Clinical, Cosmetic and Investigational Dermatology

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

A Deep Learning-Based Facial Acne Classification System

Andrea Quattrini[1], Claudio Boër[1], Tiziano Leidi[1], Rick Paydar[2]

[1]Department of Innovative Technologies, Institute of Information Technologies and Networking (SUPSI), Lugano, Switzerland; [2]CHOLLEY SA, Pambio-Noranco, Switzerland

Correspondence: Andrea Quattrini, Department of Innovative Technologies, Institute of Information Technologies and Networking (SUPSI), Polo Universitario Lugano – Campus Est, Via La Santa 1, Lugano, CH-6962, Switzerland, Tel +41765397810, Email andrea.quattrini@supsi.ch

Clinical, Cosmetic and Investigational Dermatology downloaded from https://www.dovepress.com/
For personal use only.

Introduction: Acne is one of the most common pathologies and affects people of all ages, genders, and ethnicities. The assessment of the type and severity status of a patient with acne should be done by a dermatologist, but the ever-increasing waiting time for an examination makes the therapy not accessible as quickly and consequently less effective. This work, born from the collaboration with CHOLLEY, a Swiss company with decades of experience in the research and production of skin care products, with the aim of developing a deep learning system that, using images produced with a mobile device, could make assessments and be as effective as a dermatologist.

Methods: There are two main challenges within this task. The first is to have enough data to train a neural model. Unlike other works in the literature, it was decided not to collect a proprietary dataset, but rather to exploit the enormity of public data available in the world of face analysis. Part of Flickr-Faces-HQ (FFHQ) was re-annotated by a CHOLLEY dermatologist, producing a dataset that is sufficiently large, but still very extendable. The second challenge was to simultaneously use high-resolution images to provide the neural network with the best data quality, but at the same time to ensure that the network learned the task correctly. To prevent the network from searching for recognition patterns in some uninteresting regions of the image, a semantic segmentation model was trained to distinguish, what is a skin region possibly affected by acne and what is background and can be discarded.

Results: Filtering the re-annotated dataset through the semantic segmentation model, the trained classification model achieved a final average f1 score of 60.84% in distinguishing between acne affected and unaffected faces, result that, if compared to other techniques proposed in the literature, can be considered as state-of-the-art.

Keywords: semantic segmentation, image classification, acne detection, dermatologists, computer vision

Introduction

Acne vulgaris (acne) is a chronic inflammatory disease of the sebaceous gland, caused by elevated testosterone levels and excessive colonization by Cutibacterium acnes, characterized by follicular hyperkeratinization, which leads to immune reactions and inflammation.[1] Although it is considered a predominantly youthful problem (with 95% of boys and 83% of girls at 16 years old experiencing it),[2] acne vulgaris actually also affects adult patients, but more particularly women in its two specific forms: late acne (also known as persistent, when emerging in adolescence and continuing to adulthood; accounting for 80% of cases) and late-onset (if first presenting in adulthood).[3] It is believed that genetic and hormonal factors contribute to the pathogenesis of AFA (Adult Female Acne), characterized by chronic evolution, requiring maintenance treatment, in some cases for years.[4] Acne lesions are located on the face, neck, chest and back. While not a serious disorder, acne, when it occurs in a severe form, can induce unsightly and permanent scarring. Both acne and the resulting scars can negatively affect the psyche. Psychological issues such as dissatisfaction with appearance, embarrassment, self-consciousness, lack of self-confidence, and social dysfunction such as reduced/avoidance of social interactions with peers and opposite gender, reduced employment opportunities have been documented.[5,6] For these reasons, timely diagnosis and treatment are important and desirable. The enormous spread of the disease, combined with the lack of availability of dermatologists, means that the waiting time for a visit is very long, even over a month.[7] To address this issue, we collaborated with CHOLLEY, Swiss manufacturer with decades of experience in the research and production of skin

