computers

MDPI

Article

Attention Mechanism Guided Deep Regression Model for Acne Severity Grading

Saeed Alzahrani [1,*], Baidaa Al-Bander [2] and Waleed Al-Nuaimy [1]

[1] Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3GJ, UK; wax@liverpool.ac.uk
[2] Department of Computer Engineering, University of Diyala, Baqubah 32010, Iraq; baidaa.q@gmail.com
* Correspondence: s.g.a.alzahrani@liverpool.ac.uk

Abstract: Acne vulgaris is the common form of acne that primarily affects adolescents, characterised by an eruption of inflammatory and/or non-inflammatory skin lesions. Accurate evaluation and severity grading of acne play a significant role in precise treatment for patients. Manual acne examination is typically conducted by dermatologists through visual inspection of the patient skin and counting the number of acne lesions. However, this task costs time and requires excessive effort by dermatologists. This paper presents automated acne counting and severity grading method from facial images. To this end, we develop a multi-scale dilated fully convolutional regressor for density map generation integrated with an attention mechanism. The proposed fully convolutional regressor module adapts UNet with dilated convolution filters to systematically aggregate multi-scale contextual information for density maps generation. We incorporate an attention mechanism represented by prior knowledge of bounding boxes generated by Faster R-CNN into the regressor model. This attention mechanism guides the regressor model on where to look for the acne lesions by locating the most salient features related to the understudied acne lesions, therefore improving its robustness to diverse facial acne lesion distributions in sparse and dense regions. Finally, integrating over the generated density maps yields the count of acne lesions within an image, and subsequently the acne count indicates the level of acne severity. The obtained results demonstrate improved performance compared to the state-of-the-art methods in terms of regression and classification metrics. The developed computer-based diagnosis tool would greatly benefit and support automated acne lesion severity grading, significantly reducing the manual assessment and evaluation workload.

Keywords: acne diagnosis; deep learning; density map generation; attention network; regression models; Faster-RCNN

Citation: Alzahrani, S.; Al-Bander, B.; Al-Nuaimy, W. Attention Mechanism Guided Deep Regression Model for Acne Severity Grading. *Computers* **2022**, *11*, 31. https://doi.org/10.3390/computers11030031

Academic Editors: Antonio Celesti, Ivanoe De Falco, Antonino Galletta and Giovanna Sannino

Received: 18 January 2022
Accepted: 18 February 2022
Published: 23 February 2022

1. Introduction

Acne vulgaris, or acne, is a skin condition in which dead skin cells and oil from the skin block hair follicles. This skin condition is clinically featured by blackheads and whiteheads (open and closed comedones), small and tender red bumps (papules), white or yellow squeezable spots (pustules), cyst-like fluctuant swellings (cysts), and large painful red lumps (nodules). It usually affects areas of skin with a high number of oil glands, such as the face, chest, back and shoulders [1,2]. Facial acne is most common during adolescence, but it can persist into adulthood. After severe inflammatory acne, scarring inevitably occurs. The scarring might lead to significant psychosocial consequences and potential risk factors for serious mental health issues. The resultant facial appearance can cause anxiety, low self-esteem, and, in the worst-case scenario, depression or suicidal thoughts [3,4].

Acne vulgaris is simple to diagnose; however, its polymorphic structure makes it difficult to assess its severity. As the number of acne lesions varies during the course of the condition, numerous evaluation criteria based on clinical screening and photographic documentation have been established. Grading based on clinical examination, lesion counting, and approaches requiring instruments, such as photography, fluorescent photography,

polarised light photography, video microscopy, and sebum production measurement, are developed to assess the severity of acne vulgaris. Clinical examination (grading) and lesion counting are two widely used methods for acne severity assessment [2,5]. Clinical grading is a subjective approach that entails analysing the dominating lesions, assessing the occurrences of inflammation, and measuring the degree of involvement to determine the severity of acne. On the other hand, acne lesion counting-based method involves counting the number of a certain kind of acne lesion and then evaluating the overall severity [5].

Acne severity has also been measured via photography, which involves comparing patients to a photographic standard. This method has many disadvantages, including the inability to palpate the depth of involvement and the difficulty to visualise small lesions. When it comes to determining the density of comedones, fluorescence and polarised light photography can offer some advantages over standard photography. However, there are some shortcomings, such as a substantial time commitment and the necessity for more complicated types of equipment [6]. In 2008, Hayashi et al. [7] presented a grading method to classify acne lesions into four types using standard photographs and lesion counting. On half of each patient's face, they counted the number of open and closed comedones, papules, pustules, cysts, and nodules. They categorised the eruptions into three groups: comedones, inflammatory eruptions (including papules and pustules), and severe eruptions (including cysts and nodules). They graded the severity of acne as (i) mild when the acne count is (0–5), (ii) moderate when the acne count is (6–20), (iii) severe when the acne count is (21–50), and (iv) very severe when the acne count is more than 50, based on the number of inflammatory eruptions (papules, pustules) or lesions on half of the face.

A physician's validated assessment generally determines the effectiveness of acne treatment. For assessment by the physician, the different acne lesion types involve being counted independently. Acne affects about 80% of adolescents [8], with 3% of men and 12% of women experiencing symptoms even through adulthood [9]. As a result, there are a large number of acne patients who require immediate treatment, as acne can cause scars and pigmentation as well as a sense of inferiority and depression [10]. Dermatologists need to know the severity of acne to make a precise and appropriate treatment selection [7]. However, due to the limited time available for consultation, the manual validated evaluation of acne might be difficult and time-consuming. Additionally, junior dermatologists need a reference diagnosis that is objective and trustworthy. With the development of imaging modalities, widespread availability of digital cameras, and deep learning (DL) techniques, automatic acne detection and severity evaluation systems from photographs would help dermatologists attain a more reliable and consistent assessment of acne in clinical practice trials. Recently, deep learning (DL), especially Convolutional Neural Networks (CNN) algorithms, leveraging its hierarchical feature learning ability, have made a significant breakthrough in medical imaging. With adequate training data, representation learning may potentially outperform hand-designed features [11–13].

The remainder of this paper is presented as follows: following the short clinical overview of acne vulgaris, related work is given in Section 2. The description of the dataset used in this study and the proposed methodology are described in Section 3. The experimental results and findings are reported and discussed in Section 4. Finally, the proposed work is concluded in Section 5.

2. Related Work

Remarkable progress has been made for automated acne lesion analysis in recent years covering several acne lesion analysis tasks such as acne classification [14–17], segmentation [18–21], detection and localisation [16,19,20,22,23], and severity grading [20,24–28]. The analysis of acne lesions was accomplished by image processing techniques [19,21], extracting hand-crafted features and passing them into a classier model [16,20], and automated feature learning using CNNs [15,23,26]. In this work, we address the problem of acne severity grading from facial images.

Several methods have been proposed in the literature targeting the automated severity grading of acne lesions. In [20], hand-engineered features were extracted from segmented acne areas and passed into an SVM model to classify the severity of acne lesions into four levels following the criteria established by Ramli [29]. Their method was assessed on a private dataset composed of 35 images. Alternatively, the authors in [24–26,28] exploited CNNs to extract the features automatically and subsequently, graded the severity of acne lesions following the criteria established by IGA (three levels) [30], Hayashi (four levels) [7], GEA (five levels) [31], and IGA (five levels) [30], respectively. Those developed systems were trained and evaluated on private datasets consisting of 472, 4700, 5972, and 479 images, respectively. The authors in [27] presented acne counting and grading method based on label distribution learning paradigm (LDL) with CNN to classify the acne severity into four levels following Hayashi assessment criteria [7]. They evaluated the performance of the developed method on a public dataset of 1457 images. However, the performance of these developed approaches has limitations and experiences challenges. The performance of handcrafted feature regression-based methods highly depends on the type of features extracted from a specific dataset. Furthermore, those features might be applicable in a particular dataset but may not generalise well on other datasets. On the other hand, CNN regression-based methods globally estimate outcomes from features without concerning the detailed location of understudied acnes that should be considered following the grading criteria.

