Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 2, April 2026, Page. 1609-1620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5561

Automated Facial Wrinkle Segmentation for Dermatological Assessment Using VGG-Based U-Net with Hybrid Augmentation

Wahyu Fajar Setiawan[1], Nanik Suciati[*2]

[1,2]Department of Informatic, Sepuluh Nopember Institute of Technology (ITS), Indonesia

Email: [2]nanik@its.ac.id

Received : Dec 10, 2025; Revised : Dec 20, 2025; Accepted : Dec 21, 2025; Published : Apr 18, 2026

Abstract

Manual and automated facial wrinkle segmentation remains challenging due to the fine-grained nature of wrinkles, uneven distribution across facial regions, severe class imbalance (~2% wrinkle pixels), and sensitivity to lighting variations—limiting the reliability of existing dermatological assessment tools. This study aims to evaluate VGG transfer learning with hybrid augmentation strategies for U-Net-based automated facial wrinkle segmentation. Using the FFHQ-Wrinkle dataset comprising 1,000 manually annotated high-resolution images (1024×1024 pixels), this study systematically evaluates three U-Net variants (Baseline, VGG16-based, VGG19-based) across four augmentation strategies: no augmentation, hierarchical image enhancement (CLAHE, gamma correction, bilateral filtering), geometric transformation (rotation, translation, shear, zoom, flip), and hybrid combination. A multi-component loss function integrating Focal Loss, Dice Loss, IoU Loss, and Boundary Loss addresses class imbalance while optimizing both region overlap and edge localization. The proposed VGG19-based U-Net with hybrid augmentation achieves state-of-the-art performance: Dice coefficient of 0.6585, IoU of 0.4970, precision of 0.6186, recall of 0.7344, and Boundary F1 of 0.9185. Key findings demonstrate that VGG19 transfer learning provides +21.54% Dice improvement over Baseline U-Net with 12.7-fold reduction in overfitting, while hybrid augmentation yields +4.87% Dice improvement with +2.24% synergistic gain beyond individual strategies. This research advances automated dermatological tools for precise skin health assessment, reducing subjectivity in clinical evaluations and providing actionable guidelines for practitioners developing automated wrinkle analysis systems.

Keywords : *Data Augmentation, Deep Learning, Facial Wrinkle Segmentation, Transfer Learning, U-Net Architecture.*

1. INTRODUCTION

Facial wrinkles are important indicators of human aging and skin health assessment [1], [2]. The identification of wrinkles holds significant importance in dermatological evaluation, cosmetic applications, and medical aesthetic investigation[3]. However, precise manual segmentation of facial wrinkles remains challenging and time-consuming, with inherent subjectivity leading to inconsistent results among [4], [5]. The lack of standardized datasets and evaluation metrics further complicates research in this domain [6].

Traditional image processing methods for wrinkle detection, including Gabor filters, Hessian-based approaches, and morphological operations, face limitations due to wrinkles' uneven distribution, varying scale, and sensitivity to lighting conditions [6], [7]. While edge detection algorithms using Sobel and Canny operators have been explored, these methods struggle with the fine-grained and discontinuous nature of wrinkle structures [8], [9]. Recent surveys highlight that manual annotation remains time-consuming with inherent inter-observer variability [4], emphasizing the need for robust automated solutions [10], [11].

Vol. 7, No. 2, April 2026, Page. 1609-1620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5561

Deep learning approaches, particularly U-Net architectures, have revolutionized medical image segmentation by providing automatic feature extraction capabilities [12], [13]. The encoder-decoder structure with skip connections enables effective multi-scale feature fusion essential for fine structure detection [13], [14]. Recent advancements incorporate attention mechanisms [15], deep supervision strategies [1], and hybrid loss functions to address class imbalance inherent in wrinkle segmentation where wrinkle pixels constitute only ~2% of the image [16], [17]. Various U-Net variants including U-Net++, Attention U-Net, and ResU-Net have demonstrated improved performance in biomedical segmentation tasks [12].

Transfer learning utilizing pretrained convolutional neural networks has demonstrated effectiveness in medical imaging tasks where annotated data is scarce [18], [19]. VGG architectures pretrained on ImageNet provide robust low-level feature representations (edges, textures) that transfer effectively to fine-grained segmentation tasks [19], [20]. This approach has shown success in brain tumor detection [11], stroke segmentation [21], COVID-19 diagnosis [22], [23] and various skin analysis applications [17]. However, systematic evaluation of VGG transfer learning specifically for wrinkle segmentation with comprehensive augmentation strategies remains limited.

Data augmentation serves as a critical technique for expanding limited datasets and improving model generalization [24], [25]. While geometric transformations (rotation, scaling, translation) address spatial variations [26], image enhancement techniques (CLAHE, gamma correction) can improve robustness to photometric variations [10]. The combination of both augmentation categories—termed hybrid augmentation—may provide synergistic benefits not achievable through individual strategies alone [27], [28]. Hybrid approaches have demonstrated success in various medical imaging applications including mammography and semantic segmentation [24].