Received: 23 February 2022
Accepted: 23 April 2022
Published: 11 May 2022

© 2022 Quattrini et al. This work is published and licensed by Dove Medical Press Limited. The full terms of this license are available at https://www.dovepress.com/terms.php and incorporate the Creative Commons Attribution – Non Commercial (unported, v3.0) License (http://creativecommons.org/licenses/by-nc/3.0/). By accessing the work you hereby accept the Terms. Non-commercial uses of the work are permitted without any further permission from Dove Medical Press Limited, provided the work is properly attributed. For permission for commercial use of this work, please see paragraphs 4.2 and 5 of our Terms (https://www.dovepress.com/terms.php).

care products, to develop an automated deep learning system that, based on selfie images taken with a mobile device, could carry out the same tests performed by a dermatologist, without the need to arrange an in-person visit. This would effectively solve the problem of the long waiting list for an examination and at the same time would allow CHOLLEY to incredibly improve its business network. In fact, the deep learning based dermatological examination would enable the possibility to propose, always remotely, a therapy based on CHOLLEY products, purchasable through online shop.

State of the Art

To the best of our knowledge, there is very little existing work in this area, and the topic of deep learning has not yet gained much traction in this context; Zhao et al[8] were the first to propose similar work. They collected about 1000 selfie pictures had them noted by different dermatologists accordingly to Investigator's Global Assessment (IGA) with five ordinal grades (clear-, almost clear-, mild-, moderate-, and severe-acne). Each annotation has been assigned to the whole face. Anyway, they proposed to analyse with a neural model small patches of the face automatically extracted from the position of facial landmarks.[10] The idea of analysing specific patches (placed in correspondence of cheeks, forehead and chin), in order to provide the neural model with only the sensitive information, would however exclude fundamental areas, such as the nose. In addition, assigning the same label (the face label) to each of the patches could cause wrong ground truth labels. Within[9] a dataset of 472 pictures always annotated using the IGA scale has been built; the whole face is analysed and classified in three different classes groups (IGA0-1, IGA2, IGA3-4). Best accuracy reported is 67%; anyway, since the test set is made by only 98 images, there is no real proof that the network learned the task effectively. Unfortunately, both datasets have not been made public. Junayed et al[11] used a custom model based on Deep Residual Neural Network[18] to classify different severe acne types using a dataset made public on dermnet.com. Consequently, the limitations of all the previous work are that of the scarcity in size of the dataset and the ability to ensure that (with so little data available) the network is really able to learn the task assigned to it. To solve this, in this paper we propose to:

1. Exploit the fact that there are a lot of public datasets available regarding facial analysis. Re-annotating one of these would provide enough data to train neural models.
2. Use semantic segmentation model to filter out all non-interesting regions in the selfie image. This is to avoid extracting patches and potentially excluding areas of interest.
3. Verify the correct training of the neural network by means of post-hoc attention techniques.

Methodology

The methodology presented here aims to overcome the problems highlighted in.[8] The authors themselves pointed out that the strategy of creating the training dataset likely introduced false class assignments to the annotated samples. 11 dermatologists annotated the available proprietary dataset by assigning each face a level of acne severity from 5 available levels (0-Not Acne, 1-Clear, 2–Almost Clear, 3-Mild, 4-Moderate, and 5-Severe). From those faces, whose ground truth assignment is unique, up to four different patches (depending on the orientation of the face) were subsequently extracted, with the aim of extracting the regions usually most affected by acne (cheeks, forehead and chin). Each patch extracted in this way was assigned the ground truth for the whole face. These patches, augmented by rolling skin augmentation technique, were used as training samples for a state-of-the-art neural network (a VGG-19 network).[12] The proposed strategy, starting from the same annotation technique (one class for the whole face) proposes instead to treat each face as a single sample, to maintain consistency between the annotation and the data. In order to exclude regions not belonging to the face (such as the background) and at the same time analyse the areas most affected by acne, a semantic segmentation model was implemented, based on the work in.[13] At this stage, the aim is to feed the neural model with only the information that is actually useful for the current goal. In this sense, not only the background is eliminated from the image but also all regions of the face that cannot be affected by acne problems (lips, eyes, hair, beard, …). In addition, after creating the segmentation map and filtering the image accordingly, the smallest axis aligned bounding box including that filtered face was extracted from the original image. This ensures that the highest possible resolution is propagated to the next step. The acne detection problem is then approached as a classification problem (ie, assigning each pre-processed sample to its own label) through the use of a state-of-the-art CNN. Specifically, a DenseNet121 type