To tackle the aforementioned limitations, we developed a new computer-assisted image analysis approach to grade the severity of acne lesions called dilated UNet dense regressor guided by attention mechanism. Inspired by the scenario of crowed counting from kernel density maps [32,33], region of interest density maps for acne lesions are generated to produce the count of lesions within a particular area of interest. Thus, we propose a method to count objects of interest, represented by acne lesions, and subsequently grading the severity of acne in facial images. Following [34], we adopt fully convolutional UNet, which is originally used for segmentation, to construct the regressor responsible for generating the density maps. In addition, following [35], we exploit the multi-scale dilated filters to implement the bottleneck convolutional filters of UNet. Accordingly, we developed multi-scale dilated UNet regressor for density map generation. The proposed convolutional network module uses dilated convolution filters to systematically aggregate multi-scale contextual information trying to mitigate the lose in resolution. On the top of the multi-scale dilated UNet regressor, we embed the prior information of bounding boxes as attention mechanism generated by Faster R-CNN [36], which is originally developed for object detection. In this fashion, we merge the dilated UNet dense regressor with Faster R-CNN network for density map regression allowing us to determine the count of acne lesions and subsequently grade the severity.

Beyond the bounds of acne lesion counting, the concept of object counting has been widely applied in a variety of scenarios, including cell counting in microscopic images [37], tree counting [38], animal counting [39], vehicle counting [40], and crowd counting [41]. Generally, estimating the number of any objects in a still image or a video is typically defined as a counting problem. The object counting methods can be broadly divided into two categories: detection and regression-based techniques. The counting-by-detection approaches, which use detectors to detect each object in an image or video, were widely used in early efforts addressing the object counting topic. To extract low-level features, these approaches require well-trained classifiers such as HOG, histogram-oriented gradients [42], and Haar wavelets [43]. Recent approaches leveraging CNN-based object detectors to achieve end-to-end learning paradigms, such as YOLO3 [44], SSD [45], and Faster R-CNN [36], have considerably improved counting accuracy.

Different from counting by detection, regression-based approaches obtain the count without explicitly detecting and localising each object. Global regression and density estimation are the two types of regression-based counting techniques. Global regression methods [22,27] explicitly predict the final count from images by learning the mapping

between image features. In contrast, density estimation-based methods [32,46] first estimate a density map, which is then integrated (summed) to produce the final count. Density estimation typically outperforms global regression because it makes use of more spatial information of objects in an image. However, acne lesion counting based on either regression or detection approaches is insufficient to handle both high- and low-density regions of acne lesions simultaneously. When counting using regression solely, there is a risk of overestimation when there are low densities of objects (sparse regions). Similarly, counting by purely detection methods would result in the underestimation problem on occasions with high densities of objects (dense regions). Thus, counting by detection performs comparably better in the sparse regions; on the other hand, counting by regression performs comparably better in the dense areas [47]. This motivated us to establish a system that takes advantage of regression (Dilated UNet Regressor) potentials and impressing attention to the acne lesion positions detected by the detecter (Faster R-CNN), inspired by [48,49].

In general, the contribution of the presented work can be described as follows:

- Inspired by the scenario of crowed counting from kernel density maps and leveraging the advances of deep learning models, we propose a new method for acne counting and severity grading called dilated UNet dense regressor guided by attention mechanism.
- We modify the paths of contraction-expanding (encoder–decoder paths) in the UNet segmentation model by introducing a bounding box encoder that incorporates the box information generated by Faster R-CNN.
- This embedding adaptation helps to simultaneously handle high- and low-density region of acne lesions.
- The proposed regressor exploits dilated convolutions to aggregate multi-scale contextual details systematically.
- Experiments on public facial acne image datasets demonstrate the superiority of the proposed method compared with the state-of-the-art techniques.

3. Materials and Methods

3.1. Materials

To conduct the experiments in this research work, a publicly available dataset named ACNE04 is used [27]. The number of lesion and global acne severity are annotated in the ACNE04 dataset by specialists. Images of acne lesions are collected using a digital camera with patients' consent when physicians are making a diagnosis. Images are taken at a 70-degree angle from the front of patients to meet the requirements of the Hayashi grading criteria [7]. The specialists then manually annotate the images using the annotation tool provided. The ACNE04 contains 1457 images of lesions with 18,983 bounding boxes.

3.2. Methods

In this section, we describe the proposed attention guided UNet dense regressor for addressing the task of acne counting and severity grading in detail. The developed architecture incorporates dilated UNet dense regressor for density regression with the information of bounding boxes generated from Faster R-CNN network, producing a hybrid detection–regression framework. Figure 1 presents the abstract level of the proposed architecture for acne severity grading. We will first describe the ground truth generation of kernel density maps in Section 3.2.1, then we will illustrate the architecture of the proposed system in Section 3.2.2.

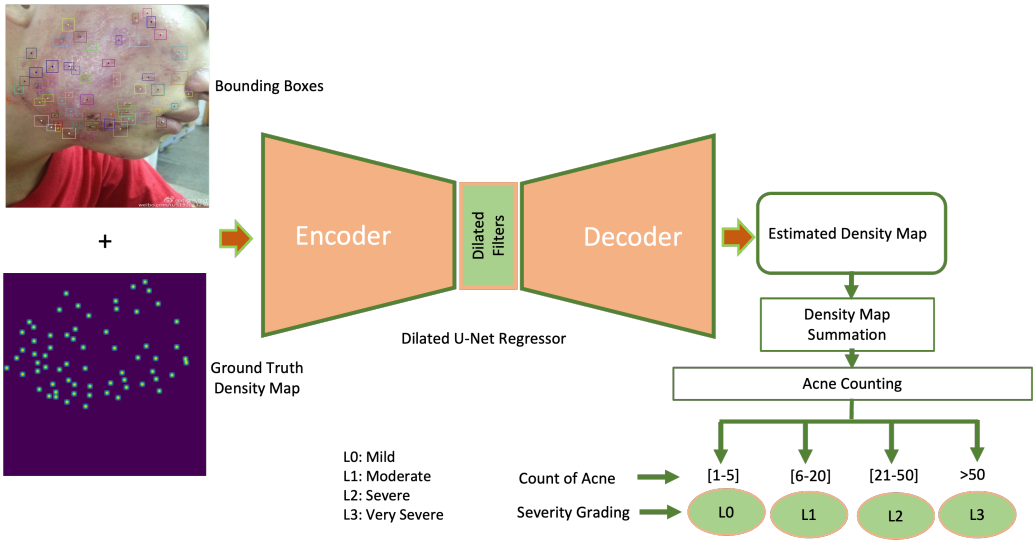


Figure 1. Block diagram of proposed acne counting and grading system.