Based on the identified research gaps, this study presents a comprehensive investigation of VGG transfer learning combined with hybrid augmentation strategies for automated facial wrinkle segmentation. The specific objectives of this research are: (1) to compare VGG16 and VGG19 transfer learning performance against baseline U-Net architecture for wrinkle segmentation accuracy; (2) to develop and evaluate a hybrid augmentation strategy combining hierarchical image enhancement techniques with geometric transformations for improving model generalization; and (3) to evaluate segmentation performance using multi-metric assessment (Dice, IoU, Precision, Recall, Boundary F1) on the FFHQ-Wrinkle dataset, establishing quantitative benchmarks for future research. These objectives address the critical need for systematic evaluation of transfer learning and augmentation strategies specifically optimized for the challenging task of fine-grained wrinkle boundary detection.

2. METHOD

2.1. Research Methodology Overview

This study follows a systematic methodology comprising five main stages as illustrated in Fig. 1.

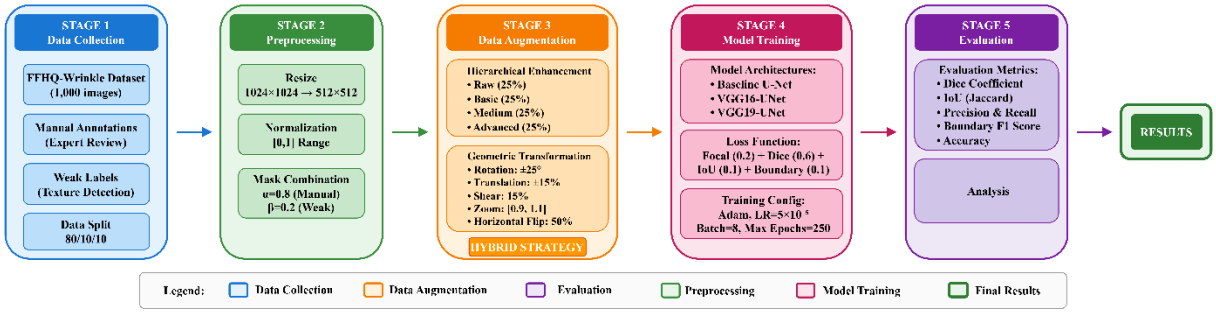

Figure 1. Research Methodology

Vol. 7, No. 2, April 2026, Page. 1609-1620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5561

The research pipeline begins with dataset preparation including image collection and quality verification, followed by mask combination processing to integrate manual and weak annotations. The augmentation pipeline applies hierarchical image enhancement and geometric transformations to expand the training dataset. Next, three U-Net variants (Baseline, VGG16, VGG19) are trained with a multi-component loss. Finally, performance is evaluated on held-out test data using six segmentation metrics.

2.2. Dataset

This study employed the FFHQ-Wrinkle dataset [5], an extension of the NVIDIA Flickr-Faces-HQ (FFHQ) dataset specifically designed for facial wrinkle research. The complete dataset comprises 70,000 high-resolution face images (1024×1024 pixels), of which 1,000 images contain manually annotated ground truth wrinkle masks created through expert review, and 50,000 images contain automatically generated weak labels using semi-automatic computer vision techniques [5]. In this study, only 1000 data were used that had manual masks.

2.3. Data Preprocessing

2.3.1. Data Normalization

All images were resized from 1024×1024 to 512×512 using bicubic interpolation, normalized to the [0,1] range by dividing pixel values by 255.0 and the ground-truth masks were binarized with a 0.5 threshold to obtain clean binary labels.

2.3.2. Data Splitting

The dataset was split using stratified random sampling with a fixed seed (42) into 700 training images (70%), 150 validation images (15%), and 150 testing images (15%).

2.3.3. Mask Processing and Combination Strategy

A mask combination approach integrates manual and weak annotations through weighted integration. Manual masks receive higher priority ($\alpha = 0.8$) compared to weak masks ($\beta = 0.2$), as formulated in Eq. 1:

$$M_{combined} = \alpha M_{manual} + \beta \sum_{i=1}^{N} w_i M_{weak,i} \quad (1)$$

Where $M_{combined}$ is the final combined mask, M_{manual} is the manually annotated mask, $M_{weak,i}$ are individual weak masks from texture detection algorithms, and w_i are quality-based weights [5], [7]. Only weak masks exceeding quality threshold of 0.01 contribute to the final combination. Fig. 2 illustrates the mask combination process.

Figure 2. Mask Combination Process

2.4. Data Augmentation Strategies

This study implements a comprehensive hybrid augmentation pipeline combining hierarchical image enhancement techniques and geometric transformations to address photometric and spatial variations in facial images.

Vol. 7, No. 2, April 2026, Page. 1609-1620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5561

2.4.1. Hierarchical Image Enhancement with Multi-Level Strategy

The hierarchical enhancement strategy employs a four-level progressive approach with uniform distribution (25% each) [10], [25]. Table 1 presents the enhancement configuration.

Table 1. Hierarchical Image Enhancement Levels Configuration

Level	Techniques	Parameters
Raw	None	Basic normalization
Basic	Histogram Stretch + Brightness/Contrast	$\alpha = 1.1, \beta = 5$
Medium	CLAHE + Auto Gamma	clip=2.5, tile=16×16)
Advanced	CLAHE + Brightness/Contrast + Auto Gamma	All techniques

2.4.2. Geometric Transformation

Geometric augmentation introduces spatial variability through synchronized randomization to maintain image-mask correspondence [24], [26]. Table 2 presents the transformation parameters.