architecture was used.[14] As the samples used represent the whole face, there is no need for customised data augmentation techniques in this case; random rotation, translation, zoom, horizontal flip and brightness variation were used to increase the generalisation capacity of the model. In Subsection Datasets, the datasets used are illustrated, while in Section Results, the results obtained and the comparison with the state of the art are proposed.

Datasets

There is not much availability of datasets for this type of task. However, the subject of face analysis by deep learning is quite active. Consequently, there are many public datasets available, but not annotated for this specific situation. Taking advantage of the knowledge of an expert in the field, some datasets used in the face field have been re-collected for our needs. Specifically, all the datasets used are listed below.

- CelebAMask-HQ is a large-scale face image dataset that has 30,000 high-resolution face images selected from the CelebA dataset by following CelebA-HQ. Each image has segmentation mask of facial attributes corresponding to CelebA. The masks of were manually annotated with the size of 512×512 and 19 classes including all facial components and accessories such as skin, nose, eyes, eyebrows, ears, mouth, lip, hair, hat, eyeglass, earring, necklace, neck, and cloth.[16]
- Flickr-Faces-HQ (FFHQ) is a high-quality image dataset of human faces, originally created as a benchmark for generative adversarial networks (GAN). The dataset consists of 70,000 high-quality PNG images at 1024×1024 resolution and contains considerable variation in terms of age, ethnicity and image background. It also has good coverage of accessories such as eyeglasses, sunglasses, etc.[15]

The CelebAMask-HQ dataset provided 19 different classes (skin, nose, eyes (left and right), eyebrows (left and right), ears (left and right), mouth, lip, hair, hat, eyeglass, earring, necklace, neck, and cloth, background). Since we were interested in extracting just two different classes (acne-affectable skin and background), skin and nose have been fused together and all the other classes have become the background class. On the other hand, part of the Flickr- Faces-HQ (FFHQ) dataset has been re-annotated by an experienced dermatologist who placed each sample in a specific class among the following (0 - No Acne, 1 - Light Acne, 2 - Moderate Acne, 3 - Severe Acne). In total, the annotation process covered 2307 images, not the whole dataset (to have consistency of evaluation only one person did the work). Unfortunately, the dataset, at the end of the annotation process was particularly unbalanced with almost 88% of the samples belonging to the "0-No Acne" class, and the remaining distributed among the "positive" classes. Due to the statistical insignificance of the sample for the other classes, these were aggregated into a single category and the problem was transformed into a binary classification problem.

Performance Metrics

The performances of the semantic segmentation model have been assessed in terms of Dice Coefficient, also called F1-Score, which is the standard way to evaluate such a model. Given, for each specific class, ground truth pixels and predicted pixels, the Dice coefficient is computed as follows.

$$Dice\ Coefficient = 2 \cdot \frac{Area\ Intersection}{Area\ Union}$$

First, the Dice Coefficient for each semantic class is computed and then the final value is the average over classes. Pixel-wise cross entropy loss has been used. This loss examines each pixel individually, comparing the class predictions (depth-wise pixel vector) to a one-hot encoded target vector. As far as the classification model is concerned, since the dataset is particularly unbalanced, the evaluation performance of the process was assessed in terms of F1 Score per class and average F1 Score. During the training process, the classical categorical cross entropy was used as a loss function and the classes were balanced using class weights.

$$F1 = \frac{TP}{TP \cdot \frac{1}{2}(FP + FN)}$$

Where TP, FP and FN are, respectively, true positives, false positives and false negatives.

