

Penyesuaian YOLOv8n untuk Deteksi Presisi Komedo Hitam, Jerawat Jamur, dan Jerawat Nodul

Tugas Akhir

diajukan untuk memenuhi salah satu syarat

memperoleh gelar sarjana

pada Program Studi Informatika

Fakultas Informatika Universitas Telkom

1301213527

Raditha Ariyani



Program Studi Sarjana Informatika

Fakultas Informatika

Universitas Telkom

Bandung

2025

LEMBAR PENGESAHAN

Penyesuaian YOLOv8n untuk Deteksi Presisi Komedo Hitam, Jerawat Jamur, dan Jerawat Nodul

Fine-Tuning YOLOv8n for Precision Detection of Blackhead, Fungal Acne, and Acne Nodule

NIM : 1301213527

Raditha Ariyani

Tugas akhir ini telah diterima dan disahkan untuk memenuhi sebagian syarat memperoleh gelar pada Program Studi Sarjana Informatika

Fakultas Informatika

Universitas Telkom

Bandung, 23 Juli 2025

Menyetujui

Pembimbing I,

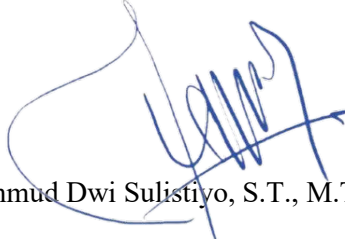


Dr. Putu Harry Gunawan, S.Si., M.Si., M.Sc

NIP: 16860043

Ketua Program Studi

Sarjana Informatika,



Mahmud Dwi Sulistiyo, S.T., M.T., Ph.D.

NIP: 13880017

LEMBAR PERNYATAAN

Dengan ini saya, Raditha Ariyani, menyatakan sesungguhnya bahwa Tugas Akhir saya dengan judul Penyesuaian YOLOv8n untuk Deteksi Presisi Komedo Hitam, Jerawat Jamur, dan Jerawat Nodul beserta dengan seluruh isinya adalah merupakan hasil karya sendiri, dan saya tidak melakukan penjiplakan yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan, serta produk dari tugas akhir bukan merupakan produk dari *Generative AI*. Saya siap menanggung resiko/sanksi yang diberikan jika di kemudian hari ditemukan pelanggaran terhadap etika keilmuan dalam Laporan TA atau jika ada klaim dari pihak lain terhadap keaslian karya,

Bandung, 23 Juli 2025

Yang Menyatakan



Raditha Ariyani

1301213527

Fine-Tuning YOLOv8n for Precision Detection of Blackhead, Fungal Acne, and Acne Nodule

1st Raditha Ariyani
School of Computing
Telkom University
Bandung, Indonesia
astxyz@student.telkomuniversity.ac.id

2nd Putu Harry Gunawan
School of Computing
Telkom University
Bandung, Indonesia
phgunawan@telkomuniversity.ac.id

3rd Narita Aquarini
École Doctorale Science Economics
Université de Poitiers Intervenant Finance
La Rochelle, France
aquarinin@excelia-group.com

4th Ida Ayu Manik Partha Sutema
Faculty of Sciences
Bali International University
Bali, Indonesia
manikparthasutema@unbi.ac.id

5th Keri Lestari
Faculty of Pharmacy
Padjadjaran University
Bandung, Indonesia
lestarikd@unpad.ac.id

Abstract—Accurate differentiation among acne subtypes, such as blackhead, fungal acne, and acne nodule, is crucial for effective dermatological treatment. However, it presents a significant challenge for automated visual analysis due to the subtle and overlapping characteristics of lesions. This research aimed to enhance the precision of automated acne detection by meticulously fine-tuning the YOLOv8n object detection model specifically for the three challenging types of acne. A curated dataset comprising 2008 images derived from public sources (Skin-90, DermNet) was utilized. The dataset was manually annotated for blackhead, fungal acne, and nodule by author under the supervision of a domain expert to ensure clinical accuracy. An iterative fine-tuning process systematically explored key hyperparameters, including image size, multi-scale training, and diverse data augmentation strategies, with performance evaluated using Precision, Recall, and mean Average Precision at an IoU of 0.5 (mAP50). The optimal configuration (Experiment 6: $\text{imgsz}=640$, multi-scale training, mixed geometric augmentations with $\text{scale}=0.5$, $\text{degrees}=10.0$, $\text{translate}=0.15$, and cls_loss weight of 0.75) achieved the highest overall precision (P_{all}) of 0.693 and a robust overall mAP50 of 0.589. Notable per-class mAP50 scores were 0.442 for blackhead, 0.750 for fungal acne, and 0.576 for acne nodule. The qualitative analysis further corroborated the model's capability for precise lesion localization, although challenges in specificity for blackheads (evidenced by some false positives) were noted. This study successfully demonstrates that a systematically fine-tuned YOLOv8n model can achieve high precision in differentiating challenging acne subtypes, offering a promising approach for developing more reliable computer-aided diagnostic tools in dermatology.