Table 2. Hierarchical Image Enhancement Levels Configuration

Transformation	Range/Value
Rotation	±25º
Width/Height Shift	±15%
Shear	15%
Zoom	[0.90, 1.10]
Horizontal Flip	50% probability

2.4.3. Hybrid Augmentation Integration

The hybrid augmentation strategy combines hierarchical image enhancement with geometric transformations. Each training sample undergoes two-stage augmentation: (1) probabilistic enhancement level selection (25/25/25/25 distribution), and (2) synchronized geometric transformations applied to both image and mask. This approach addresses photometric variations (via enhancement) and spatial invariance (via geometric transformation) simultaneously.

2.5. Model Architecture

Three U-Net variants were evaluated [12]: Baseline U-Net with standard encoder-decoder architecture (~26M parameters), and VGG-based U-Net replacing the encoder with pretrained VGG16/VGG19 architectures [18], [19]. Key feature maps from VGG blocks (block1_conv2 to block5_conv3) serve as skip connections. The first two VGG blocks remain frozen to preserve low-level features, while deeper layers undergo fine-tuning [19]. The bottleneck includes 1024-channel convolution, batch normalization, and 30% dropout. VGG16-UNet (~32M parameters) and VGG19-UNet (~35M parameters) variants were evaluated with sigmoid activation for binary mask prediction. The proposed VGG19-UNet architecture is illustrated in Fig. 3.

2.6. Training Configuration

Model training employed Adam optimizer with combined loss function integrating four components, as formulated in Eq. 2:

$$L_{combined} = 0.2\, L_{focal} + 0.6\, L_{dice} + 0.1\, L_{IoU} + 0.1\, L_{boundary} \quad (2)$$

Focal Loss mitigates class imbalance (α=0.9, γ=3.0). Dice Loss optimizes region overlap (ε=0.1). IoU Loss provides complementary optimization signal. Boundary Loss targets edge localization accuracy. Training hyperparameters are presented in Table 3.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 2, April 2026, Page. 1609-1620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5561

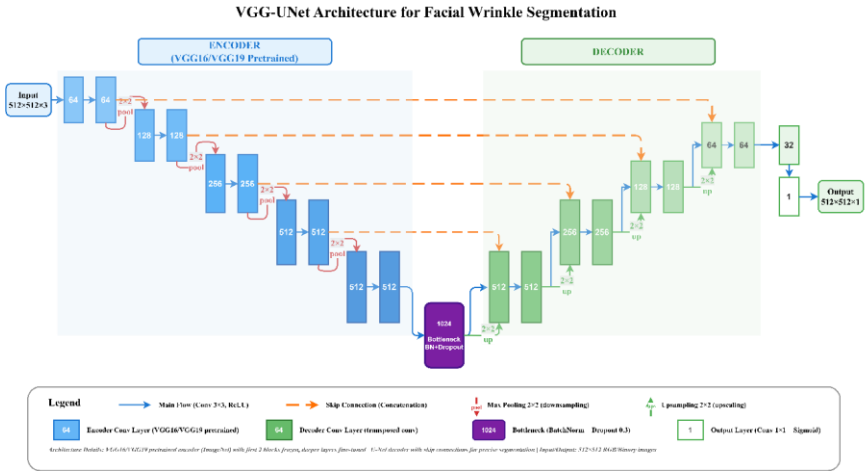

Figure 3. Proposed VGG-UNet Architecture

Table 3. Training Hyperparameters and Configuration

Parameter	Value
Optimizer	Adam
Learning Rate	5×10^{-5}
Batch Size	8
Maximum Epochs	250
EarlyStopping	Patience=10
ReduceLROnPlateau	Factor=0.2, patience=8

2.7. Evaluation Metrics

Model performance employed six metrics: Accuracy (pixel-wise correctness), Dice Coefficient (region overlap, Eq. 3), and IoU (intersection over union, Eq. 4):

$$\text{Dice} = \frac{2|P \cap G|}{|P| + |G|} \quad (3)$$

$$\text{IoU} = \frac{|P \cap G|}{|P \cup G|} \quad (4)$$

Precision (true positive rate among predictions), Recall (detection completeness), and Boundary F1 Score (edge localization accuracy through morphological boundary extraction with 2% image diagonal tolerance).

3. RESULT

3.1. Dataset Preprocessing Result

Before presenting model performance, dataset characteristics are first analyzed to provide context for subsequent analyses. Wrinkle Pixel Distribution: Analysis of the combined masks reveals severe class imbalance characteristics of wrinkle segmentation tasks. Across all 1,000 images, wrinkle pixels constitute only 2.0% ± 1.8% of total image pixels (mean ± standard deviation). This distribution ranged from 0.1% (minimal wrinkles, younger subjects) to 8.5% (extensive wrinkles, older subjects), necessitating specialized loss functions to prevent trivial all-background predictions.