Table 1 Sermantic Segmentation Training Results

	Nose	Skin	Background	Average
F1Score [%]	94.1	92.4	71.9	86.1

Results

Segmentation model has been trained on a HPC Cluster, equipped with two NVIDIA Tesla T4. SGD has been used as a optimizer with an initial learning rate equal to 1e-2, momentum and weight decay, respectively, equal to 0.9 and 1e-5. After 80.000 iterations, the final average Dice Coefficient was 86.1%. Results of the training process are summarized in Table 1.

In Figure 1, on the left an example of an original image. In the middle, a segmentation map, where blue is the color for the background and green the color for the acne-affectable skin. On the right, the final rescaled preprocessed image. As for the classification model, the model was trained for 50 epochs by transfer learning from a pretrained model on ImageNet. Adam with an initial learning rate of 3e-4 and label smoothing was used as the optimiser to help the model generalise, given the modest size of the dataset. As the amount of data in this dataset is quite small, the training was performed on CPU on a 2018 MacBook Pro 2.2 GHz Intel Core i7 6 core. In Table 2 are the results of the classification process, split into the two categories. Although the dataset is highly unbalanced, the neural model, although not too deep, is able to distinguish samples with good reliability, even in the least populated class.

To verify the quality and correctness of the results, Grad-CAM[17] was applied to the samples with positive acne, Figures 2–5. Grad-CAM uses the gradients of any target concept, flowing into the final convolutional layer to produce a coarse localization map highlighting the important regions in the image for predicting the concept. The warmer the colour of the heatmap, the more important that area was in the predictive process of the neural model. It can be seen that in the examples in question, the area affected by acne is actually the one that activated the network.

Discussion

Teledermatology is an extremely modern discipline that makes it possible to provide real-time consultancy services to patients wherever they are. In this sense, the progress made in the last decade by the AI (Artificial Intelligence) sector has enabled a number of studies such as this one on teledermatology that were once unthinkable. Timing the start of an appropriate therapy has a positive impact both on recovery time and on prevention of psychological consequences. However, there is still a long and complex way to go before we have an artificial intelligence-based instrument capable of analysing dermatological pathologies at the same level as an experienced dermatologist. Although this study shows that, with good data availability, a neural network can learn to recognise a specific

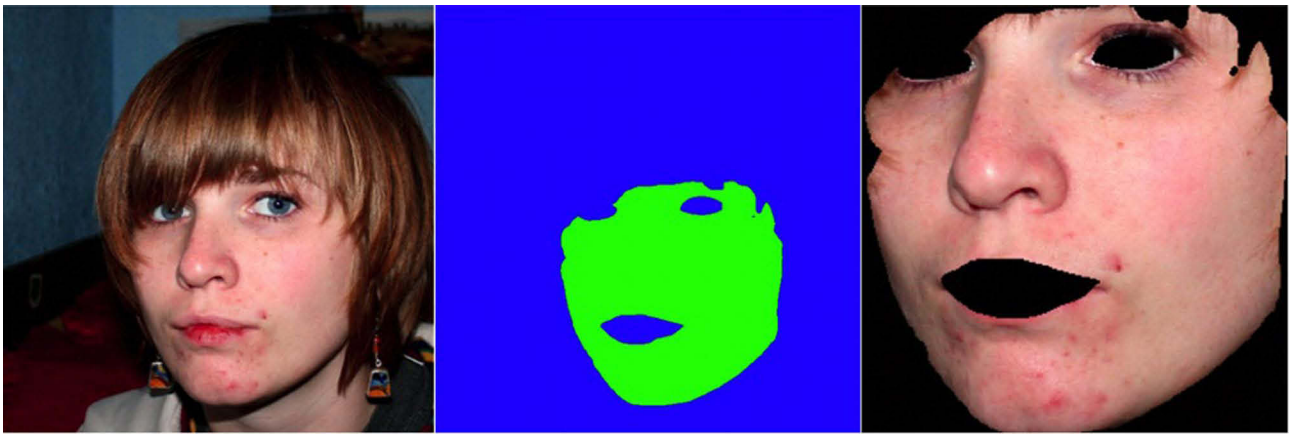

Figure 1 An example of preprocessing based on semantic segmentation, first subfigure is part of the FFHQ dataset (https://www.flickr.com/photos/cliche/3086726400/), author: Kate Brady, licensed under CC BY 2.0 (https://creativecommons.org/licenses/by/2.0/)/central subfigure: segmentation map computed on the first figure/right subfigure: final image used with the classification model.

