Improving Acne Detection and Severity Grading Using Deep Learning on Smartphone Images

Authors: Anayna Nidhi Singh, Narasimha Karthik G

---

1. Project Information

Primary Paper Implemented in the Project:
Huynh, Q.T., Nguyen, P.H., et al. "Automatic Acne Object Detection and Acne Severity Grading Using Smartphone Images and Artificial Intelligence." *Diagnostics*, 2022.

---

2. Problem Statement

Acne is a highly prevalent skin condition affecting around 9.38% of the global population. While often associated with adolescence, acne can persist into adulthood, leading to physical and psychological impacts. The condition varies in severity and presentation, including inflammatory lesions like papules, pustules, and nodules, as well as non-inflammatory lesions like blackheads and whiteheads. If untreated, severe cases can lead to permanent scarring and long-term psychological effects, such as anxiety and low self-esteem.

Current acne diagnosis and severity grading typically require a dermatologist's expertise, which limits accessibility, especially in remote or underserved areas. An automated acne detection and grading system could help democratize access to initial assessments, enabling users to monitor their acne and seek timely professional care if necessary. Given the ubiquity of smartphones, a system that uses smartphone-captured images to detect acne lesions and grade their severity could significantly improve patient engagement in managing acne.

---

3. Proposed Solution and Implementation Details

Algorithm Overview

The solution comprises a **two-stage system**:

1. **Acne Object Detection Model**
   - **Model Architecture:** We implemented a Faster R-CNN architecture with a ResNet50 backbone, a widely used approach for object detection tasks in computer vision. Faster R-CNN utilizes Region Proposal Networks (RPN) to generate region proposals and then classifies these proposals to identify objects within the image. The ResNet50 backbone provides a robust feature extraction mechanism, enabling the model to capture spatial and contextual information effectively. A pre-trained Faster R-CNN model with a ResNet50

backbone and Feature Pyramid Network (FPN) was fine-tuned to adapt to the acne classification task. The number of output classes was modified to include the four acne types and a background class. Specifically, the box predictor layer of the Faster R-CNN model was replaced with a customized layer to output predictions for the desired number of classes.

- **Objective:** The model aims to detect and classify four distinct acne types (blackheads/whiteheads, papules/pustules, nodules/cysts, and acne scars). Additionally, the model identifies the background class to separate areas without acne. The bounding box annotations allow precise localization of acne types in the input images.
- **Training and Evaluation:** The model was trained on the annotated ACNE04 dataset, which includes bounding box annotations for each acne type. The performance of the model is measured using mean Average Precision (mAP), a standard metric for object detection tasks. The model was trained using the Stochastic Gradient Descent (SGD) optimizer with a learning rate of 0.005, momentum of 0.9, and a weight decay of 0.0005. A 10-epoch training schedule was adopted with batch processing and loss monitoring.

2. **Acne Severity Grading Model**
   - **Model Architecture:** We implemented a LightGBM-based machine learning model to automatically grade acne severity based on a set of features derived from a Faster R-CNN object detection model. This model takes as input the output counts and spatial information of detected acne lesions from the object detection model and uses these features to assess acne severity. The extract_features method calculates lesion-based features from the detection model output, which include:
     - Lesion Counts and Confidence Scores: Average, max, and min confidence scores of detected lesions.
     - Lesion Sizes: Average, max, and min sizes of detected lesions.
     - Face Segmentation Ratios: Ratios of lesions on the upper and lower halves of the face.
     - Lesion Density: Based on lesion count relative to face area.

     The LightGBM model is trained on extracted features and lesion counts, which are mapped to the IGA grades

   - **Objective:** This model grades the overall acne severity according to the Investigator's Global Assessment (IGA) scale, a five-level severity grading scale commonly used in dermatology. These levels are Clear (0), Almost Clear (1), Mild (2), Moderate (3), and Severe (4).
   - **Training and Evaluation:** The model is trained on a dataset annotated for severity, which is labeled with counts and confidence metrics for each detected acne type. Evaluation is conducted using accuracy and Area Under the ROC Curve (AUC) as primary performance metrics. The classification report includes precision, recall, and F1-scores for each severity class to provide a more detailed analysis of model performance.

