-8856(98)00091-2.
[19] Tang J. A multi-direction GVF snake for the segmentation of skin cancer images. Pattern Recognition. 42(6) (2009), 1172–1179. doi: 10.1016/j.patcog.2008.09.007.
[20] Li C, Kao CY, Gore JC, Ding Z. Minimization of region-scalable fitting energy for image segmentation. IEEE Transactions on Image Processing. 17(10) (2008), 1940–1949. doi: 10.1109/tip.2008.2002304.
[21] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, Oct. 2017, pp. 618–626. doi: 10.1109/iccv.2017.74.
[22] Chattopadhay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). 2018, pp. 839–847. doi: 10.1109/wacv.2018.00097.
[23] Sun X, Yang J, Sun M, Wang K. A Benchmark for Automatic Visual Classification of Clinical Skin Disease Images. European Conference on Computer Vision. 2016, pp. 206–222. doi: 10.1007/978-3-319-46466-4_13.
[24] Deng J, Dong W, Socher R, Li LJ, Li K, Li FF. ImageNet: A Large-Scale Hierarchical Image Database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009, pp. 248–255. doi: 10.1109/cvpr.2009.5206848.
[25] Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934. 2020.
[26] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC. SSD: Single Shot MultiBox Detector. Lecture Notes in Computer Science. 2016, 21–37. doi: 10.1007/978-3-319-46448-0_2.
[27] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 39(6) (2017), 1137–1149. doi: 10.1109/tpami.2016.2577031.
[28] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017, pp. 7263–7271. doi: 10.1109/cvpr.2017.690.
[29] Everingham M, Van Gool L, Williams CK, Winn J, Zisserman A. The pascal visual object classes (VOC) challenge. International Journal of Computer Vision. 88(2) (2010), 303–338. doi: 10.1007/s11263-009-0275-4.
[30] Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft COCO: Common Objects in Context. European Conference on Computer Vision. 2014, pp. 740–755. doi: 10.1007/978-3-319-10602-1_48.
[31] Huang J, Rathod V, Sun C, Zhu M, Korattikara A, Fathi A, Fischer I, Wojna Z, Song Y, Guadarrama S, Murphy K. Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017, pp. 7310–7311. doi: 10.1109/cvpr.2017.351.
[32] Abadi M, et al. Tensorflow: A System for Large-Scale Machine Learning. 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16). 2016, pp. 265–283.
[33] Unel FO, Ozkalayci BO, Cigla C. The Power of Tiling for Small Object Detection. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, June 2019, pp. 582–591. doi: 10.1109/cvprw.2019.00084.
[34] Tuchayi SM, Makrantonaki E, Ganceviciene R, Dessinioti C, Feldman SR, Zouboulis CC. Acne vulgaris. Nature Reviews Disease Primers. 1(1) (2015), 1–20. doi: 10.1038/nrdp.2015.29.