Table 2 Classification Training Results

	No Acne	Acne	Average
Precision [%]	94.1	30.43	59.89
Recall [%]	82.51	43.75	63.13
F1 Score [%]	85.79	35.89	60.84

pathology, it must be considered that many others would have to be integrated to have a complete assessment of the patient's skin health. Last but not least, having the user take a selfie to be analysed can introduce additional problems related to image quality (over- and under-exposure, reflections, …), all of which are avoided in a dermatological visit.

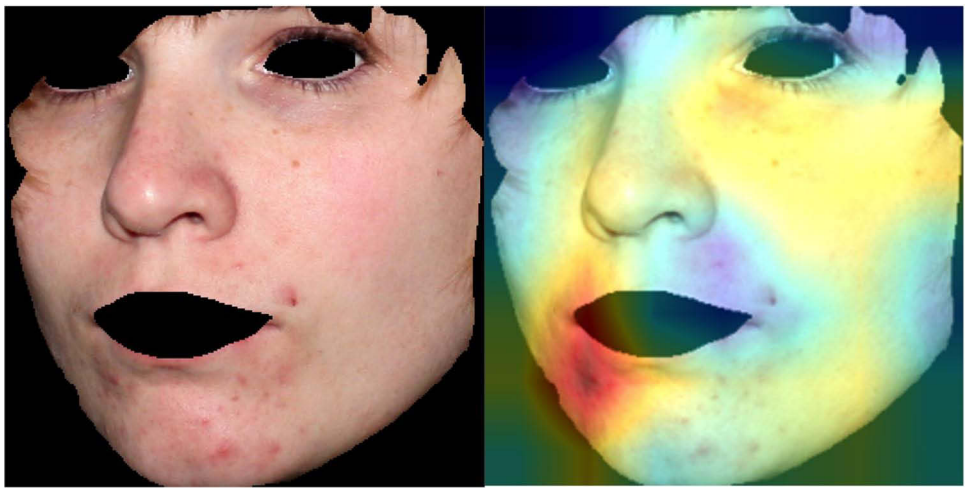

Figure 2 Grad-CAM of a positive sample, part of the FFHQ dataset (https://www.flickr.com/photos/cliche/3086726400/), author: Kate Brady, licensed under CC BY 2.0 (https://creativecommons.org/licenses/by/2.0/), modified through the application of Grad-CAM algorithm.

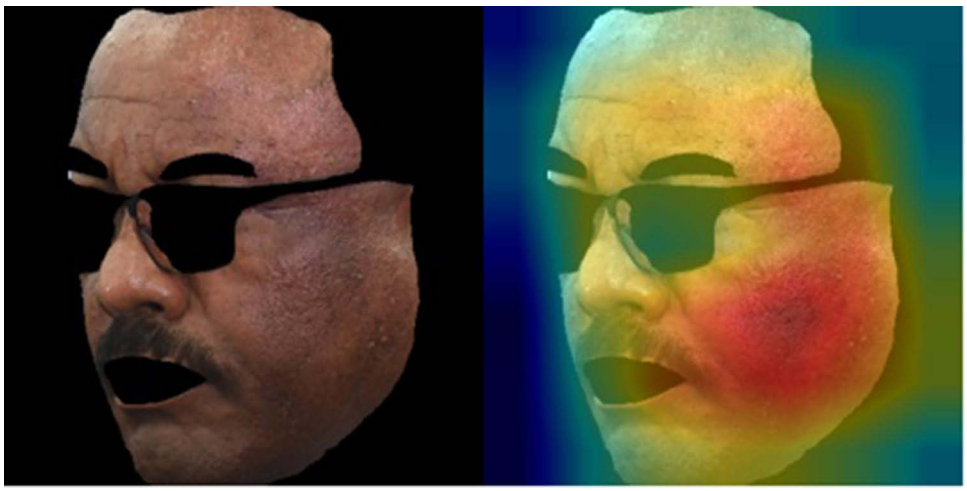

Figure 3 Grad-CAM of a positive sample, part of the FFHQ dataset (https://www.flickr.com/photos/au_unistphotostream/32895808952/), author: AMISOM Public Information, licensed under CC0 (https://creativecommons.org/publicdomain/zero/1.0/).