This two-stage architecture allows the acne severity grading model to leverage the specific counts of each acne type, closely mimicking the dermatologist's diagnostic process.

---

4. Dataset

The ACNE04 dataset is a high-resolution, annotated resource designed to advance research in automated acne detection and severity grading. It features diverse images that capture a wide range of skin tones, lighting conditions, and facial orientations, making it an ideal dataset for training machine learning models that can generalize across real-world scenarios. The dataset includes over 1,572 images and 41,000 labeled acne lesions categorized into four distinct types: blackheads/whiteheads, papules/pustules, nodules/cysts, and acne scars. These annotations enable the model to learn the nuanced visual characteristics of each lesion type, from the smaller and less inflamed blackheads/whiteheads to the larger, more severe nodules/cysts, and the subtle, harder-to-detect acne scars.

Severity grading in the dataset is based on the Investigator's Global Assessment (IGA) scale, ranging from 0 (clear) to 4 (severe), providing valuable data for models to predict the overall severity of acne. This dual labeling of lesion types and severity levels supports a two-stage approach, where object detection results feed into a severity grading model. For training and evaluation, the dataset is split into 70% training and 30% testing, following best practices to ensure robust model development and reliable performance metrics such as mean Average Precision (mAP) for detection and F1 scores for severity grading.

The dataset reflects real-world acne presentation, with acne scars comprising the majority of lesions (55.46%), followed by blackheads/whiteheads (37.47%), papules/pustules (6.4%), and nodules/cysts (0.67%). Severity grades are similarly imbalanced, with Grade 1 (almost clear) being the most common at 56.18% and Grade 4 (severe) being the least frequent at 2.16%. This class imbalance necessitates techniques such as data augmentation and class weighting to mitigate model bias. Overall, the ACNE04 dataset's detailed annotations and thoughtful design make it a vital tool for developing accurate, generalizable machine learning models for acne analysis.

---

5. Results and Discussion

1. **Acne Object Detection (Faster R-CNN)**
   The object detection model, based on Faster R-CNN with a ResNet-50 backbone, achieved a Mean Average Precision (mAP) of 24.81%. The model was evaluated on four classes of acne lesions: Blackheads/Whiteheads, Papules/Pustules, Nodules/Cysts, and Acne Scars. Below are the class-wise detection metrics:

Class	Precision	Recall	F1-Score	TP	FP	FN
Blackheads/Whiteheads	0.000000	0.0	0.000000	0	0	863
Papules/Pustules	0.694797	1.0	0.819918	17174	7544	0
Nodules/Cysts	0.482352	1.0	0.650793	7120	7641	0
Acne Scars	0.670172	1.0	0.802519	7965	3920	0

Table 1: *Table displaying the accuracy metrics of the acne object detection model for each of the four acne classes.*

- The detection summary on the first sample image taken from the ACNE04 dataset revealed two detected acne lesions:

Diagram 1: Image showing the results of evaluating R-CNN with Resnet-50 backbone for Object (Acne) Detection on sample image from ACNE04 dataset. Output generated from Python code.

- Total detections after NMS: 2

Type	Confidence
1	0.8467
1	0.3067

Table 2: Table displaying the acne object detection summary results from the above image as generated by Python code. It detected 2 acne belonging to Type-1(Papules/Pustules) with some confidence values.