Keywords—acne detection, yolov8n, fine-tuning, precision detection

I. INTRODUCTION

Personal appearance is important to consider, as it significantly influences an individual's character. An unsatisfactory appearance, such as that caused by acne, can lower self-confidence, cause anxiety, and even depression. Acne (acne vulgaris) is a prevalent skin disease, especially in adolescents, that results from inflammation of the pilosebaceous follicles (hair follicles) and sebaceous glands (oil-producing glands) [1], [2]. Acne is divided into two types: inflammatory lesions, such as papules, pustules, nodules, and cysts; and non-inflammatory lesions, such as blackheads (both open and closed comedones) [3]. Acne affects approximately 85% of the human population aged 11-30 years and commonly

appears on the face, back, chest, and shoulder areas, which are areas with a high concentration of sebaceous glands [4]–[6]. Although acne is often perceived primarily as a cosmetic problem, it has a considerable impact on physical appearance in adolescents, affecting their overall quality of life, both in social relationships and emotional well-being, making it a much more complex issue than just a skin disorder [7], [8].

For effective treatment, dermatologists must accurately identify and understand the specific type of acne a patient is experiencing. However, due to the limited time often available for consultation, manual skin examination can be challenging and time-consuming [4]. One promising approach to overcome this problem is to utilize artificial intelligence (AI) technology, specifically deep learning-based object detection models.

Deep learning, a subfield of machine learning, has been widely adopted for analyzing skin images, including acne, due to its capability to detect subtle patterns and features in facial or skin images that may be difficult for the human eye to discern [1], [2], [4], [5]. Applying deep learning in acne detection facilitates faster and potentially more accurate diagnoses. With the ability to automatically process images and recognize various types of acne, this technology can assist dermatologists in choosing appropriate treatments and providing more efficient diagnostic solutions.

Several studies have explored the use of object detection models for acne detection. Early work utilizing YOLOv5 showed promising potential, as evidenced by research reporting good accuracy [9]. However, other studies that employed YOLOv5 for multiclass detection, such as Sangha and Rizvi, reported mAP@0.5 scores of around 26.50% [1]. These figures indicated significant room for improvement, particularly in distinguishing between various types of acne lesions with high fidelity.

Recent advancements in the YOLO architecture, such as YOLOv8, have enhanced object detection performance in terms of both speed and accuracy [10], [11]. The ACNE8M system, which utilized a YOLOv8-based model, achieved notable results with a mean Average Precision (mAP) score of 0.69 and accuracy of 0.976 for detecting and differentiating a wide range of acne types and related conditions [12].

While these results highlight the power of YOLOv8 and its potential for comprehensive acne detection, achieving highly precise automated visual differentiation of specific acne types solely from images remains a significant challenge. Although conditions like fungal acne, acne nodules, and blackheads are medically distinct, their manifestations in images can present subtle or overlapping visual characteristics that challenge even advanced object detection models in achieving the high precision required for reliable clinical support.

Therefore, recognizing these existing challenges and the potential of recent architectures, this research aims to address a specific gap in automated acne analysis. While comprehensive systems like ACNE8M target a broad range of dermatological conditions, a focused investigation into the intricate fine-tuning required to enhance the precision for the visually similar and often misdiagnosed triad of fungal acne, acne nodules, and blackheads remains less explored. The primary objective of this study is to meticulously fine-tune the computationally efficient YOLOv8n model through an iterative exploration of key parameters, including image size, multi-scale training, and data augmentation strategies. The selection of the ‘n’ (nano) variant was deliberate, aiming for a model that is not only precise but also lightweight, making it viable for practical deployment on resource-constrained devices such as mobile phones. The goal is to identify an optimal configuration that significantly improves the reliability and precision of differentiating these specific subtypes, thereby offering a robust foundation for a faster and more efficient diagnostic support tool. Such a tool could potentially reduce diagnostic time, support tele dermatology consultations, and provide accessible preliminary analysis for patients, ultimately facilitating more timely and appropriate treatment.

II. RELATED WORK

Various methodologies have been implemented for automated acne detection. Research utilizing the YOLOv5 architecture, such as the study by Pinasty and Hakim, demonstrated its potential for multi-class acne detection—specifically for blackheads, papules, pustules, nodules, and whiteheads—reporting high overall performance metrics [9]. In contrast, another investigation employing YOLOv5 by Sangha and Rizvi yielded a notably lower mAP@0.5 of 26.50% when differentiating multiple acne classes, underscoring the variability and challenges in this domain [1].

In light of these challenges, the continuous evolution of the YOLO architecture offers significant advancements. The recent iteration, YOLOv8, substantially improves upon its predecessors in terms of speed and accuracy [13]. A notable study developing the ACNE8M system capitalized on these capabilities, successfully detecting a comprehensive set of 12 acne-related conditions, including *pityrosporum folliculitis* (fungal acne), using a large YOLOv8l-seg model and achieving a mean Average Precision (mAP) of 0.69 [12]. While the ACNE8M study demonstrates the broad diagnostic potential of a large-scale YOLOv8 model, our research provides a complementary and more granular analysis. We focus specifically on the fine-tuning process of a lightweight model (YOLOv8n) to optimize detection precision for the challenging subset of

fungal acne, acne nodules, and blackheads, a combination not explicitly addressed in prior works.

This review of existing literature reveals that despite the significant progress achieved in deep learning-based object detection for acne, a critical need persists for enhanced precision in the automated visual differentiation of specific and often challenging acne subtypes, namely fungal acne, nodules, and blackheads. The strategic fine-tuning of advanced architectures like YOLOv8, particularly efficient variants such as YOLOv8n as explored in this study, presents a promising avenue to effectively address these remaining challenges and improve the reliability of automated acne detection systems.