3.2.1. Generation of Ground Truth Kernel Density Maps

Due to severe overlapping and variation in the size of the acne, individual acne detectors might encounter problems in locating facial skin lesions in dense regions. Hence, the challenge of acne counting is handled as estimating a kernel density function whose integral over each image region yields the number of acne in that image. Thus, the resulted density map would preserve information indicating the presence of lesions in a specific area. To estimate the acne density map from an input facial image, the UNet density map regressor is first trained on training facial images along with their ground truth density maps. The quality of generated ground truth density map for a given training image determines the performance of the developed method. To generate a map of acne density for training data, it is required to provide point-annotations for acne lesions. As the data used in this study were provided with bounding boxes around each acne, we first determined the centre point value of each bounding box around acne lesions producing pixels dot-annotation. To generate the density map $F(x)$ given a point at pixel x_i from total R acne lesions, the method for generating density maps used in [32] is followed by convolving $\delta(x - x_i)$ with Gaussian kernel G_σ. The Gaussian kernel is set with fixed spread parameter σ of 4 and kernel size of 15 by blurring each acne annotation point as follows:

$$F(x) = \sum_{i=1}^{R} \delta(x - x_i) * G_\sigma(x) \quad (1)$$

3.2.2. Dilated UNet Dense Regressor Guided by Attention Mechanism

The overall structure of the proposed dilated UNet dense regressor architecture with attention module is shown in Figure 2. We adapt the UNet encoder–decoder segmentation model by integrating bounding box information at the level of the skip connections. The outcome of the bounding boxes acts as an attention assistant module. Using element-wise multiplication, feature maps extracted at different scales from the contraction path are fused with features extracted from bounding boxes, then passed to the expanding path.

When it comes to adding attention blocks at the skip connection level, the Attention-UNet developed in [50] is similar to our model by inserting convolutional filters in the middle of the encoder and decoder paths. However, the structure of attention models used to focus on relevant features as well as the strategies to which each model establishes the constraints differ considerably. While we utilise bounding boxes to guide the network on where to seek through the network until reaching the bottleneck, Attention-UNet [50] employs inputs provided by the bottleneck output and moves upward through the skip-connections. Inserting the convolutional filters in the middle of the encoder and decoder

paths in our model helps the model to adjust what it learns by concentrating on the attention areas. This results in the enhancement of feature detection within specific regions of the facial image.

UNet [34] is a segmentation network architecture built upon fully convolutional neural networks (FCNs). Unlike FCNs, UNet adopts the symmetry structure of encoder and decoder (contraction and expanding paths). The UNet architecture consists of three sections: the contraction, the bottleneck, and the expansion section. UNet's contracting path (shown on the left in Figure 2) is similar to that of a standard CNN, with a combination of convolutional and max-pooling layers. It gradually decreases the size of feature maps while increasing the number of feature channels, allowing the model to learn both global and local features. The output size of the encoder path (contracting path) passing to the bottleneck is 1/16 of the original input size. If we keep adding convolutional and pooling layers to the bottleneck, the output size would be further downsized, making it difficult to produce high-quality density maps. Inspired by the work [35], dilated convolutional layers is deployed in the bottleneck to extract more saliency information while preserving the output resolution. A small-size kernel with a $k \times k$ filter is typically enlarged to $k + (k-1)(r-1)$ with a dilated stride parameter r in dilated convolution scheme. As a result, it enables flexible aggregation of multi-scale contextual information while maintaining the same resolution. A 2-D dilated convolution can be formulated as follows:

$$y(m,n) = \sum_{i=1}^{M} \sum_{j=1}^{N} x(m + r \times i, n + r \times j) w(i,j) \tag{2}$$

where $y(m,n)$ is the resulted dilated convolution from input $x(m,n)$ and filter weights $w(i,j)$ with the dimensions M and N, respectively. The parameter r represents the dilation rate. If the dilation rate $r = 1$, a dilated convolution returns back into a standard convolution. The third section of UNet, the expanding path (the right part in Figure 2), contains a succession of convolution and deconvolution components that can step-wise up-sample the feature maps to their original size and minimise the feature channels. The skip connections between the contracting and expanding paths combine and concatenate features from both sides, forcing the model to collect both local and global information.

This dilated UNet dense regressor is augmented with features of the parallel bounding boxes generated by Faster R-CNN in the skip connections between the encoder and decoder segmentation model. This helps to embed bounding box information as an attention mechanism for acne lesions at different scales in the model. The regression-based model (UNet) works well on dense acne lesions on the facial images, whereas the detection-based model (Faster R-CNN) provides better detection on sparse acne lesions. Thus, integrating the detection attention model in one framework with a regression model helps guide and bring the attention of the regressor to the sparse acne lesions that could be missed by dense regressor. The bounding boxes are fed independently to two convolutional layers (attention module) for location feature extraction. The bounding boxes provided to the attention model is a binary map representing the attention region that corresponds to the location of the acne lesions. The intersection of the un-pooled map from a level contracting layer and the feature map of acne lesions from the attention module is produced and concatenated with the features from the up-sampling layers within each skip connection. Finally, a 1×1 convolutional layer is applied to map the resultant feature vector to the density maps. The difference between the predicted density map and the ground truth is estimated using Euclidean distance. The following is the definition of the loss function:

$$L(\Theta) = \frac{1}{2B} \sum_{i=1}^{B} \| Z(X_i; \Theta) - Z_i^{GT} \|_2^2 \tag{3}$$

where B refers to the training batch size and $Z(X_i;\Theta)$ refers to the output produced in our model with Θ learnable parameters. X_i denotes the input image, and Z_i^{GT} is the ground truth of the input image X_i.

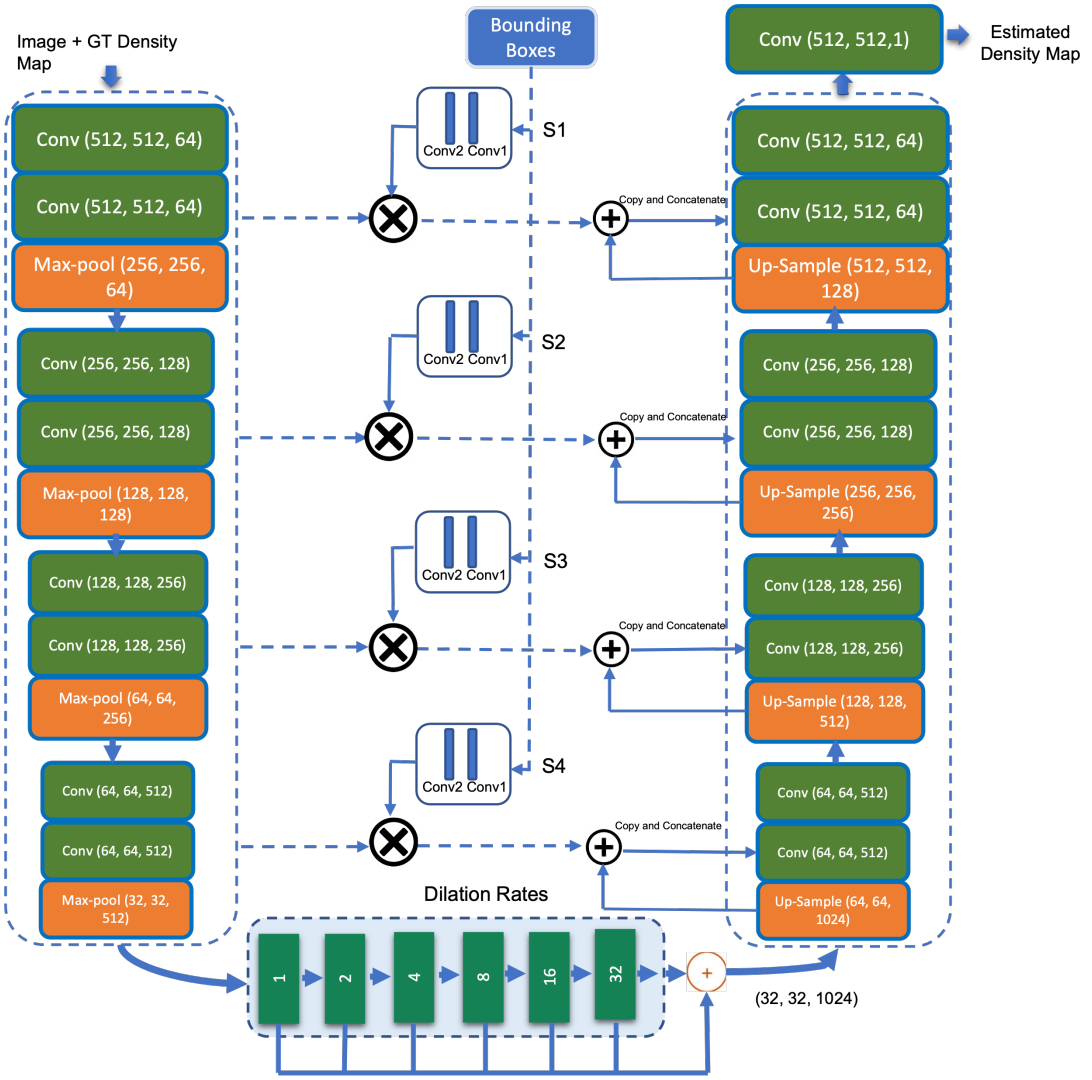
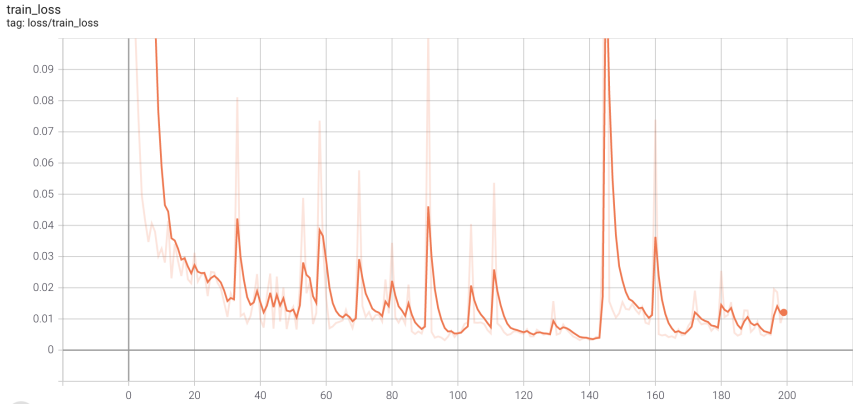