- The sample image demonstrates the model's capability to localize acne lesions accurately with bounding boxes. However, the confidence scores suggest that the model is less confident about some detections (e.g., 30.67%), likely due to overlaps or ambiguity in lesion appearance.
- Now testing the model on sample image 2:

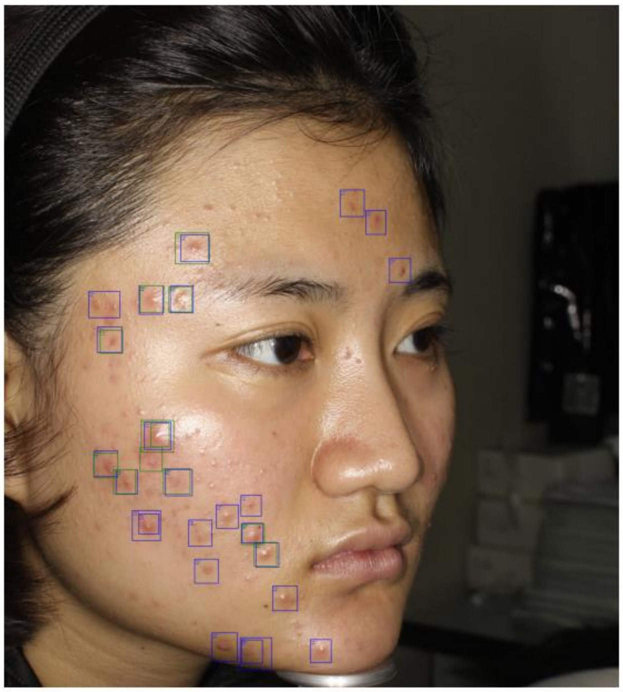

Diagram 2: Results of evaluating R-CNN with Resnet-50 backbone for Object (Acne) Detection on another sample image from ACNE04 dataset. Output generated from Python code. In the above image, the total number of acne that was generated was 37 with 25 acne classified as type 1 and 12 acne classified as type 2 with different confidence values.

- Total detections after NMS: 37

Model Performance

- The model showed strong recall for most acne types, particularly for Papules/Pustules, Nodules/Cysts, and Acne Scars, achieving 100% recall. This indicates that the model successfully identifies all true positives for these classes.
- However, precision is notably lower, especially for classes with high numbers of false positives (e.g., Papules/Pustules and Nodules/Cysts). This suggests that the model frequently misclassified other regions as acne lesions.
- The model struggled significantly with detecting Blackheads/Whiteheads, as evidenced by zero precision and recall scores for this class. This might be due to the subtle

appearance of such lesions in images, leading to poor feature representation in the detection network.
- The Mean Average Precision (mAP) of 24.81% reflects the challenges faced in detecting smaller and less prominent acne types, as well as differentiating between closely related classes.
- The F1-scores for Papules/Pustules and Acne Scars indicate a better balance between precision and recall, suggesting that these types are more distinguishable for the model.

2. **Acne Severity Grading (LightGBM)**

   The proposed acne severity grading model, based on the LightGBM classifier, demonstrates a strong performance across various metrics. A total of 1,457 samples were used to train and evaluate the model, with the following class distribution:
   - Grade 0: 513 samples (35.2%)
   - Grade 1: 607 samples (41.7%)
   - Grade 2: 128 samples (8.8%)
   - Grade 3: 80 samples (5.5%)
   - Grade 4: 129 samples (8.9%)

- **Performance:**
  - **Overall Accuracy:** 91%
  - **Classification report:**

Grade	Precision	Recall	F1-Score
0	0.99	0.98	0.98
1	0.93	0.95	0.94
2	0.69	0.69	0.69
3	0.67	0.51	0.58
4	0.82	0.89	0.86
Macro Average	-	-	0.81
Weighted Average	-	-	0.91

Table 3: Table showing classification metrics for the acne severity grading model using LightGBM

The confusion matrix highlights that the model performs well in distinguishing Grade 0 and Grade 1, which constitute the majority of the dataset. The performance on Grade 3 is relatively lower, likely due to class imbalance and overlap in features with neighboring grades.

  - **ROC Curve Analysis:** High AUC values across all grades confirm robustness.