III. RESEARCH METHODOLOGY

A. YOLOv8

YOLOv8 represents the latest advancement in the You Only Look Once (YOLO) series of models, developed by Ultralytics and released on January 10, 2023 [14]. It builds upon the successes of its predecessors, including YOLOv5, by offering significant improvements in speed, accuracy, and overall efficiency in object detection tasks [15], [16]. This makes YOLOv8 particularly well-suited for demanding applications like medical imaging, where precise and rapid detection is crucial [13].

The YOLOv8 architecture, like previous YOLO versions, is generally composed of three main parts: the Backbone, Neck, and Head. The Backbone extracts rich features from the input image, utilizing a modified version of the CSPDarknet53 architecture. The Neck then processes these features, often utilizing components like PANet (Path Aggregation Network) to aggregate multi-scale information, which enhances the model’s ability to detect objects of varying sizes. The Head in YOLOv8 directly concatenates features without enforcing uniform channel dimensions, contributing to a reduction in parameter count and overall tensor size. Finally, the Head makes the ultimate predictions, including bounding box coordinates, class probabilities, and confidence scores for each detected object [11], [14], [15], [17]. YOLOv8 introduces an anchor-free approach to bounding box prediction, simplifying the prediction process and reducing the number of hyperparameters. It also improves its adaptability to objects with varying aspect ratios and scales [14].

YOLOv8 offers several models with varying sizes and complexities, such as nano (YOLOv8n), small (YOLOv8s), medium (YOLOv8m), large (YOLOv8l), and extra-large (YOLOv8x). This provides flexibility for different computational resources and performance requirements [14]. For this research, the YOLOv8n variant was specifically chosen due to its optimal balance between detection accuracy and computational efficiency, rendering it suitable for potential deployment where resources may be constrained while still leveraging the advanced architectural features of the YOLOv8 series for precise lesion differentiation. Its enhanced performance compared to earlier versions, including YOLOv5, makes it more stable and efficient for single- and multi-class object detection, even with smaller datasets [11], [15]. For instance, the larger YOLOv8x model achieved an average precision (AP) of 53.9% with a 640-pixel image size, surpassing YOLOv5’s AP of 50.7% at the same input size [11]. In the

context of acne detection, the advanced capabilities of the selected YOLOv8n architecture are expected to improve the differentiation of visually similar acne types, such as fungal acne, acne nodules, and blackheads, which pose a significant challenge for standard detection models.

Further architectural advancements in YOLOv8 include replacing the C3 layer in YOLOv5 with a new C2f layer. The C2f layer concatenates the outputs of all Bottleneck layers, unlike the C3 layer, which only utilizes the output of the last Bottleneck layer [15]. This new C2f module integrates high-level features with contextual information to enhance detection accuracy [10], [15]. Using mixed-precision training allows YOLOv8 to leverage 16-bit floating-point precision, significantly accelerating training on compatible GPUs while maintaining accuracy and reducing memory consumption [14]. The YOLOv8 series encompasses various models optimized for specific computer vision tasks, including object detection, instance segmentation, pose/key point detection, oriented object detection, and classification. These models are compatible with various operational modes, such as Inference, Validation, Training, and Export, streamlining their use throughout deployment and development [14], [17].

Similar to how the model detects objects in general images by localizing them with bounding boxes and assigning class probabilities, YOLOv8 can be fine-tuned to identify and classify different types of acne lesions, such as fungal acne, nodules, and blackheads. This capability is crucial in dermatology for faster and more accurate diagnoses, as depicted in Fig. 1.

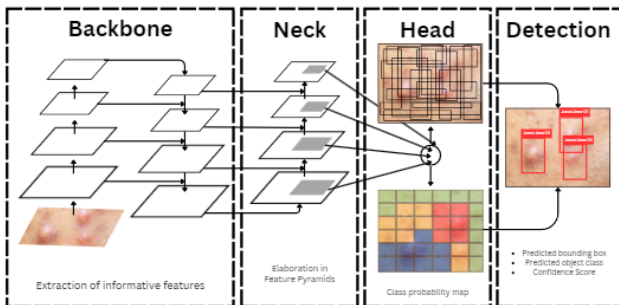


Fig. 1. Implementation of acne detection in YOLOv8.

This integration highlights YOLOv8's versatility beyond general object recognition, demonstrating its potential to significantly enhance diagnostic efficiency and accuracy in specialized medical fields.

B. Datasets

For this study, the dataset was curated by combining images from two publicly available sources: Skin-90¹ and DermNet², resulting in an initial collection of 836 raw images. This aggregation strategy aimed to provide a diverse representation of various acne manifestations. Subsequently, all images underwent a detailed manual annotation process using Roboflow³, a specialized platform for

dataset preparation. To create the ground truth, the annotation was performed by the author under the supervision of a domain expert following a strict set of predefined guidelines to ensure consistency. For each lesion identified, a tight bounding box was drawn to precisely enclose its visible extent. A corresponding class label—blackhead, fungal acne, or nodule—was then assigned based on the visual characteristics described in the subsequent paragraph. Labels were exported in YOLOv8 format, which provides normalized bounding box coordinates `<class-ID>` `<x_center>` `<y_center>` `<width>` `<height>` and class identifiers for each detected object. The dataset was organized into standard train, validation, and test directories, each containing respective subdirectories for images and labels. An illustrative example of the annotation file content is provided in Table I. The dataset was annotated into three distinct classes of acne: blackhead (class ID 0), fungal acne (class ID 1), and nodule (class ID 2).