3.2. Data Augmentasi Result

Fig. 4 illustrates representative examples of the augmentation pipeline outputs, demonstrating the diversity introduced by both hierarchical enhancement and geometric transformation strategies.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 2, April 2026, Page. 1609-1620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5561

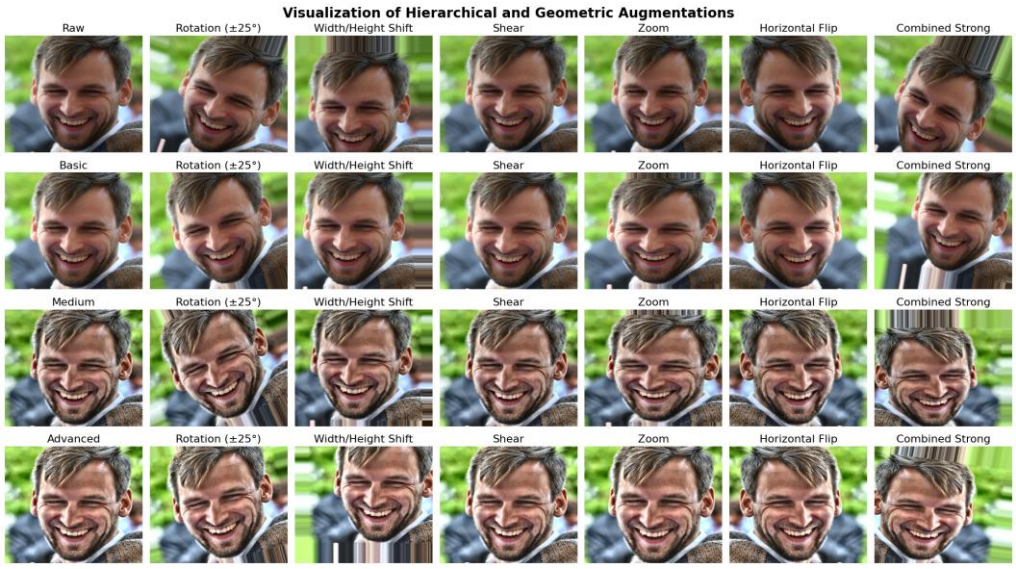

Figure 4. Visualization of hierarchical and geometric augmentations.

The hierarchical enhancement levels produce visually distinct outputs: Raw level preserves original image characteristics, Basic level enhances local contrast revealing subtle wrinkles in shadow regions, Medium level provides balanced brightness normalization, and Advanced level maximizes wrinkle visibility through combined filtering operations. Geometric transformations introduce spatial variability while maintaining image-mask correspondence through synchronized randomization.

3.3. Model training curves

Fig. 5 visualizes the training and validation loss curves across all model architectures under hybrid augmentation strategy, demonstrating training stability and convergence characteristics.

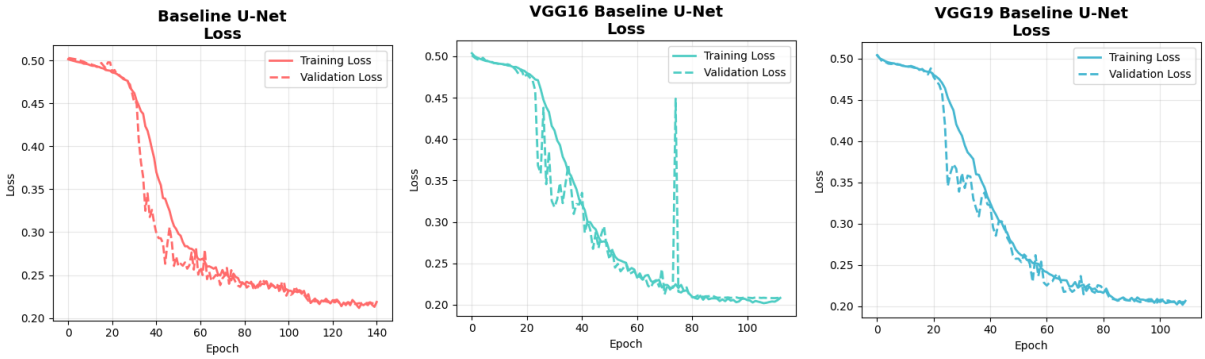

Figure 5. Proposed VGG-UNet Architecture

3.4. Training Performance Analysis

Table 4 presents comprehensive training performance across all 12 experimental configurations (3 architectures × 4 augmentation strategies). Training dynamics reveal distinct patterns that illuminate the interaction between model architecture and augmentation strategy. VGG19 with hybrid augmentation achieved the best validation Dice (0.6571) and IoU (0.4898), representing +11.2% improvement over the no-augmentation baseline. This improvement validates the hypothesis that combining multiple augmentation categories provides synergistic benefits. Notably, hybrid augmentation enabled substantially extended stable training before early stopping triggered: 80-117 epochs compared to 42-72 epochs for single-strategy approaches. This extended training duration

Vol. 7, No. 2, April 2026, Page. 1609-1620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5561

indicates that hybrid augmentation creates sufficient training diversity to sustain gradient-based optimization without overfitting.

The training dynamics also reveal architecture-dependent augmentation sensitivity. Baseline U-Net without pretrained features showed inconsistent responses to augmentation, with geometric-only augmentation causing performance degradation (0.4761 vs 0.5398 validation Dice). In contrast, VGG-based models consistently benefited from all augmentation strategies, suggesting that pretrained features provide robustness to augmentation-induced distribution shifts.