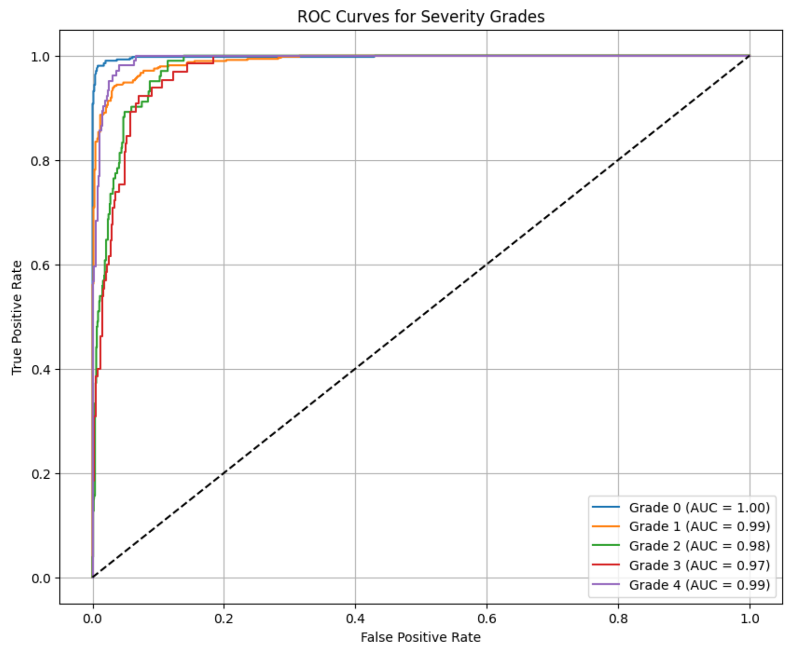

Diagram 3: ROC Curves for Severity Grading Model

Grade	AUC	Performance	Interpretation
0	1.00	Perfect	Model perfectly distinguishes Grade 0
1	0.99	Excellent	Nearly perfect discrimination for Grade 1
2	0.98	Excellent	Very high accuracy in identifying Grade 2
3	0.97	Excellent	Strong performance in detecting Grade 3
4	0.99	Excellent	Nearly perfect discrimination for Grade 4

Table 4: Table showing ROC curves for different grades and its interpretation

- **Confusion Matrix**

  The confusion matrix highlights that the model performs well in distinguishing Grade 0 and Grade 1, which constitute the majority of the dataset. The performance on Grade 3 is relatively lower, likely due to class imbalance and overlap in features with neighboring grades.

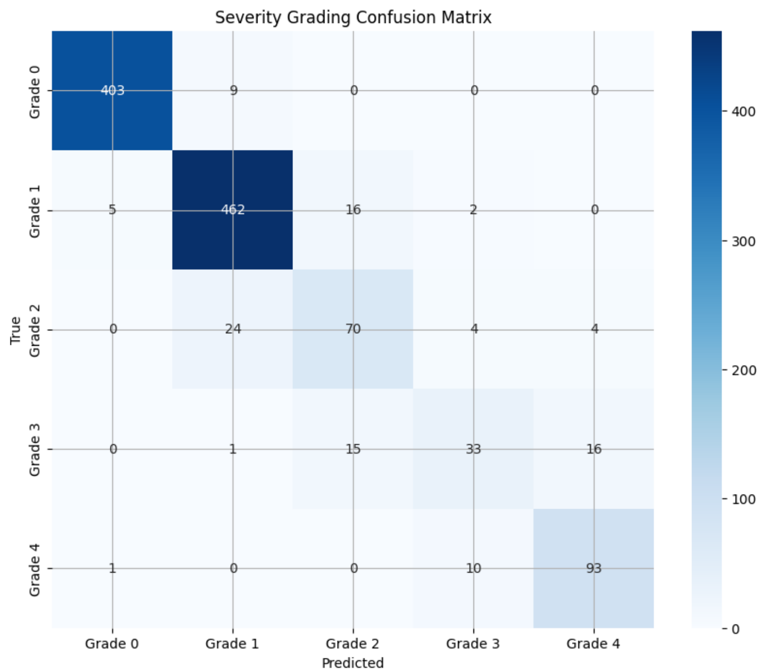

Diagram 4: Diagram showing confusion matrix for the acne severity grading model

- **Feature Importance:**

The most significant features contributing to the model's performance include:

- Total Lesions: 8129.0
- Maximum Confidence: 4710.2
- Minimum Confidence: 3632.2
- Lesion Density: 801.6
- Average Confidence: 525.1

These results indicate that lesion count, confidence scores of detected lesions, and lesion density significantly influence the model's predictions.