Blackheads are non-inflammatory open comedones appearing as small, dark follicular plugs from oxidized sebum and cellular debris [6]. Fungal acne (Malassezia folliculitis), distinct from common acne, typically presents as uniform 2-4 mm erythematous papules and pustules, commonly found on seborrheic areas such as the face (especially the forehead), chest, and back, and is often characterized by its pruritic nature [18]. Nodules represent a severe inflammatory form of acne, manifesting as larger than 5 mm, firm, and often tender lesions developing deeper within the skin [6]. Fig. 2 illustrates visual examples of these classes from the dataset.

To address the common challenge of data scarcity in medical imaging and to enhance model robustness, a comprehensive data augmentation strategy was employed, a key technique for improving generalization from limited datasets [19]. Preprocessing and data augmentation techniques were applied to enrich the dataset and enhance model robustness. Preprocessing involved auto-orientation and resizing all images to a uniform 640 × 640 pixel resolution using a stretch method. Data augmentation was then performed on the training images, generating three augmented versions for each original training sample. The augmentation techniques included horizontal flipping, random cropping (0-15% zoom), rotations ($\pm 10^\circ$), and brightness adjustments ($\pm 15\%$). These procedures expanded the total dataset to 2008 images. This final augmented dataset was then partitioned into training, validation, and testing subsets, allocating 1758 images (88%) for training, 164 images (8%) for validation, and 86 images (4%) for testing. This partitioning ensures a robust framework for model training, hyperparameter tuning, and an unbiased final evaluation.

TABLE I. EXAMPLE CONTENT OF A YOLOv8 ANNOTATION TEXT FILE

Class ID	x_center	y_center	width	height
0	0.18203125	0.15234375	0.0375	0.0484375
1	0.43125	0.48046875	0.8625	0.5609375
2	0.74921875	0.81328125	0.234375	0.21953125

C. Evaluation Metrics

To evaluate the performance of the fine-tuned YOLOv8n model, metrics specifically designed for object detection tasks

¹<https://www.kaggle.com/datasets/dinartas/skin90>

²<https://dermnetnz.org/>

³<https://roboflow.com/>



Fig. 2. Visual examples for blackhead, fungal acne, and acne nodule.

were utilized. Unlike simple classification accuracy, which only measures the correctness of labels without considering localization, object detection requires metrics that evaluate both the classification and the localization (bounding box) accuracy. Therefore, the primary metrics employed in this study are Precision, Recall, and mean Average Precision at an Intersection over Union (IoU) threshold of 0.5 (mAP@0.5 or mAP50).

Precision measures the accuracy of the positive predictions made by the model. It is defined as the ratio of correctly identified positive detections (True Positives, TP) to the total number of positive detections predicted by the model (TP + False Positives, FP) [20]. A high precision value indicates that the model makes few false positive errors. The formula for Precision is:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

Recall, also known as sensitivity, measures the model's ability to identify all relevant class instances. It is calculated as the ratio of correctly identified positive detections (TP) to the total number of actual positive instances in the dataset (TP + False Negatives, FN) [20]. A high recall value signifies that the model misses a few actual lesions. The formula for Recall is:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

Average Precision (AP) for each class is calculated as the area under the Precision-Recall (PR) curve. The PR curve plots precision against recall for various confidence thresholds [21]. AP provides a single value summarizing the performance of a model for a specific class. The formula for AP is generally represented as an integral:

$$\text{AP} = \int_0^1 p(r) dr \quad (3)$$

where $p(r)$ is the precision as a function of recall r .

Mean Average Precision (mAP) is the average of AP values across all N classes. For this study, mAP@0.5 (mAP50) is used, indicating an IoU threshold of 0.5 for a detection to be considered a True Positive. A higher mAP50 value generally indicates a more accurate and reliable detection model [20]. The formula for mAP is:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i \quad (4)$$

These metrics were calculated for each class individually and as an overall average across all classes on the validation set during the iterative fine-tuning process and on the test set for final model evaluation.

IV. RESULTS AND ANALYSIS

A. Fine-Tuning Process and Key Findings

The initial investigation, Experiment 1, served as a baseline to establish the foundational performance of the YOLOv8n model. This experiment employed standard configurations, including an input image size (imgsz) of 640x640, a batch size of 16, and default light data augmentations. As detailed in Table II, this baseline configuration yielded an overall mAP50 of 0.582 (Precision: 0.693, Recall: 0.513). Class-specific analysis revealed a satisfactory initial detection for fungal acne (mAP50: 0.797) and acne nodules (mAP50: 0.582). However, blackhead (mAP50: 0.367) was identified as the most challenging class, indicating a need for enhanced feature representation and more robust augmentation strategies in subsequent experiments.

Building upon the baseline observations, the second experiment introduced significant enhancements to the augmentation strategy and loss configuration. Both HSV augmentations (hsv_h=0.02, hsv_s=0.8, hsv_v=0.5) and geometric augmentations (degrees=10.0, translate=0.15, scale=0.6) were applied with increased intensity. Furthermore, the cls_loss weight was adjusted to 0.75 to emphasize classification accuracy, and patience for early stopping was set to 60. This configuration, while maintaining imgsz=640 and batch_size=16 without multi_scale training, resulted in a notable improvement in overall performance, achieving the highest overall mAP50 of 0.594 (Precision:0.676, Recall:0.552), as shown in Table II. Performance in mAP50 for blackhead increased to 0.403, and fungal acne reached its peak mAP50 of 0.806. These results confirmed the efficacy of richer augmentations and adjusted loss weighting.