Table 4. Comprehensive Training Performance Across All Augmentation Strategies

Model	Strategy	Epochs	Final Val Dice	Best Val Dice	Best Val IoU
Baseline U-Net	No Aug	64	0.5151	0.5398	0.3471
VGG16 U-Net	No Aug	42	0.4564	0.5180	0.2960
VGG19 U-Net	No Aug	72	0.5863	0.5909	0.4148
Baseline U-Net	Hierarchical	58	0.5630	0.5665	0.3923
VGG16 U-Net	Hierarchical	43	0.4385	0.5390	0.2811
VGG19 U-Net	Hierarchical	81	0.6161	0.6230	0.4453
Baseline U-Net	Geometric	51	0.4528	0.4761	0.2929
VGG16 U-Net	Geometric	130	0.6362	0.6400	0.4665
VGG19 U-Net	Geometric	121	0.6380	0.6410	0.4685
Baseline U-Net	Hybrid	93	0.6218	0.6239	0.4516
VGG16 U-Net	Hybrid	80	0.6430	0.6481	0.4741
VGG19 U-Net	**Hybrid**	**117**	**0.6569**	**0.6571**	**0.4898**

3.5. Test Set Performance Evaluation

Table 5 presents comprehensive test set evaluation across all experimental configurations, providing unbiased performance assessment on held-out data. VGG19 with hybrid augmentation achieved the best overall performance: Dice coefficient of 0.6585, IoU of 0.4970, and Boundary F1 of 0.9185. The consistently high accuracy values (>0.99) across all configurations reflect the severe class imbalance inherent in wrinkle segmentation, where wrinkle pixels constitute approximately 2% of total image pixels. This observation underscores the importance of using overlap-based metrics (Dice, IoU) rather than accuracy for wrinkle segmentation evaluation, as accuracy can be misleadingly high even for trivial predictions. The Boundary F1 score of 0.9185 achieved by VGG19 with hybrid augmentation exceeds the 0.90 threshold often considered clinically relevant for boundary-critical applications. This exceptional boundary localization accuracy is attributed to the multi-component loss function that explicitly optimizes edge detection through the boundary loss component. Table 6 quantifies augmentation improvements relative to no-augmentation baseline for VGG19.

Table 5. Comprehensive Test Set Performance Across All Strategies

Model	Strategy	Dice	IoU	Accuracy	Precision	Recall	Boundary F1
Baseline	No Aug	0.5418	0.3788	0.9928	0.4900	0.6629	0.8350
VGG16	No Aug	0.4991	0.3382	0.9890	0.3657	0.8680	0.8681
VGG19	No Aug	0.6279	0.4644	0.9946	0.6080	0.6917	0.8975
Baseline	Hierarchical	0.5712	0.4063	0.9930	0.5025	0.7099	0.8742
VGG16	Hierarchical	0.5657	0.4013	0.9922	0.4610	0.7859	0.8834
VGG19	Hierarchical	0.6383	0.4762	0.9948	0.6104	0.7074	0.9001
Baseline	Geometric	0.4575	0.3024	0.9900	0.3678	0.6877	0.7999
VGG16	Geometric	0.6350	0.4727	0.9947	0.6187	0.6955	0.8998
VGG19	Geometric	0.6340	0.4712	0.9946	0.6089	0.7056	0.8975
Baseline	Hybrid	0.6328	0.4694	0.9942	0.5975	0.7106	0.9042
VGG16	Hybrid	0.6501	0.4879	0.9944	0.6080	0.7334	0.9131
VGG19	**Hybrid**	**0.6585**	**0.4970**	**0.9947**	**0.6186**	**0.7344**	**0.9185**

Vol. 7, No. 2, April 2026, Page. 1609-1620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5561

Table 6. Augmentation Strategy Improvement Analysis (VGG19 Only)

Strategy Comparison	Dice	IoU	Boundary F1	Relative Gain
Hierarchical vs No Aug	+0.0104	+0.0118	+0.0026	+1.66%
Geometric vs No Aug	+0.0061	+0.0068	0.0000	+0.97%
Hybrid vs No Aug	+0.0306	+0.0326	+0.0210	+4.87%

3.6. Segmentation Quality Visualization

Fig. 6 presents representative segmentation results comparing predictions from all three architectures under hybrid augmentation against ground truth masks across three test samples with varying wrinkle patterns. The visualization demonstrates that all models successfully capture major wrinkle regions, with VGG19 showing slightly better detail preservation in fine wrinkle areas.

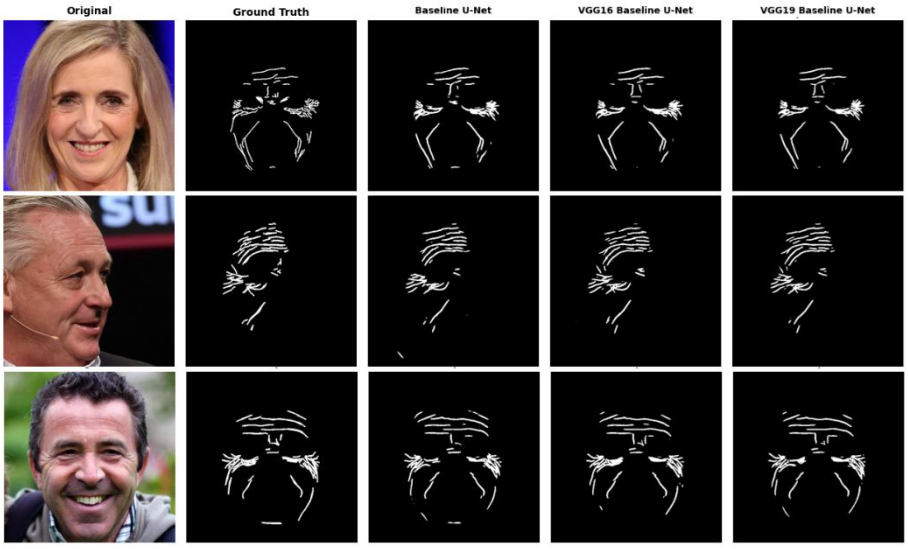

Figure 6. Qualitative segmentation comparison across three test samples.