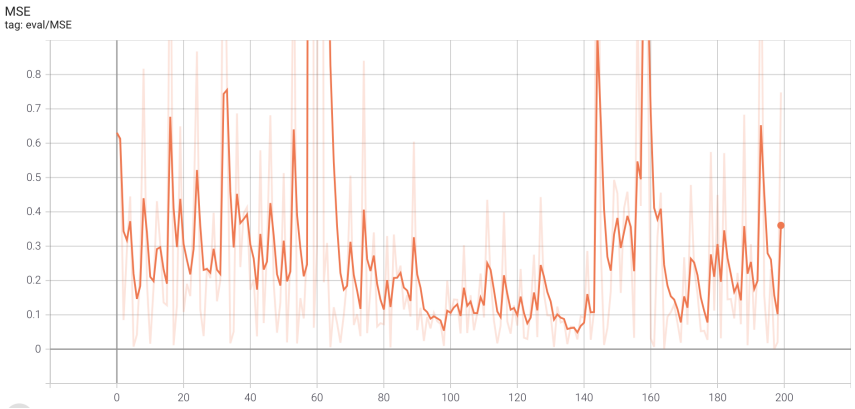
Figure 2. Block diagram of proposed dilated UNet dense regressor with attention module.

4. Results and Discussion

In this section, we present and discuss the experimental results of the proposed acne severity grading method. The public dataset used for assessment of our model [27] is split into 80% for training and validation (1165 images) and 20% for testing (292 images). The resolution of the facial images are fixed with 512 × 512 pixels. The best performance of the proposed attention guided regressor was obtained after training the network for 200 epochs using the Adam optimisation method on a batch size of 4 and the learning rate 0.0001. The data augmentation is applied to avoid over-fitting. The learning curves during the learning phase depicting the training and validation loss in terms of MSE are shown in Figure 3. Our developed model was run on an NVIDIA GTX TITAN X 12GB GPU card. The proposed algorithm is implemented based on the Tensorflow framework.



(**a**)



(**b**)

Figure 3. Learning curves of the proposed attention guided dilated UNet dense regressor. (**a**) Training loss in terms of MSE. (**b**) Validation loss in terms of MSE.

Table 1 presents the resulted confusion matrix from the proposed model architecture, where L_0, L_1, L_2, and L_3 refer to the four severity grading levels introduced as mild, moderate, severe, and very severe labels, respectively, based on the number of inflammatory eruptions (papules, pustules) and lesions. It can be noticed that images with L_0, i.e., acne count is $\leq$5, are accurately diagnosed and graded. The remaining grading levels, L_1 (6–20), L_2 (21–50), and L_3 (>50), show that the misclassification in the label prediction always occurs between two successive labels. For instance, L_1 is only falsely predicted as L_0. Similarly, L_2 is falsely predicted L_1, and L_3 is misclassified as L_2. This is a foreseen prediction due to the overlapping and similarity of appearance of acne lesions with a close severity level [27].

To elaborate the performance of our method in terms of the identification of each severity level, Table 2 exhibits the performance evaluation in terms of precision, recall (sensitivity), specificity, and accuracy. The last column shows the number of existing examples per each class label. In terms of precision, L_3 attains the best performance achieving precision of 100%, followed by L_1, L_2, and L_0, respectively. The images with severity level L_0 are identified with 100% sensitivity, proving the superiority of detection in terms of true positive detection over other severity levels. Otherwise, L_1 is predicted with the lowest sensitivity, reporting only 69%. According to the true negative rate, the severity level L_3 yields the best performance with specificity 100%, whereas the severity level L_0 produces the lowest results achieving specificity of 79%. The images with severity level L_3 (26 images) gain accuracy of 99%, whereas the images with severity level L_1 show accuracy of 84% (127 images). However, due to the imbalanced label distribution, the accuracy metric solely could be misleading in measuring the model performance [51].

Table 1. Confusion matrix of the proposed attention mechanism guided dilated UNet dense regressor.

	Predicted			
	L0	**L1**	**L2**	**L3**
L0	103	0	0	0
L1	39	88	0	0
L2	0	8	28	0
L3	0	0	4	22

Table 2. Performance evaluation of each class detection in the proposed attention mechanism guided dilated UNet dense regressor.

Class	**Pre**	**Sen**	**Spe**	**Acc**	**Support**
L0	0.73	1	0.79	0.87	103
L1	0.92	0.69	0.95	0.84	127
L2	0.88	0.78	0.98	0.96	36
L3	1	0.85	1	0.99	26

Table 3 displays a comparison of the performance of the proposed acne grading method against methods existing in the literature. In addition to precision, sensitivity, specificity, and accuracy evaluation metrics, Mean Absolute Error (MAE) and Mean Square Error (MSE) are also used. These metrics can be defined as follows:

$$MAE = \frac{1}{K}\sum_{i=1}^{K}\left|C_i - C_i^{GT}\right| \tag{4}$$

$$MSE = \sqrt{\frac{1}{K}\sum_{i=1}^{K}\left|C_i - C_i^{GT}\right|^2} \tag{5}$$

where K refers to the number of testing images, C_i^{GT} represents the ground truth count of acne lesions, and C_i is the estimated count of acne, which is resulted from calculating the total pixel values corresponding to acne lesions in the density map. The number of acne lesions in an image can be counted by integrating the densities over the image region [52]. The concept of object counting from density map was originally introduced in [52], where the integral (sum) over a region yields the number of objects in that region. This can be defined using the following formula:

$$C_i = \sum_{l=1}^{L}\sum_{w=1}^{W} Z_{l,w} \tag{6}$$

where $Z_{l,w}$ refer to the pixel values of density map; L and W are the dimensions of density map.