Subsequently, the third experiment aimed to explore the impact of multi_scale training by enabling the multi_scale=True parameter while retaining imgsz=640 and the aggressive augmentation settings (degrees=10.0, translate=0.15, scale=0.6) from the preceding experiment. The introduction of multi_scale training led to an increase in mAP50 for blackhead to 0.423 and a substantial improvement for acne nodules, which achieved its peak mAP50 of 0.624. However, a significant degradation in performance was observed for fungal acne, with its mAP50 decreasing to 0.692. Consequently, the overall mAP50 slightly declined to 0.580. This outcome highlighted the potential of multi_scale training for classes with size variations but also revealed a critical sensitivity of fungal acne to the combination of multi_scale input and aggressive geometric augmentations.

The fourth experiment was conducted to address the degradation in fungal acne performance observed previously with multi_scale=True and aggressive augmentations. While maintaining imgsz=640, multi_scale=True, batch_size=16, and cls_loss=0.75, the intensity of geometric augmentations was reduced (degrees=5.0, translate=0.1, scale=0.5). This adjustment resulted in a significant recovery of fungal acne performance, achieving an mAP50 of 0.764. Blackhead (0.411) and nodule (0.587) detections remained robust. Notably, this configuration yielded the highest overall Recall (0.572) with an overall mAP50 of 0.587, indicating a well-balanced model across classes, albeit with a comparatively lower Precision of 0.596.

TABLE II. SUMMARY OF ITERATIVE YOLOV8N FINE-TUNING EXPERIMENTS: PARAMETERS AND PERFORMANCE METRICS. BEST OVERALL VALUES AND KEY PER-CLASS PEAKS ARE HIGHLIGHTED IN BOLD

Exp.	imgsz	MS	Scale	Deg.	Transl.	Batch	cls	LR	Time (h)	mAP _{all} ⁵⁰	P _{all}	R _{all}	BH ₅₀	FU ₅₀	NO ₅₀
1	640	No	0.5	0.0	0.1	16	0.50	Auto	0.505	0.582	0.693	0.513	0.367	0.797	0.582
2	640	No	0.6	10.0	0.15	16	0.75	Auto	0.715	0.594	0.676	0.552	0.403	0.806	0.573
3	640	Yes	0.6	10.0	0.15	16	0.75	Auto	0.894	0.580	0.658	0.533	0.423	0.692	0.624
4	640	Yes	0.5	5.0	0.1	16	0.75	Auto	0.813	0.587	0.596	0.572	0.411	0.764	0.587
5	800	Yes	0.6	10.0	0.15	8	0.75	Auto	1.352	0.557	0.691	0.552	0.446	0.687	0.538
6	640	Yes	0.5	10.0	0.15	16	0.75	Auto	0.959	0.589	0.693	0.532	0.442	0.750	0.576

Notes: Exp.: Experiment. MS: Multi-Scale Training. Deg.: Degrees. Transl.: Translate. cls: Classification loss weight. LR: Effective learning rate (Auto \approx 0.001429 for batch 16, may vary for batch 8). Time: Training time in hours. P: Precision. R: Recall. BH: Blackhead. FU: Fungal Acne. NO: Acne Nodules. Subscript ₅₀ denotes mAP at IoU 0.50.

To maximize blackhead detection, the fifth experiment utilized imgsz=800 with multi_scale=True. Due to VRAM constraints, batch_size was reduced to 8. Aggressive geometric augmentations (degrees=10.0, translate=0.15, scale=0.6) and cls_loss=0.75 were employed, with an automatically determined learning rate (approximately 0.001429). This configuration led to the absolute peak mAP50 for blackhead (0.446) and the highest overall Precision (0.691) observed up to that point. However, this was achieved because of significantly reduced performance for fungal acne (0.687) and acne nodules (0.538), resulting in a lower overall mAP50 of 0.557. This experiment underscored the pronounced trade-offs when optimizing for a single, challenging class.

The final iteration, the sixth experiment, aimed to achieve high precision while maintaining a good balance, returning to imgsz=640 with multi_scale=True and batch_size=16. Geometric augmentations were mixed: degrees (10.0) and translate (0.15) remained aggressive, while the scale was set to a gentler 0.5. This configuration yielded an overall mAP50 of 0.589 (the second highest) and, crucially for the paper’s focus, the highest overall Precision (0.693). Blackhead detection was excellent (mAP50: 0.442, the best for imgsz=640), fungal acne performance was well-recovered (mAP50: 0.750), and nodule detection was solid (mAP50: 0.576). This demonstrates that careful tuning of augmentation parameters, particularly the scale factor in conjunction with multi_scale training, can lead to superior precision while maintaining strong overall detection capabilities.

B. Precision-Recall Curve Analysis

The precision-recall (PR) curves from Experiment 6 were analyzed to further evaluate the model’s detection performance, particularly the trade-off between precision and recall. Fig. 3 presents the PR curves for each class (blackhead, fungal acne, acne nodules) and the mean Average Precision (mAP) at an IoU threshold of 0.5.

Fig. 3 shows that the fungal acne class exhibits a PR curve closest to the ideal top-right corner, achieving an Average Precision (AP) of 0.750. This indicates a robust and well-balanced performance between precision and recall for this class. The acne nodules class also performs well with an AP of 0.576; its curve remains relatively stable at high precision levels before declining at higher recall values.

For the blackhead class, the PR curve yields an AP of 0.442. Although this AP value is lower than the other two classes, the curve’s shape suggests that the model can still attain reasonable precision across various recall levels for these challenging, typically small, and numerous instances.