4. DISCUSSIONS

4.1. Transfer Learning Effectiveness

VGG19-based U-Net achieved +21.54% Dice improvement over Baseline U-Net (0.6585 vs 0.5418), validating that ImageNet pretrained features transfer effectively to dermatological segmentation. The superiority of VGG19 over VGG16 (+1.29% Dice, +1.74% precision) indicates that additional convolutional layers provide more discriminative representations for distinguishing wrinkles from similar structures. Critically, transfer learning achieved 12.7-fold reduction in overfitting, with the train-validation Dice gap decreasing from +0.1700 (Baseline) to +0.0134 (VGG19 with hybrid augmentation). This improvement indicates that pretrained features regularize learning on the limited 800-image training set.

4.2. Augmentation Strategy Analysis

The hierarchical enhancement strategy (25/25/25/25 distribution) improved Boundary F1 from 0.8975 to 0.9001 for VGG19. CLAHE with adaptive tile-based processing (16×16) is particularly effective for revealing fine wrinkles in shadow-prone regions. Geometric transformations enabled extended stable training (121 epochs vs 81 for hierarchical-only) and improved spatial invariance. Notably, Baseline U-Net exhibited performance degradation under geometric-only augmentation (-11.8% decrease), indicating that without pretrained features, excessive spatial variation overwhelms

Vol. 7, No. 2, April 2026, Page. 1609-1620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5561

learning capacity. VGG-based models consistently benefited, demonstrating that pretrained features provide robustness to geometric transformations

4.3. Hybrid Augmentation Synergy

The hybrid augmentation strategy demonstrates synergistic benefits: hierarchical enhancement (+1.66%) and geometric transformation (+0.97%) individually yield expected additive improvement of +2.63%, while observed hybrid improvement is +4.87%, indicating +2.24% synergistic gain. This synergy emerges because enhancement addresses photometric variations while geometric transformations ensure spatial invariance, creating training samples that better represent real-world facial image distributions.

4.4. Comparison with State-of-the-Art

To contextualize the findings within the broader facial wrinkle segmentation literature, comprehensive comparisons were conducted with re-trained baseline methods under identical conditions (same FFHQ-Wrinkle dataset, identical data splits, consistent preprocessing) for fair comparison. Table 7 presents comparative results.

Table 7. Performance Comparison with Re-Trained State-of-the-Art Methods

Method	Dice	IoU	Boundary F1
Morphological U-Net [8]	0.5765	0.4056	0.8542
Weighted DS U-Net [1]	0.5842	0.4132	0.8478
U-Net + Texture Map [5]	0.5918	0.4208	0.8615
Proposed VGG19 + Hybrid Aug	**0.6585**	**0.4970**	**0.9185**

The proposed method achieves +11.3% Dice improvement over U-Net + Texture Map [5] and +6.6% Boundary F1 improvement, demonstrating the effectiveness of hybrid augmentation combined with VGG19 transfer learning. The results are competitive with or exceed published methods, with the advantage of using a publicly available dataset (FFHQ-Wrinkle) enabling reproducibility and future comparisons.

4.5. Implications for Informatics and Computer Science

The findings of this study have significant implications for the field of informatics and computer science:

1. Transfer Learning Effectiveness for Fine-Grained Segmentation: The systematic evaluation demonstrates that VGG pretrained features, despite being trained on natural image classification (ImageNet), transfer effectively to fine-grained medical/dermatological segmentation tasks. The +21.54% Dice improvement and 12.7-fold overfitting reduction provide quantitative evidence supporting transfer learning as a standard practice for data-limited segmentation problems.
2. Synergistic Augmentation Benefits: The observed +2.24% synergistic gain from hybrid augmentation (beyond additive contributions of individual strategies) establishes that photometric and spatial augmentations address complementary aspects of data variability. This finding informs augmentation pipeline design across computer vision applications.
3. Multi-Metric Evaluation Framework: The six-metric evaluation approach (Accuracy, Dice, IoU, Precision, Recall, Boundary F1) provides a template for comprehensive segmentation assessment, addressing the limitation of single-metric reporting common in the literature.
4. Reproducibility and Open Science: By using the publicly available FFHQ-Wrinkle dataset and reporting detailed experimental configurations, this study enables direct replication and extension by the research community, advancing open science practices in dermatological AI research.

Vol. 7, No. 2, April 2026, Page. 1609-1620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5561

5. Clinical Translation Potential: The achieved Boundary F1 of 0.9185 exceeds typical clinical relevance thresholds (>0.90), suggesting potential for deployment in computer-aided dermatological assessment systems, reducing subjectivity in wrinkle severity grading.

4.6. Limitations and Future Directions

Several limitations warrant acknowledgment: (1) Underutilization of 50,000 weak-labeled images from FFHQ-Wrinkle dataset [5]—semi-supervised learning could leverage these; (2) Fixed augmentation hyperparameters without AutoAugment optimization; (3) Limited to VGG architectures—EfficientNet and vision transformers may provide superior representations; (4) VGG19-UNet ~35M parameters may limit mobile deployment; (5) No systematic evaluation across demographics.