Figure 4 Grad-CAM of a positive sample, part of the FFHQ dataset (https://www.flickr.com/photos/rhythmicdiaspora/37814947704/), author: rhythmic diaspora, licensed under CC BY 2.0 (https://creativecommons.org/licenses/by/2.0/), modified through the application of Grad-CAM algorithm.

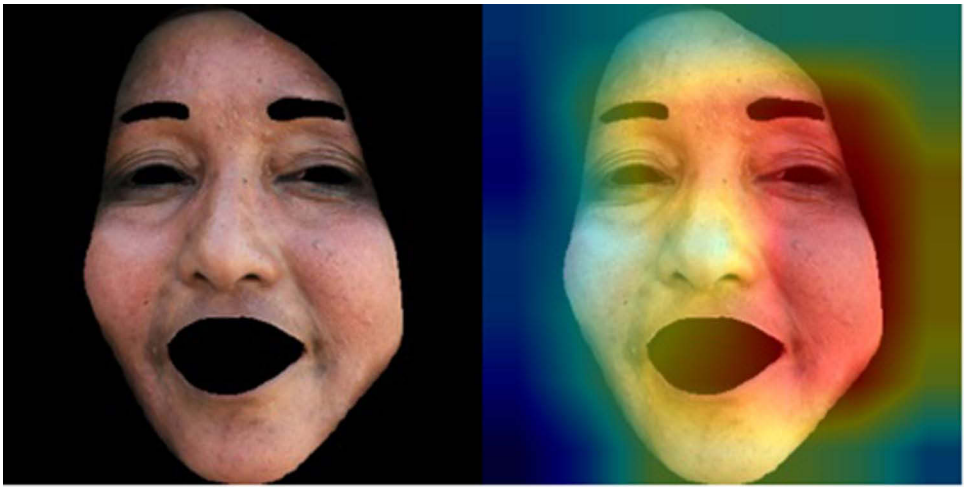

Figure 5 Grad-CAM of a positive sample, part of the FFHQ dataset (https://www.flickr.com/photos/au_unistphotostream/31369618020/), author: AMISOM Public Information, licensed under CC0 (https://creativecommons.org/publicdomain/zero/1.0/).

Conclusion

The study carried out on the use of deep learning techniques for the analysis of facial problems related to acne gave good insights. Firstly, thanks to CHOLLEY's work, the potential to exploit public datasets and re-annotate them proved effective. Although not all of the Flickr-Faces-HQ (FFHQ) dataset was annotated due to time constraints, about 12% of the processed sample has acne issues and can be used effectively for training neural models. If the same percentage was maintained over the entire dataset, this would result in approximately 8400 high-resolution images of faces with acne. No other dataset collected so far has such a large amount of data. Instead, using the CelebAMask-HQ dataset, a BiSeNet semantic segmentation model was trained to segment two different classes: acne-sensitive area and background. The acne-sensitive area includes all the face with the exception of eyes, eyebrows, lips, hair and beard. The high dimensionality of the dataset allowed for effective training resulting in a final mIoU value of 95.3%. The re-annotated Flickr-Faces-HQ (FFHQ) dataset, filtered through the semantic segmentation model, was finally trained to distinguish between the "Acne" and "No Acne" classes, resulting in an average F1 Score of 60.84%. To verify the effectiveness of the annotated model, Grad-CAM was used and confirmed the validity of the result.

Ethics and Consent

This project does not need an approval from an Ethic Committee, since the project does not fall under the Human Research Act (HRA), according to Articles 2 and 3.