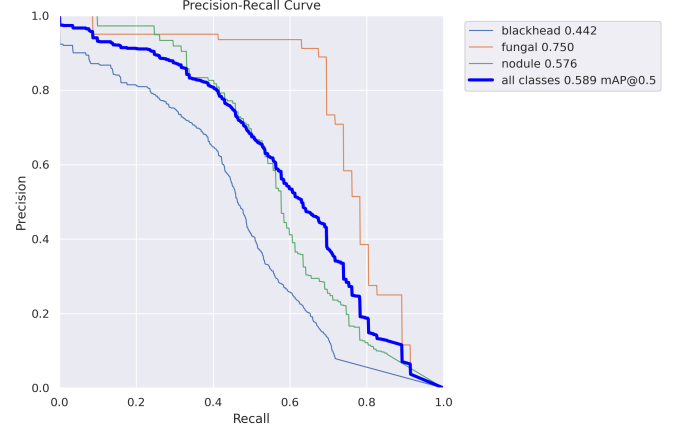


Fig. 3. PR curve of experiment 6.

The more rapid decline in precision with increasing recall for blackhead is characteristic of such object categories.

The mean PR curve (mAP@0.5) reached 0.589, depicting a solid overall model performance. The PR curves for Experiment 6 substantiate the quantitative findings, confirming the model’s capability to achieve high precision levels (overall Precision: 0.693). This robust precision underscores the model’s suitability for ‘Precision Detection’ tasks, where the accurate identification of acne lesions is critical.

To visually substantiate the performance and precision of the optimized YOLOv8n model from Experiment 6, Fig. 4 presents representative detection examples on test images. These images showcase the model’s capabilities in identifying each target acne class: blackhead, fungal acne, and acne nodules. This model was selected for qualitative review due to its achievement of the highest overall precision (P_{all}: 0.693).

Fig. 4a illustrates the model’s performance in detecting blackhead lesions. The model demonstrates a commendable ability to identify numerous blackhead instances, including those that appear subtle or faint. Bounding boxes are generally well-placed, even for detections with confidence scores varying within the 0.3 to 0.6 range. However, the image also reveals false positives, where non-blackhead features or skin textural variations were incorrectly detected as blackhead lesions. While these correct detections of subtle lesions are noteworthy, the presence of false positives contributes to the overall mAP50 of 0.442 for this class (Table II). Although the best for imgsz=640 with multi_scale=True configurations, this score reflects the inherent challenge of blackhead detection due to small object size, high intra-class variance, and potential confusion with other skin features, thereby offering scope for further refinement in specificity.

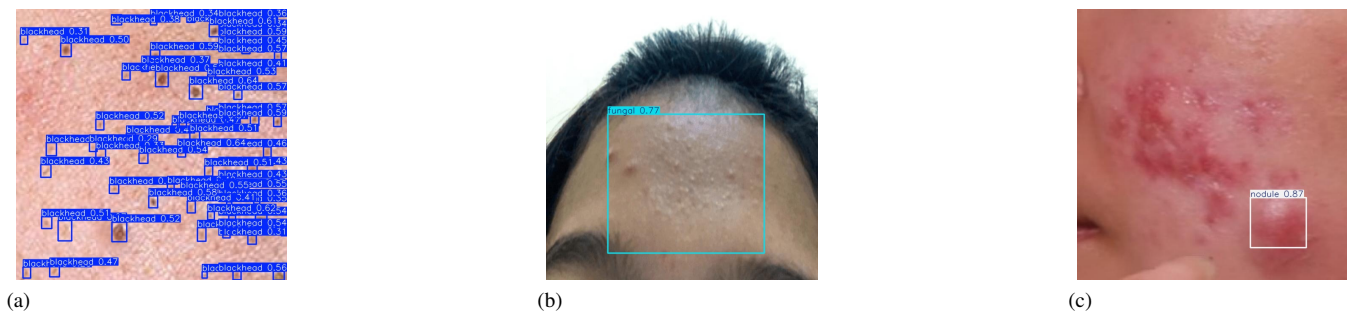


Fig. 4. Examples of precision detection by Experiment 6 on test images, showcasing performance on (a) blackheads, (b) fungal acne, and (c) acne nodules.

Subsequently, Fig. 4b demonstrates the model’s successful detection of fungal acne. The model effectively localizes and delineates the affected area on the forehead, exhibiting the characteristic, often clustered, lesions associated with fungal acne. This qualitative result aligns with this class’s strong mAP50 of 0.750, indicating the model’s proficiency in recognizing its distinct visual patterns.

Fig. 4c provides a representative example of the successful detection of acne nodules. The model identifies and accurately borders a distinct nodule lesion. This performance is consistent with the mAP50 of 0.576 for nodules, illustrating the model’s capability to address this more prominent and often deeper-seated lesion type.

These qualitative examples in Fig. 4 collectively reinforce the quantitative metrics presented in Table II. The model from Experiment 6 exhibits a strong capability for identifying the three types of acne lesions. The high overall precision (P_{all} : 0.693) is supported by generally accurate localizations, although the analysis of blackhead detection highlights that even with good visual examples, the challenge of minimizing false positives remains pertinent for achieving higher mAP scores in complex classes. These visual insights underscore the model’s potential as an effective tool for precise acne lesion identification.

C. Computational Complexity Analysis

In addition to detection accuracy, the computational complexity of the fine-tuned models was evaluated to assess their practicality for real-world deployment. The analysis focused on training duration and inference speed, which are critical factors for model development and application.