The acne severity grading model demonstrates excellent performance for Grades 0 and 1, with challenges primarily observed in Grades 2 and 3 due to limited samples and overlapping features. The imbalanced dataset likely contributed to reduced recall for Grade 3, as evident in the confusion matrix and F1-scores. Enhancing the dataset with more examples for the minority classes (Grades 2 and 3) could improve the model's ability to generalize across all severity levels.

The ROC analysis confirms that the model maintains a strong discrimination capacity, even for the minority classes, as reflected in high AUC scores. Further refinement, such as class weighting or augmentation techniques, could mitigate the observed discrepancies.

In real-world scenarios, this model can aid dermatologists by providing accurate severity grading based on lesion counts and their confidence levels, offering a quick, objective assessment. Future work can focus on integrating additional contextual features like patient demographics and clinical history to further enhance performance.

- Testing the model on a sample image from the ACNE04 dataset:

Diagram 5: Image showing the results of evaluating acne severity grading model on sample image from ACNE04 dataset. Output generated from Python code.

- Total detections: 16
- Confidence levels for predicted acne type of all the 16 predicted acne:

Type	Confidence
1	0.9465
1	0.9310
1	0.8954
1	0.8874
1	0.8805
1	0.8729
1	0.8648
1	0.8578
1	0.8543
1	0.8363
1	0.7997
1	0.7796
1	0.7535
1	0.7140
1	0.6816
1	0.5961

Table 5: Confidence levels for acne type of all the 16 predicted acne for image in Diagram 5.

- Predicted Severity Grade: 1
- Grade Probabilities:

Grade	Probabilities
0	0.1344
1	0.7351
2	0.0272
3	0.0179
4	0.0854

Table 6: Table showing the probabilities of being classified as a severity grade based on confidence levels of all 16 acne types predicted earlier.

6. Challenges and Future Work

Challenges:

1. **Data Imbalance:** Severe grades underrepresented.
   - Solution: Augmentation and weighted classes.
2. **Bounding Box Accuracy:** Small or faint lesions (e.g., blackheads) often missed.
   - Solution: Finer annotations, focused augmentation, and contrast enhancement.
3. **Performance Optimization:** Faster R-CNN's computational cost.
   - Solution: Hyperparameter tuning and leveraging GPU acceleration.

Future Improvements:

1. Pretraining on similar datasets for improved feature learning.
2. Enhancing dataset diversity (e.g., more Grade 3/4 samples).
3. Refining Non-Maximum Suppression (NMS) thresholds to reduce false positives.

---

7. Conclusion

This project demonstrates the potential of AI-based systems for **acne detection and severity grading** using smartphone images. While the models show promising results, particularly in distinguishing mild and severe acne grades, challenges such as dataset imbalance and detection precision for subtle lesions remain. By addressing these issues, this approach could be scaled for real-world applications, enhancing accessibility to dermatological care.

---

8. References

[1]Q. T. Huynh et al., “Automatic Acne Object Detection and Acne Severity Grading Using Smartphone Images and Artificial Intelligence,” Diagnostics, vol. 12, no. 8, p. 1879, Aug. 2022, doi: https://doi.org/10.3390/diagnostics12081879.

[2]S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: https://doi.org/10.1109/tpami.2016.2577031.

[3]G. Ke et al., “LightGBM: A Highly Efficient Gradient Boosting Decision Tree.” NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, Dec 2017, doi: https://dl.acm.org/doi/pdf/10.5555/3294996.3295074.