For comparison purposes, results reported from state-of-the-art acne grading models summarised in Table 3 are broadly classified into regression-based machine learning approaches [42,53,54], regression-based deep learning approaches [55–57], detection-based approaches [36,44], and label distribution learning approach [27]. In the regression-based machine learning approaches including SIFT-Hand Crafted Features [53], HOG-Hand Crafted Features [42], and GABOR-Hand Crafted Features [54], the features SIFT, HOG, and GABOR, respectively, are extracted manually from facial images and classified by an SVM model into four severity levels. Regression-based machine learning approaches show poor performance in all evaluation metrics. In regression-based deep learning approaches including VGGNet [55], Inceptionv3 [56], and ResNet [57], the features are extracted automatically and fed to a fully connected neural network for classifying the severity into four levels. Contrary to the regression-based machine learning approaches, regression-based

deep learning approaches achieve substantially improved performance, where ResNet [57] attains precision of 75.81%, specificity of 91.85%, sensitivity of 75.35%, and accuracy of 78.42%. MAE and MSE metrics do not apply to regression-based methods because they use a classifier to identify the levels of acne lesion severity rather than grading based on counting the acne lesions.

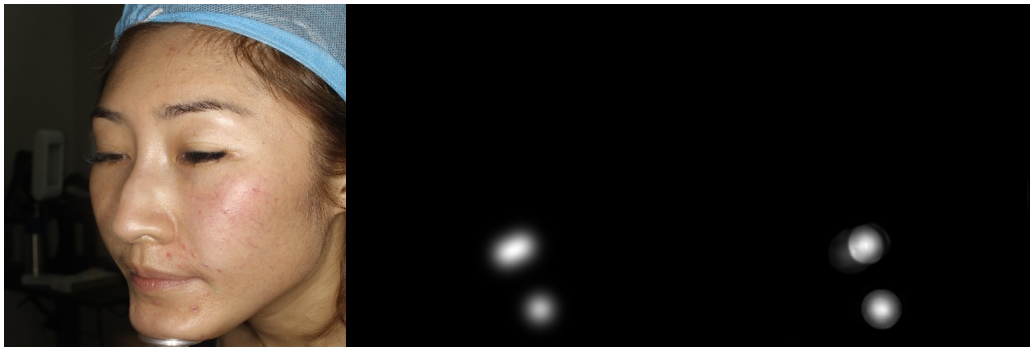
Table 3. Comparison with the existing acne lesion detection and grading methods on the same dataset. NA: Not Applicable, R-ML: Regression based Machine Learning (SVM), Regression-DL: Regression based Deep Learning, D: Detection, and LD: Label Distribution.

Method/Criteria	Method Description	MAE	MSE	Pre	Spe	Sen	Acc
SIFT-Hand Crafted Features [53]	R-ML	NA	NA	42.59	78.44	39.09	45.89
HOG-Hand Crafted Features [42]	R-ML	NA	NA	39.1	77.91	38.1	41.3
GABOR-Hand Crafted Features [54]	R-ML	NA	NA	45.35	79.89	41.78	48.22
VGGNet [55]	R-DL	NA	NA	72.65	90.6	72.71	75.17
Inceptionv3 [56]	R-DL	NA	NA	74.26	90.95	72.77	76.44
ResNet [57]	R-DL	NA	NA	75.81	91.85	75.35	78.42
YOLOv3 [44]	D	6.69	11.35	67.01	85.96	51.68	63.7
F-RCNN [36]	D	6.7	11.51	56.91	90.32	61.01	73.97
LDL [27]	LD	2.93	5.42	84.37	**93.8**	81.52	84.11
Proposed Method	**Attention Guided Regressor**	**1.76**	**3.57**	**88.25**	93	**83**	**91.5**

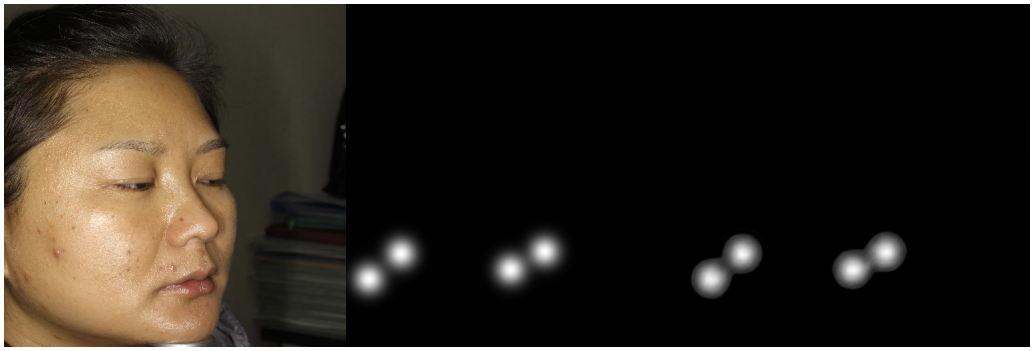
Moreover, detection-based methods including YOLOv3 [44] and F-RCNN [36] perform well in a sparse region where the acne lesions are not dense. However, they fail in the detection when the size of the acne lesions is small and overlapped. For instance, F-RCNN [36] yields MAE of 6.7, MSE of 11.51, precision of 56.91%, specificity of 90.32%, sensitivity of 61.01%, and accuracy of 73.97%. In the most recent acne severity grading method named LDL [27], the acne severity grading was realised following the scheme of label distribution learning (LDL) that considers the ambiguous information among levels of acne severity. The authors reported MAE, MSE, precision, specificity, sensitivity, and accuracy of 2.93, 5.42, 84.37%, 93.8%, 81.52%, and 84.11%, respectively. Our proposed attention guided regressor model surpasses the state-of-the-art methods in all evaluation metrics except specificity, where LDL [27] achieved better performance. The developed method shows MAE of 1.76, MSE of 3.57, precision of 88.35%, specificity of 93%, sensitivity of 83%, and accuracy of 91.5%. In terms of subjective evaluation, an example of images shown in Figure 4 illustrates the correct acne lesion detection and severity grading in the resulted attention density maps using the attention mechanism guided regression model, whereas Figure 5 depicts the misprediction of acne lesions in the resulted attention density maps.

The Figures illustrate the attention density maps through the four levels of acne severity. These results show that our model contributes to significantly estimating improved density and localisation maps. It can also be noticed the misprediction that occurred in the resulted maps is not substantial and can be tolerated. The misprediction in the density maps could be improved when training the model on a larger dataset. The presented objective and subjective performance indicate the importance of properly integrating regression and detection methods in one framework. It also reveals the significance of embedding the prior knowledge onto the model architecture while training. Hence, the proposed attention mechanism incorporated into regressor architecture would help to highlight salient features that are passed through the skip connections. This leads us to believe that the proposed model is a viable solution when dealing with diverse object distribution in specific regions. Furthermore, the dilated convolution is shown to be a good choice, which uses sparse kernels to replace implementing several layers of the pooling and convolutional filters. In summary, this paper presents an improved deep learning method based on integrating

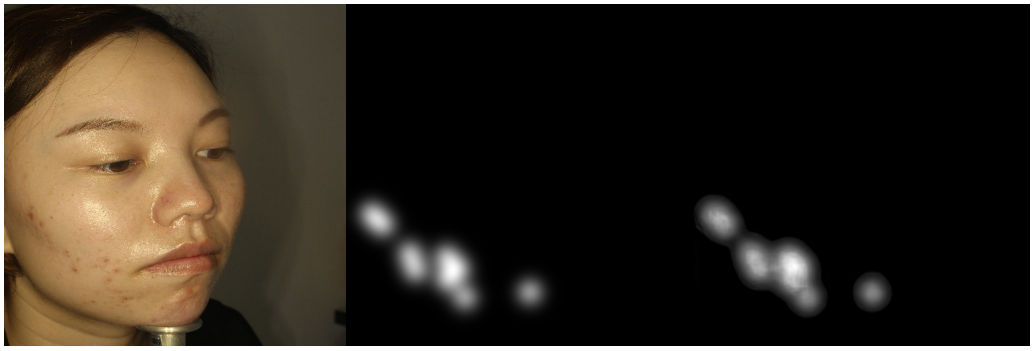
regression and detection-based approaches for acne severity grading from facial images. As a result, the acne lesions are correctly counted, and the severity is accurately graded by the proposed method.