The training time for each experiment is detailed in Table II. A clear trend was observed, where configurations with a larger input image size (imgsz=800) and a smaller batch size (e.g., Experiment 5) required significantly longer training durations (1.352 hours) compared to configurations with imgsz=640 and a larger batch size. For instance, the baseline model (Experiment 1) was completed in approximately 0.5 hours. This highlights the computational trade-off associated with training on higher-resolution images.

To evaluate deployment efficiency, the inference speed of the best-performing model (from Experiment 6) was measured on an NVIDIA GeForce RTX 4060 GPU. The performance is summarized in Table III. The model achieved a total processing time of 10.9 milliseconds per image,

with the core inference step taking only 5.3 ms. This rapid processing speed demonstrates the model’s suitability for real-time or near real-time applications, such as in a mobile-based diagnostic support tool.

TABLE III. INFERENCE SPEED OF THE OPTIMIZED MODEL (EXPERIMENT 6)

Metric	Time per Image (ms)
Preprocessing	0.4
Inference	5.3
Postprocessing	5.2
Total	10.9

It is also noted that, since all experiments utilized the YOLOv8n architecture, the model size remained consistent at approximately 3.0 million parameters, with a computational load of 8.1 GFLOPs for a 640x640 input image. This analysis confirms that the proposed fine-tuning methodology not only enhances detection precision but also maintains the high computational efficiency characteristic of the YOLOv8 nano model, making it a viable and practical solution.

D. Discussion

The fine-tuning process culminated in Experiment 6, yielding an optimized YOLOv8n model with a high overall precision of 0.693 and a robust mAP50 of 0.589. When contextualized with related work, these results represent a focused advancement in precision-oriented acne detection. For instance, our model’s multi-class mAP50 of 0.589 significantly surpasses the 26.50% reported by Sangha and Rizvi [1] for a similar multi-class task using YOLOv5. While the study by Pinasty and Hakim [9] also achieved high metrics with YOLOv5, our research demonstrates the successful application of the newer YOLOv8n architecture on a challenging triad that includes fungal acne, a clinically distinct subtype not addressed in their study. In comparison to the comprehensive ACNE8M system [12], which achieved a higher mAP of 0.69 with a large YOLOv8l-seg model, our research offers a complementary contribution by demonstrating that a computationally efficient, lightweight model can be specifically optimized for high precision on a visually ambiguous subset of lesions. This focus on precision with an efficient model is crucial for practical deployment in real-time or mobile-based clinical support tools. Finally, the persistent challenge in detecting small objects, such as

blackheads (mAP50: 0.442), aligns with findings from studies like Huang et al. [22], suggesting a clear direction for future enhancements in small-object detection techniques.

V. CONCLUSION

The primary contribution of this research lies in the systematic and iterative fine-tuning of the YOLOv8n model for the precise detection of three visually challenging acne subtypes: blackhead, fungal acne, and acne nodules. This study demonstrated that through methodical adjustments to key hyperparameters, a lightweight and efficient model can be optimized for high-precision tasks. Experiment 6, which utilized $\text{imgsz}=640$, multi-scale training, and a tailored mixed geometric augmentation, was identified as the optimal configuration. This model achieved the highest overall precision (P_{all} : 0.693) and a strong overall mAP50 of 0.589, with robust per-class scores for blackhead (0.442), fungal acne (0.750), and acne nodules (0.576). The key finding is that careful balancing of augmentation parameters, particularly the scale factor in conjunction with multi-scale training, is critical for enhancing precision while maintaining strong performance across all classes. These results provide a valuable fine-tuning methodology and offer a robust foundation for developing more reliable computer-aided diagnostic tools in dermatology. To promote reproducibility and encourage further innovation in this area, the source code and final model weights associated with this study have been made publicly available at GitHub⁴.

Despite these promising outcomes, this study acknowledges several limitations, including the scope of its public-source dataset and the challenge of detecting small lesions, such as blackheads (mAP50 of 0.442), where some false positives were observed. Future research should prioritize expanding the training dataset with diverse, clinically validated images to enhance model robustness and generalizability. Further investigations could also explore larger YOLOv8 variants, alternative architectures, and advanced techniques to specifically minimize false positives, ultimately paving the way for a more comprehensive and reliable diagnostic tool for clinical deployment.