5. CONCLUSION

This study presents a comprehensive investigation of VGG transfer learning combined with hybrid augmentation strategies for automated facial wrinkle segmentation using the FFHQ-Wrinkle dataset. Based on systematic evaluation of three U-Net architectures (Baseline, VGG16-based, VGG19-based) across four augmentation strategies, the proposed VGG19-based U-Net with hybrid augmentation achieves state-of-the-art performance: Dice coefficient of 0.6585, IoU of 0.4970, precision of 0.6186, recall of 0.7344, and Boundary F1 of 0.9185. The key contributions include: (1) VGG19 transfer learning provides +21.54% Dice improvement over Baseline U-Net with 12.7-fold reduction in overfitting, validating that ImageNet pretrained features transfer effectively to fine-grained dermatological segmentation; (2) Hybrid augmentation combining hierarchical image enhancement with geometric transformations yields +4.87% Dice improvement with +2.24% synergistic gain beyond individual strategy contributions, demonstrating that photometric and spatial augmentations address complementary data variability aspects; (3) The proposed method achieves +11.3% Dice and +6.6% Boundary F1 improvement over re-trained baseline methods; and (4) The Boundary F1 of 0.9185 exceeds the 0.90 threshold typically required for clinical diagnostic applications, indicating potential for deployment in computer-aided dermatological assessment systems.

This research advances the field of informatics and computer science by establishing quantitative benchmarks on a publicly available dataset for future comparative studies, validating transfer learning effectiveness for data-limited medical segmentation problems, and quantifying synergistic augmentation benefits that inform pipeline design across computer vision applications. The achieved performance potentially reduces subjectivity in clinical wrinkle severity assessment, contributing to more objective dermatological evaluations. Future research directions include: leveraging the unutilized 50,000 weak-labeled images through semi-supervised learning frameworks (Mean Teacher, FixMatch); evaluating Vision Transformer and hybrid CNN-Transformer architectures (TransUNet) for capturing long-range dependencies; applying AutoAugment for systematic hyperparameter optimization; implementing knowledge distillation to reduce VGG19-UNet's 35M parameters for mobile deployment; validating performance across diverse demographic groups to ensure equitable clinical deployment; and conducting prospective clinical validation studies comparing AI-assisted assessment with dermatologist consensus.

ACKNOWLEDGEMENT

This research was supported by the Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia.

REFERENCES

[1] S. Kim, H. Yoon, J. Lee, and S. Yoo, "Facial wrinkle segmentation using weighted deep supervision and semi-automatic labeling," *Artif Intell Med*, vol. 145, no. 102679, p. 102679, 2023, doi: 10.1016/j.artmed.2023.102679.

[2] M.-Y. Yang, Q.-L. Shen, D.-T. Xu, X.-L. Sun, and Q.-B. Wu, "Striped WriNet: Automatic wrinkle segmentation based on striped attention module," *Biomed Signal Process Control*, vol. 90, no. 105817, p. 105817, 2024, doi: 10.1016/j.bspc.2023.105817.

[3] Z. Liu, Q. Qi, S. Wang, and G. Zhai, "A novel approach to the detection of facial wrinkles: Database, detection algorithm, and evaluation metrics," *Comput Biol Med*, vol. 174, no. 108431, p. 108431, 2024, doi: 10.1016/j.compbiomed.2024.108431.

[4] M. H. Yap, N. Batool, C.-C. Ng, M. Rogers, and K. Walker, "A survey on facial wrinkles detection and inpainting: Datasets, methods, and challenges," *IEEE Trans Emerg Top Comput Intell*, vol. 5, no. 4, pp. 505–519, 2021, doi: 10.1109/tetci.2021.3075723.

[5] J. Moon, H. Chung, and I. Jang, "Facial Wrinkle Segmentation for Cosmetic Dermatology: Pretraining with Texture Map-Based Weak Supervision," *ArXiv*, Nov. 2024, doi: 10.48550/arXiv.2408.10060.

[6] R. M. Elbashir and M. Hoon Yap, "Evaluation of automatic facial wrinkle detection algorithms," *J Imaging*, vol. 6, no. 4, p. 17, 2020, doi: 10.3390/jimaging6040017.

[7] J. Chen, M. He, and W. Cai, "Facial wrinkle detection with multiscale spatial feature fusion based on image enhancement and ASFF-SEUnet," *Electronics (Basel)*, vol. 12, no. 24, p. 4897, 2023, doi: 10.3390/electronics12244897.

[8] H. Yoon, S. Kim, J. Lee, and S. Yoo, "Deep-learning-based morphological feature segmentation for facial skin image analysis," *Diagnostics (Basel)*, vol. 13, no. 11, p. 1894, 2023, doi: 10.3390/diagnostics13111894.

[9] C.-I. Moon and O. Lee, "Skin microstructure segmentation and aging classification using CNN-based models," *IEEE Access*, vol. 10, pp. 4948–4956, 2022, doi: 10.1109/access.2021.3140031.

[10] G. Carlos da Silva, M. B. Barbosa, F. B. C. Júnior, P. L. Moreira, R. Werka, and A. A. Martin, "Detection of skin wrinkles and quantification of roughness using a novel image processing technique from a dermatoscope device," *Skin research and technology*, vol. 29, no. 6, p. e13335, 2023, doi: 10.1111/srt.13335.