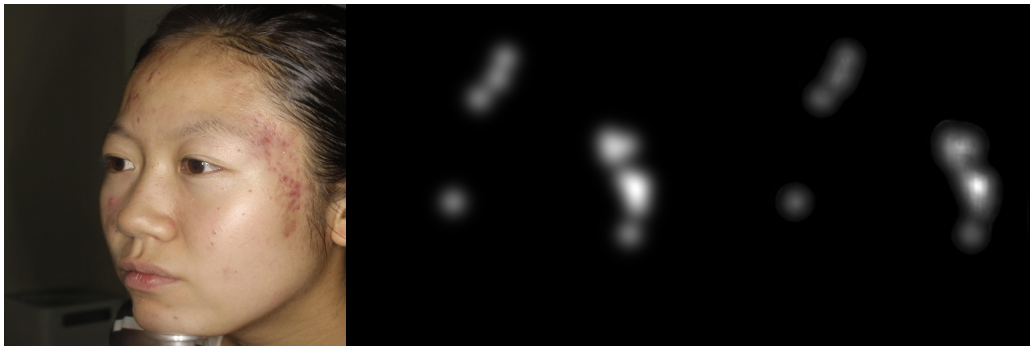
(**a**)



(**b**)

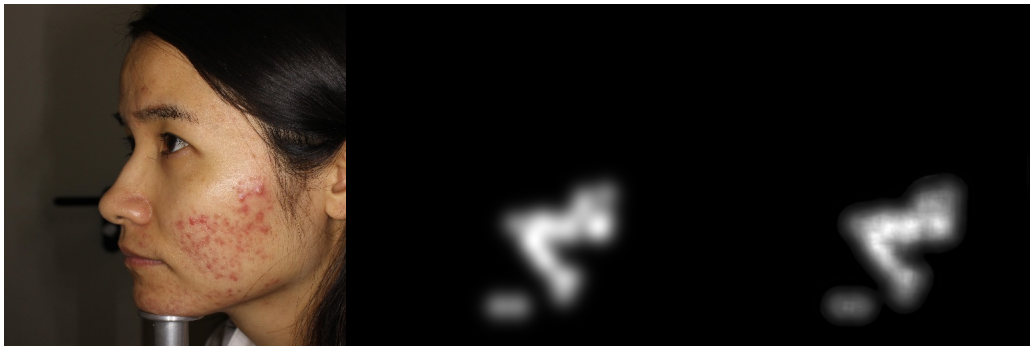


(**c**)



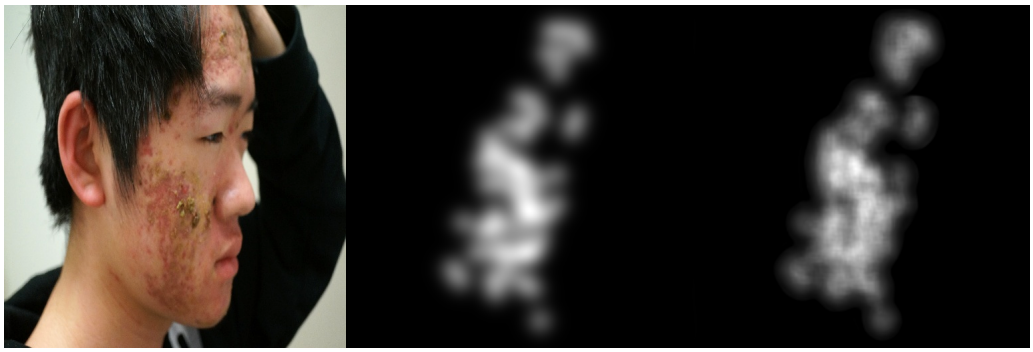
(**d**)

Figure 4. *Cont.*

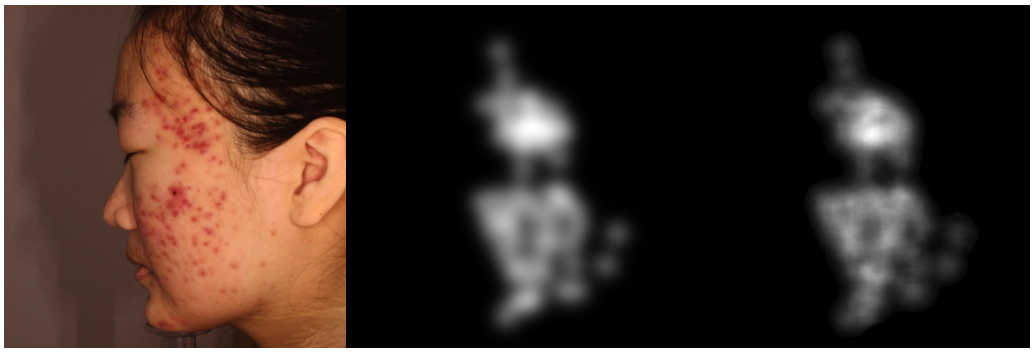


(**e**)

(**f**)

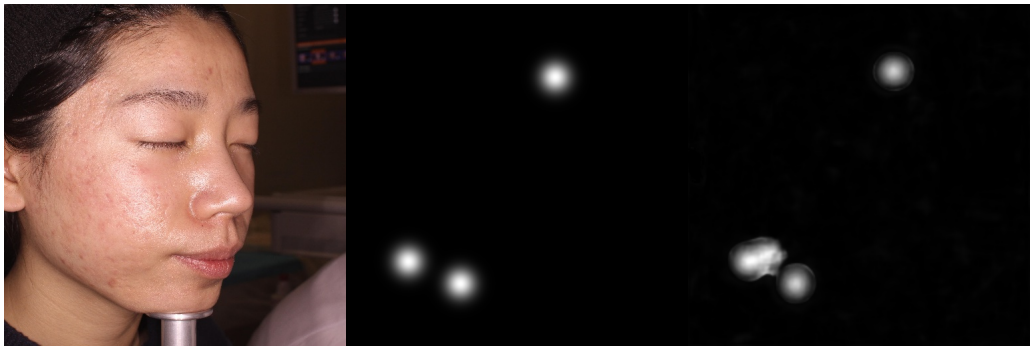


(**g**)

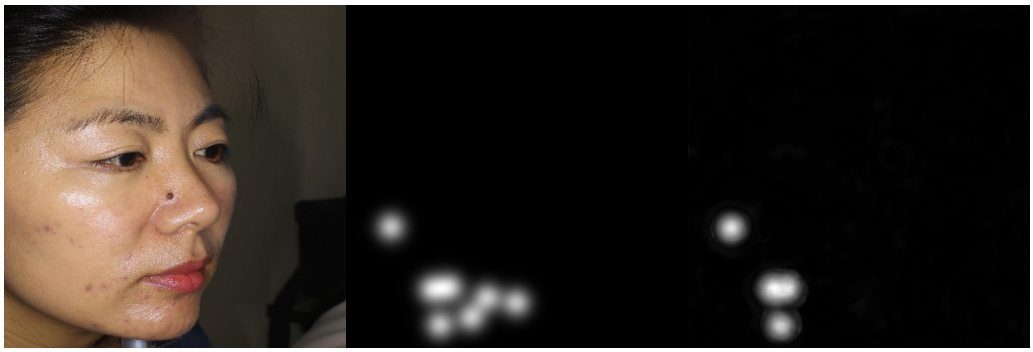


(**h**)