Acknowledgments

The research in this paper has been funded by Innosuisse – Swiss Innovation Agency, Application Number: 48961.1 INNO-ICT.

Disclosure

Prof. Dr. Tiziano Leidi reports grants from Innosuisse - Swiss Innovation Agency, during the conduct of the study. The authors report no other conflicts of interest in this work.

References

1. Leccia MT, Auffret N, Poli F, Claudel JP, Corvec S, Dreno B. Tropical acne treatments in Europe and the issue of antimicrobial resistance. *J Eur Acad Dermatol Venereol*. 2015;29:1485–1492. doi:10.1111/jdv.12989
2. Heng AHS, Chew FT. Systematic review of the epidemiology of acne vulgaris. *Sci Rep*. 5754;10:1–29.
3. Chiilicka K, Rogowska AM, Szygula R, Dziendziora-Urbinska I, Taradai J. A comparison of the effectiveness of azelaic and pyruvic acid peels in the treatment of female adult acne: a randomized controlled trial. *Sci Rep*. 2020;10:12612. doi:10.1038/s41598-020-69530-w
4. Dréno B, Thiboutot D, Layton AM, Berson D, Perez M, Kang S. Global alliance to improve outcomes in acne large-scale international study enhances understanding of an emerging acne population: adult females. *J Eur Acad Dermatol Venereol*. 2015;29:1096–1106. doi:10.1111/jdv.12757
5. Tan JK. Psychosocial impact of acne vulgaris: evaluating the evidence. *Skin Therapy Lett*. 2004;9:1–3, 9.
6. Magin P, Adams J, Heading G, Pond D, Smith W. Psychological sequelae of acne vulgaris: results of a qualitative study. *Can Fam Physician*. 2006;52:978–979.
7. Tan JKL, Bhate K. A global perspective on the epidemiology of acne. *Br J Dermatol*. 2015;172:3–12. doi:10.1111/bjd.13462
8. Zhao T, Zang H, Spoelstra J. A computer vision application for assessing facial acne severity from selfie images. *Preprint*. 2019. doi:10.48550/arXiv.1907.07901
9. Ziying VL, Farhan A, Cuong PN, et al. Automated grading of acne vulgaris by deep learning with convolutional neural networks. *Skin Res Technol*. 2019;2:187–192.
10. Wu Y, Ji Q. Facial landmark detection: a literature survey. *Int J Comput Vision*. 2018;127:115–142. doi:10.1007/s11263-018-1097-z
11. Junayed MS, Jeny AA, Atik ST, Neehal N. AcneNet - a deep CNN based classification. Approach for acne classes. 12th International Conference on Information Communication Technology and System (ICTS). 2019.
12. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiV:1409.1556*. 2014. doi:10.48550/arXiv.1409.1556
13. Yu C, Wang J, Peng C, et al. BiSeNet: bilateral segmentation network for real-time semantic segmentation. *ECCV*. 2018;2018:334–349.
14. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. *ECCV*. 2016;2016:630–645.
15. Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. *arXiv:1812.04948*. 2018. doi:10.48550/arXiv.1812.04948
16. Lee C, Liu Z, Wu L, Luo P. MaskGAN: towards diverse and interactive facial image manipulation. *arXiv*. 2019. doi:10.48550/arXiv.1907.11922.
17. Selvaraju R, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *ICCV*. 2017. doi:10.1007/s11263-019-01228-7
18. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. arXiv:1512.03385; 2015. Available from: https://arxiv.org/abs/1512.03385. Accessed May 01, 2022.

Clinical, Cosmetic and Investigational Dermatology Dovepress

Publish your work in this journal

Clinical, Cosmetic and Investigational Dermatology is an international, peer-reviewed, open access, online journal that focuses on the latest clinical and experimental research in all aspects of skin disease and cosmetic interventions. This journal is indexed on CAS. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

Submit your manuscript here: https://www.dovepress.com/clinical-cosmetic-and-investigational-dermatology-journal