REFERENCES

- [1] A. Sangha and M. Rizvi, "Detection of acne by deep learning object detection," *medRxiv*, 2021. [Online]. Available: <https://www.medrxiv.org/content/early/2021/12/11/2021.12.05.21267310>
- [2] Q. T. Huynh, P. H. Nguyen, H. X. Le, L. T. Ngo, N.-T. Trinh, M. T.-T. Tran, H. T. Nguyen, N. T. Vu, A. T. Nguyen, K. Suda, K. Tsuji, T. Ishii, T. X. Ngo, and H. T. Ngo, "Automatic acne object detection and acne severity grading using smartphone images and artificial intelligence," *Diagnostics*, vol. 12, no. 8, 2022. [Online]. Available: <https://www.mdpi.com/2075-4418/12/8/1879>
- [3] R. V. Reynolds, H. Yeung, C. E. Cheng, F. Cook-Bolden, S. R. Desai, K. M. Druby, E. E. Freeman, J. E. Keri, L. F. Stein Gold, J. K. Tan, M. M. Tollefson, J. S. Weiss, P. A. Wu, A. L. Zaenglein, J. M. Han, and J. S. Barbieri, "Guidelines of care for the management of acne vulgaris," *Journal of the American Academy of Dermatology*, vol. 90, no. 5, pp. 1006.e1–1006.e30, May 2024. [Online]. Available: <https://doi.org/10.1016/j.jaad.2023.12.017>
- [4] F. V. Hendryanna, Y. W. Syaifudin, M. A. Hendrawan, N. Funabiki, and I. Siradjuddin, "Recognizing acne vulgaris severity levels: An application of faster r-cnn and yolo methods on medical images," *AIP Conference Proceedings*, vol. 3077, no. 1, p. 040006, 07 2024. [Online]. Available: <https://doi.org/10.1063/5.0201131>
- [5] E. Maligna and M.-A. Kurochkina, "Development of the mobile application for assessing facial acne severity from photos," in *2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus)*, 2021, pp. 1790–1793.
- [6] M. Vasam, S. Korutla, and R. A. Bohara, "Acne vulgaris: A review of the pathophysiology, treatment, and recent nanotechnology based advances," *Biochemistry and Biophysics Reports*, vol. 36, p. 101578, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2405580823001590>
- [7] N. Hazarika and M. Archana, "The psychosocial impact of acne vulgaris," *Indian Journal of Dermatology*, vol. 61, no. 5, pp. 515–520, Sep-Oct 2016.
- [8] B. Dréno, "What is new in the pathophysiology of acne, an overview," *Journal of the European Academy of Dermatology and Venereology : JEADV*, vol. 31, no. Suppl 5, pp. 8–12, Sep 2017.
- [9] S. Pinasty and R. B. F. Hakim, "Automatic detection of acne types using the YOLOv5 method," *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIMD)*, vol. 8, no. 1, pp. 236–248, March 2025, p-ISSN: 2614-3372.
- [10] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, p. 1680–1716, Nov. 2023. [Online]. Available: <http://dx.doi.org/10.3390/make5040083>
- [11] M. A. R. Alif and M. Hussain, "Yolov1 to yolov10: A comprehensive review of yolo variants and their application in the agricultural domain," 2024. [Online]. Available: <https://arxiv.org/abs/2406.10139>
- [12] P. K. Nguyen, T. Le, B. A. Nguyen, and P. A. Nguyen, "Acne8m - an acnes detection and differential diagnosis system using ai technologies," *VNUHCM Journal of Science and Technology Development*, vol. 27, no. 3, pp. 3550–3561, Sep. 2024. [Online]. Available: <https://stdj.scienceandtechnology.com.vn/index.php/stdj/article/view/4293>
- [13] A. Widayani, A. Putra, A. Maghrieib, D. Adi, and H. Faishal, "Review of application yolov8 in medical imaging," *Indonesian Applied Physics Letters*, vol. 5, pp. 23–33, 05 2024.
- [14] M. Yaseen, "What is yolov8: An in-depth exploration of the internal features of the next-generation object detector," 2024. [Online]. Available: <https://arxiv.org/abs/2408.15857>
- [15] M. Sohan, T. Ram, and V. Ch, *A Review on YOLOv8 and Its Advancements*, 01 2024, pp. 529–545.
- [16] R. Varghese and S. M., "Yolov8: A novel object detection algorithm with enhanced performance and robustness," in *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, 2024, pp. 1–6.
- [17] P. Hidayatullah, N. Syakrani, M. R. Sholahuddin, T. Gelar, and R. Tubagus, "Yolov8 to YOLO11: A comprehensive architecture in-depth comparative review," *CoRR*, vol. abs/2501.13400, 2025. [Online]. Available: <https://doi.org/10.48550/arXiv.2501.13400>
- [18] M. A. S. Henning, R. Hay, C. Rodriguez-Cerdeira, J. C. Szepietowski, B. M. Piraccini, M. P. Ferreirós, M. Arabatzis, A. Sergeev, P. Nenoff, L. Kotrekova, R. J. Nowicki, J. Faergemann, V. Padovese, A. Prohic, M. Skerlev, P. Schmid-Grendelmeier, B. Sigurgeirsson, G. Gaitanis, P. Lecerf, and D. M. L. Saunte, "Position statement: Recommendations on the diagnosis and treatment of malassezia folliculitis," *Journal of the European Academy of Dermatology and Venereology*, vol. 37, no. 7, pp. 1268–1275, 2023. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jdv.18982>
- [19] L. Alzubaidi, J. Bai, A. Al-Sabaawi, J. Santamaría, A. S. Albahri, B. S. N. Al-dabbagh, M. A. Fadhel, M. Manoufali, J. Zhang, A. H. Al-Timemy, Y. Duan, A. Abdullah, L. Farhan, Y. Lu, A. Gupta, F. Albu, A. Abbosh, and Y. Gu, "A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications," *Journal of Big Data*, vol. 10, no. 1, p. 46, April 2023. [Online]. Available: <https://doi.org/10.1186/s40537-023-00727-2>
- [20] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168–192, January 2021. [Online]. Available: <https://doi.org/10.1016/j.aci.2018.08.003>
- [21] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2020, pp. 237–242.
- [22] H. Huang, B. Wang, J. Xiao, and T. Zhu, "Improved small-object detection using yolov8: A comparative study," *Applied and Computational Engineering*, vol. 41, pp. 80–88, 02 2024.

⁴<https://github.com/eursamajor/Fine-Tuning-YOLOv8n>