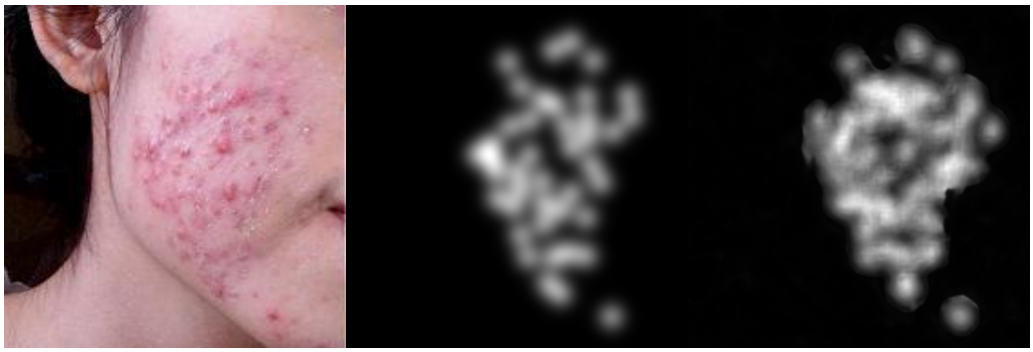
Figure 4. Image examples show correctly acne lesion detection and severity grading in the resulted attention density maps using attention mechanism guided regression model. From left to right: image, ground truth, and predicted attention density map of acne lesions. (**a**) Level 0: Example 1. (**b**) Level 0: Example 2. (**c**) Level 1: Example 1. (**d**) Level 1: Example 2. (**e**) Level 2: Example 1. (**f**) Level 2: Example 2. (**g**) Level 3: Example 1. (**h**) Level 3: Example 2.



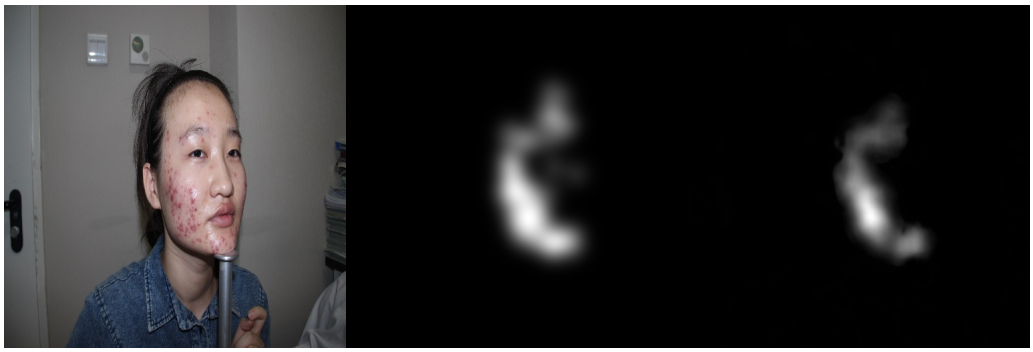
(**a**)



(**b**)



(**c**)



(**d**)

Figure 5. Image examples show misprediction of acne lesions in the resulted attention density maps. From left to right: image, ground truth, and predicted attention density map of acne lesions. (**a**) Level 0. (**b**) Level 1. (**c**) Level 2. (**d**) Level 3.

5. Conclusions

This work proposed an attention mechanism integrated with dilated UNet regressor for acne counting and severity grading from two-dimensional facial images. By incorporating the attention mechanism represented by bounding boxes generated by Faster R-CNN with density map generated by dense regressor, following a fully supervised learning scheme, the proposed method yielded better acne grading performance than the state-of-the-art methods. Integrating bounding boxes information guides the proposed method to simultaneously locate the sparse and dense acne lesion regions for the density map regression task, targeting towards improving its robustness to diverse distributions of facial acne lesions. For future work, we suggest implementing and training the developed model within a weakly-supervised framework, pushing forward to weakly supervised learning fashion due to unavailability of large amounts of annotated data within the medical domain and the fact that partial annotations are more common.

Author Contributions: Conceptualization, S.A. and B.A.-B.; methodology, S.A., B.A.-B. and W.A.-N.; software, S.A.; validation, S.A., B.A.-B. and W.A.-N.; formal analysis, S.A.; investigation, S.A. and B.A.-B.; writing—original draft preparation, S.A. and B.A.-B.; writing—review and editing, S.A., B.A.-B. and W.A.-N.; supervision, W.A.-N.; project administration, W.A.-N. All authors have read and agreed to the published version of the manuscript.

Funding: Saeed Alzahrani was funded by the Kingdom of Saudi Arabia government.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: According to the authors who collected and made the data publicly available, informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Public dataset was used in this study. These data can be found in https://github.com/xpwu95/LDL, accessed on 17 January 2022.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Aslam, I.; Fleischer, A.; Feldman, S. Emerging drugs for the treatment of acne. *Expert Opin. Emerg. Drugs* **2015**, *20*, 91–101. [CrossRef] [PubMed]
2. Adityan, B.; Kumari, R.; Thappa, D.M. Scoring systems in acne vulgaris. *Indian J. Dermatol. Venereol. Leprol.* **2009**, *75*, 323. [PubMed]
3. Barnes, L.E.; Levender, M.M.; Fleischer, A.B.; Feldman, S.R. Quality of life measures for acne patients. *Dermatol. Clin.* **2012**, *30*, 293–300. [CrossRef]
4. Goodman, G. Acne and acne scarring: The case for active and early intervention. *Aust. Fam. Phys.* **2006**, *35*, 503–504. [CrossRef]
5. Witkowski, J.A.; Parish, L.C. The assessment of acne: An evaluation of grading and lesion counting in the measurement of acne. *Clin. Dermatol.* **2004**, *22*, 394–397. [CrossRef]
6. Burke, B.M.; Cunliffe, W. The assessment of acne vulgaris—The Leeds technique. *Br. J. Dermatol.* **1984**, *111*, 83–92. [CrossRef] [PubMed]
7. Hayashi, N.; Akamatsu, H.; Kawashima, M.; Group, A.S. Establishment of grading criteria for acne severity. *J. Dermatol.* **2008**, *35*, 255–260.
8. Dreno, B.; Poli, F. Epidemiology of acne. *Dermatology* **2003**, *206*, 7. [CrossRef]
9. Goulden, V.; Stables, G.; Cunliffe, W. Prevalence of facial acne in adults. *J. Am. Acad. Dermatol.* **1999**, *41*, 577–580.
10. Williams, H.C.; Dellavalle, R.P.; Garner, S. Acne vulgaris. *Lancet* **2012**, *379*, 361–372. [CrossRef]
11. Alzahrani, S.; Al-Bander, B.; Al-Nuaimy, W. A Comprehensive Evaluation and Benchmarking of Convolutional Neural Networks for Melanoma Diagnosis. *Cancers* **2021**, *13*, 4494. [CrossRef]
12. Suzuki, K. Overview of deep learning in medical imaging. *Radiol. Phys. Technol.* **2017**, *10*, 257–273. [CrossRef] [PubMed]
13. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef] [PubMed]
14. Amini, M.; Vasefi, F.; Valdebran, M.; Huang, K.; Zhang, H.; Kemp, W.; MacKinnon, N. Automated facial acne assessment from smartphone images. In *Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XVI*; International Society for Optics and Photonics: Bellingham, WA, USA, 2018; Volume 10497, p. 104970N.
15. Junayed, M.S.; Jeny, A.A.; Atik, S.T.; Neehal, N.; Karim, A.; Azam, S.; Shanmugam, B. AcneNet-A Deep CNN Based Classification Approach for Acne Classes. In Proceedings of the 2019 12th International Conference on Information & Communication Technology and System (ICTS), Surabaya, Indonesia, 18 July 2019; pp. 203–208.