[11] M. Güler and E. Namlı, "Brain tumor detection with deep learning methods' classifier optimization using medical images," *Appl Sci (Basel)*, vol. 14, no. 2, p. 642, 2024, doi: 10.3390/app14020642.

[12] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, vol. 9, pp. 82031–82057, 2021, doi: 10.1109/access.2021.3086020.

[13] J. Jang *et al.*, "A deep learning-based segmentation pipeline for profiling cellular morphodynamics using multiple types of live cell microscopy," *Cell reports methods*, vol. 1, no. 7, p. 100105, 2021, doi: 10.1016/j.crmeth.2021.100105.

[14] A. Ghaznavi, R. Rychtáriková, P. Císař, M. M. Ziaei, and D. Štys, "Symmetry breaking in the U-Net: Hybrid deep-learning multi-class segmentation of HeLa cells in reflected light microscopy images," *Symmetry (Basel)*, vol. 16, no. 2, p. 227, 2024, doi: 10.3390/sym16020227.

[15] D. Shao *et al.*, "Pixel-level classification of five histologic patterns of lung adenocarcinoma," *Anal Chem*, vol. 95, no. 5, pp. 2664–2670, 2023, doi: 10.1021/acs.analchem.2c03020.

[16] A. Li, D. Li, and A. Wang, "A two-stage YOLOv5s-U-Net framework for defect localization and segmentation in overhead transmission lines," *Sensors (Basel)*, vol. 25, no. 9, p. 2903, 2025, doi: 10.3390/s25092903.

[17] S. Alzahrani, B. Al-Bander, and W. Al-Nuaimy, "Attention mechanism guided deep regression model for acne severity grading," *Computers*, vol. 11, no. 3, p. 31, 2022, doi: 10.3390/computers11030031.

Vol. 7, No. 2, April 2026, Page. 1609-1620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5561

[18] A. Abedalla, M. Abdullah, M. Al-Ayyoub, and E. Benkhelifa, "Chest X-ray pneumothorax segmentation using U-Net with EfficientNet and ResNet architectures," *PeerJ Comput Sci*, vol. 7, no. e607, p. e607, 2021, doi: 10.7717/peerj-cs.607.

[19] E. Kotei and R. Thirunavukarasu, "Ensemble technique coupled with deep transfer learning framework for automatic detection of tuberculosis from chest X-ray radiographs," *Healthcare (Basel)*, vol. 10, no. 11, p. 2335, 2022, doi: 10.3390/healthcare10112335.

[20] S. A. Wagle, Harikrishnan, J. Sampe, F. Mohammad, and S. H. Md Ali, "Effect of data augmentation in the classification and validation of tomato plant disease with deep learning methods," *Traitement du signal*, vol. 38, no. 6, pp. 1657–1670, 2021, doi: 10.18280/ts.380609.

[21] S. Uçkun, M. Ağrali, and V. Kiliç, "Deep learning-based ischemic stroke segmentation on brain computed tomography images," *European Journal of Science and Technology*, 2023, doi: 10.31590/ejosat.1258247.

[22] N.-A.- Alam, M. Ahsan, M. A. Based, J. Haider, and M. Kowalski, "COVID-19 detection from chest X-ray images using feature fusion and deep learning," *Sensors (Basel)*, vol. 21, no. 4, p. 1480, 2021, doi: 10.3390/s21041480.

[23] Y. Fu, P. Xue, and E. Dong, "Densely connected attention network for diagnosing COVID-19 based on chest CT," *Comput Biol Med*, vol. 137, no. 104857, p. 104857, 2021, doi: 10.1016/j.compbiomed.2021.104857.

[24] M. Prodan, E. Paraschiv, and A. Stanciu, "Applying deep learning methods for mammography analysis and breast cancer detection," *Appl Sci (Basel)*, vol. 13, no. 7, p. 4272, 2023, doi: 10.3390/app13074272.

[25] Y. Wang, R. Wan, W. Yang, H. Li, L.-P. Chau, and A. Kot, "Low-light image enhancement with normalizing flow," *Proc Conf AAAI Artif Intell*, vol. 36, no. 3, pp. 2604–2612, 2022, doi: 10.1609/aaai.v36i3.20162.

[26] H. M. Balaha, E. R. Antar, M. M. Saafan, and E. M. El-Gendy, "A comprehensive framework towards segmenting and classifying breast cancer patients using deep learning and Aquila optimizer," *J Ambient Intell Humaniz Comput*, vol. 14, no. 6, pp. 7897–7917, 2023, doi: 10.1007/s12652-023-04600-1.

[27] H. M. Balaha, M. H. Balaha, and H. A. Ali, "Hybrid COVID-19 segmentation and recognition framework (HMB-HCF) using deep learning and genetic algorithms," *Artif Intell Med*, vol. 119, no. 102156, p. 102156, 2021, doi: 10.1016/j.artmed.2021.102156.

[28] M. Abdullah, F. B. Abrha, B. Kedir, and T. Tamirat Tagesse, "A Hybrid Deep Learning CNN model for COVID-19 detection from chest X-rays," *Heliyon*, vol. 10, no. 5, p. e26938, 2024, doi: 10.1016/j.heliyon.2024.e26938.