16. Abas, F.S.; Kaffenberger, B.; Bikowski, J.; Gurcan, M.N. Acne image analysis: Lesion localization and classification. In *Medical Imaging 2016: Computer-Aided Diagnosis*; International Society for Optics and Photonics: Bellingham, WA, USA, 2016; Volume 9785, p. 97850B.
17. Shen, X.; Zhang, J.; Yan, C.; Zhou, H. An automatic diagnosis method of facial acne vulgaris based on convolutional neural network. *Sci. Rep.* **2018**, *8*, 5839. [CrossRef]
18. Malik, A.; Humayun, J.; Kamel, N.; Yap, F.B. Novel techniques for enhancement and segmentation of acne vulgaris lesions. *Ski. Res. Technol.* **2014**, *20*, 322–331. [CrossRef]
19. Chantharaphaichi, T.; Uyyanonvara, B.; Sinthanayothin, C.; Nishihara, A. Automatic acne detection for medical treatment. In Proceedings of the 2015 6th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES), Hua Hin, Thailand, 22–24 March 2015; pp. 1–6.
20. Alamdari, N.; Tavakolian, K.; Alhashim, M.; Fazel-Rezai, R. Detection and classification of acne lesions in acne patients: A mobile application. In Proceedings of the 2016 IEEE International Conference on Electro Information Technology (EIT), Grand Forks, ND, USA, 19–21 May 2016; pp. 0739–0743.
21. Liu, Z.; Zerubia, J. Towards automatic acne detection using a MRF model with chromophore descriptors. In Proceedings of the 21st European Signal Processing Conference (EUSIPCO 2013), Marrakech, Morocco, 9–13 September 2013; pp. 1–5.
22. Maroni, G.; Ermidoro, M.; Previdi, F.; Bigini, G. Automated detection, extraction and counting of acne lesions for automatic evaluation and tracking of acne severity. In Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, USA, 27 November–1 December 2017; pp. 1–6.
23. Min, K.; Lee, G.H.; Lee, S.W. ACNet: Mask-Aware Attention with Dynamic Context Enhancement for Robust Acne Detection. *arXiv* **2021**, arXiv:2105.14891.
24. Melina, A.; Dinh, N.N.; Tafuri, B.; Schipani, G.; Nisticò, S.; Cosentino, C.; Amato, F.; Thiboutot, D.; Cherubini, A. Artificial Intelligence for the Objective Evaluation of Acne Investigator Global Assessment. *J. Drugs Dermatol.* **2018**, *17*, 1006–1009. [PubMed]
25. Seité, S.; Khammari, A.; Benzaquen, M.; Moyal, D.; Dréno, B. Development and accuracy of an artificial intelligence algorithm for acne grading from smartphone photographs. *Exp. Dermatol.* **2019**, *28*, 1252–1257. [CrossRef] [PubMed]
26. Zhao, T.; Zhang, H.; Spoelstra, J. A Computer Vision Application for Assessing Facial Acne Severity from Selfie Images. *arXiv* **2019**, arXiv:1907.07901.
27. Wu, X.; Wen, N.; Liang, J.; Lai, Y.K.; She, D.; Cheng, M.M.; Yang, J. Joint Acne Image Grading and Counting via Label Distribution Learning. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 10642–10651.
28. Lim, Z.V.; Akram, F.; Ngo, C.P.; Winarto, A.A.; Lee, W.Q.; Liang, K.; Oon, H.H.; Thng, S.T.G.; Lee, H.K. Automated grading of acne vulgaris by deep learning with convolutional neural networks. *Ski. Res. Technol.* **2020**, *26*, 187–192. [CrossRef]
29. Ramli, R.; Malik, A.S.; Hani, A.F.M.; Jamil, A. Acne analysis, grading and computational assessment methods: An overview. *Ski. Res. Technol.* **2012**, *18*, 1–14. [CrossRef] [PubMed]
30. MedicineWise. Investigator's Global Assessment (IGA) of Acne Severity. Available online: https://www.nps.org.au/radar/articles/investigators-global-assessment-iga-of-acne-severity-additional-content-adapalene-with-benzoyl-peroxide-epiduo-for-severe-acne-vulgaris (accessed on 18 January 2022).
31. Dreno, B.; Poli, F.; Pawin, H.; Beylot, C.; Faure, M.; Chivot, M.; Auffret, N.; Moyse, D.; Ballanger, F.; Revuz, J. Development and evaluation of a Global Acne Severity scale (GEA scale) suitable for France and Europe. *J. Eur. Acad. Dermatol. Venereol.* **2011**, *25*, 43–48. [CrossRef] [PubMed]
32. Zhang, Y.; Zhou, D.; Chen, S.; Gao, S.; Ma, Y. Single-image crowd counting via multi-column convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 589–597.
33. Li, Y.; Zhang, X.; Chen, D. CSRNET: Dilated convolutional neural networks for understanding the highly congested scenes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1091–1100.
34. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
35. Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv* **2015**, arXiv:1511.07122.
36. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [CrossRef]
37. Xie, W.; Noble, J.A.; Zisserman, A. Microscopy cell counting and detection with fully convolutional regression networks. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **2018**, *6*, 283–292. [CrossRef]
38. Yao, L.; Liu, T.; Qin, J.; Lu, N.; Zhou, C. Tree counting with high spatial-resolution satellite imagery based on deep neural networks. *Ecol. Indic.* **2021**, *125*, 107591. [CrossRef]
39. Norouzzadeh, M.S.; Nguyen, A.; Kosmala, M.; Swanson, A.; Palmer, M.S.; Packer, C.; Clune, J. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E5716–E5725. [CrossRef]

40. Song, H.; Liang, H.; Li, H.; Dai, Z.; Yun, X. Vision-based vehicle detection and counting system using deep learning in highway scenes. *Eur. Transp. Res. Rev.* **2019**, *11*, 1–16. [CrossRef]
41. Babu Sam, D.; Surya, S.; Venkatesh Babu, R. Switching convolutional neural network for crowd counting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017, pp. 5744–5752.
42. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.
43. Viola, P.; Jones, M.J. Robust real-time face detection. *Int. J. Comput. Vis.* **2004**, *57*, 137–154. [CrossRef]
44. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
45. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
46. Boominathan, L.; Kruthiventi, S.S.; Babu, R.V. Crowdnet: A deep convolutional network for dense crowd counting. In Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands, 15–19 October 2016; pp. 640–644.
47. Wu, X.; Zheng, Y.; Ye, H.; Hu, W.; Ma, T.; Yang, J.; He, L. Counting crowds with varying densities via adaptive scenario discovery framework. *Neurocomputing* **2020**, *397*, 127–138. [CrossRef]
48. Ibrahim, M.S.; Vahdat, A.; Ranjbar, M.; Macready, W.G. Semi-supervised semantic image segmentation with self-correcting networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 12715–12725.
49. El Jurdi, R.; Petitjean, C.; Honeine, P.; Abdallah, F. Bb-unet: U-net with bounding box prior. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 1189–1198.
50. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention u-net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
51. Akosa, J. Predictive accuracy: A misleading performance measure for highly imbalanced data. In Proceedings of the SAS Global Forum, Orlando, FL, USA, 2–5 April 2017; Volume 12.
52. Lempitsky, V.; Zisserman, A. Learning to count objects in images. *Adv. Neural Inf. Process. Syst.* **2010**, *23*, 1324–1332.
53. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
54. Mehrotra, R.; Namuduri, K.R.; Ranganathan, N. Gabor filter-based edge detection. *Pattern Recognit.* **1992**, *25*, 1479–1494. [CrossRef]
55. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
56. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
57. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.