Academicia Review-A Multidisciplinary Online Journal
ISSN (Online): 3070-6726
Website: https://academicia.org
Volume 2, Issue 3, March, 2026

Hybrid Attention U-Net with Evolutionary Optimization for Skin Lesion Segmentation on ISIC 2018 Dataset

Author: **Nazarov Jasurbek Ilkhomjon ugli**
Affiliation: Kokand University
Corresponding Email: jasurbek3113nazarov@gmail.com

Abstract

Skin lesion segmentation is a fundamental step in automated melanoma detection systems. Although deep learning architectures such as U-Net have significantly advanced the field, challenges persist due to variations in lesion size, shape, color, and the presence of imaging artifacts. In this paper, we propose a Hybrid Attention U-Net (HA-UNet) that integrates channel-wise and spatial attention mechanisms to improve feature representation. To further enhance segmentation accuracy and reduce manual hyperparameter tuning, we employ Differential Evolution (DE) as an evolutionary optimization strategy. The proposed method is evaluated on the ISIC 2018 dataset, consisting of 2,594 dermoscopic images with corresponding ground truth masks. Experimental results show that HA-UNet with DE optimization achieves a Dice similarity coefficient of 0.908 and a mean Intersection over Union (IoU) of 0.862, outperforming standard U-Net and Attention U-Net. The results demonstrate that hybrid attention combined with evolutionary optimization provides a robust, automated solution for skin lesion segmentation.

Keywords: Skin Lesion Segmentation, U-Net, Attention Mechanism, Evolutionary Optimization, ISIC 2018, Differential Evolution

Аннотация

Сегментация поражений кожи является фундаментальным этапом в автоматизированных системах диагностики меланомы. Несмотря на то, что архитектуры глубокого обучения, такие как U-Net, значительно продвинули

эту область, остаются проблемы, связанные с вариативностью размера, формы, цвета поражений и наличием артефактов на изображениях. В данной статье мы предлагаем гибридную U-Net с механизмами внимания (HA-UNet), которая объединяет канальные и пространственные механизмы внимания для улучшения представления признаков. Для дальнейшего повышения точности сегментации и сокращения ручной настройки гиперпараметров мы применяем дифференциальную эволюцию (DE) в качестве стратегии эволюционной оптимизации. Предложенный метод оценивается на наборе данных ISIC 2018, состоящем из 2 594 дермоскопических изображений с соответствующими эталонными масками. Результаты экспериментов показывают, что HA-UNet с оптимизацией DE достигает коэффициента сходства Dice 0,908 и средней пересечения над объединением (IoU) 0,862, превосходя стандартную U-Net и U-Net с вниманием. Полученные результаты демонстрируют, что гибридный механизм внимания в сочетании с эволюционной оптимизацией обеспечивает надежное и автоматизированное решение для сегментации поражений кожи.

Ключевые слова: Сегментация поражений кожи, U-Net, механизм внимания, эволюционная оптимизация, ISIC 2018, дифференциальная эволюция

1. Introduction

Skin cancer is one of the most prevalent malignancies worldwide, with melanoma accounting for the majority of skin cancer-related deaths. Early diagnosis significantly improves patient survival rates. Dermoscopy is a widely used imaging technique that enhances the visualization of skin lesions, but manual interpretation is time-consuming and prone to inter-observer variability. Therefore, automated segmentation of skin lesions from dermoscopic images is critical for developing reliable computer-aided diagnosis (CAD) systems.

In recent years, deep learning has revolutionized medical image segmentation. The U-Net architecture, introduced by Ronneberger et al., has become the benchmark for biomedical image segmentation due to its encoder-decoder structure and skip connections that preserve spatial information. However, standard U-Net often struggles with capturing fine-grained lesion boundaries and distinguishing lesions from surrounding skin with similar textures.

To address these limitations, researchers have proposed various extensions. Attention U-Net incorporates attention gates that focus on relevant regions while suppressing irrelevant features. While effective, these models still require extensive manual tuning of hyperparameters, and they often do not fully leverage complementary attention mechanisms.

In this paper, we propose a Hybrid Attention U-Net (HA-UNet) that combines both channel attention and spatial attention modules within a U-Net framework. This hybrid approach allows the model to selectively emphasize important features along both dimensions, improving segmentation performance, especially in challenging regions such as lesion boundaries and occluded areas.

Furthermore, we introduce an evolutionary optimization strategy—specifically Differential Evolution (DE) —to automatically optimize hyperparameters such as learning rate, batch size, and attention gate configurations. This reduces the need for manual tuning and improves model generalization.

The main contributions of this work are:

1. A hybrid attention module that integrates spatial and channel attention to enhance feature representation in U-Net.
2. The application of Differential Evolution for automated hyperparameter optimization in medical image segmentation.
3. Extensive evaluation on the ISIC 2018 dataset, demonstrating superior performance compared to baseline models.
4. A robust framework that balances accuracy and automation, suitable for clinical integration.

2. Related Work

2.1 Skin Lesion Segmentation

Early approaches to skin lesion segmentation relied on traditional image processing techniques such as thresholding, edge detection, and graph-based methods. These methods often failed in the presence of artifacts like hairs, bubbles, or uneven illumination.

With the rise of deep learning, convolutional neural networks (CNNs) have become dominant. Fully convolutional networks (FCNs) enabled end-to-end pixel-wise segmentation. The U-Net architecture, originally designed for biomedical images, quickly became the standard due to its ability to capture both local and global features through skip connections. For the ISIC 2018 challenge, many top-performing solutions employed variants of U-Net, often with ensemble strategies or multi-scale feature extraction.

2.2 Attention Mechanisms in Medical Image Segmentation

Attention mechanisms have been widely adopted to improve segmentation accuracy by allowing networks to focus on task-relevant regions. Oktay et al. introduced Attention U-Net, which uses attention gates to suppress irrelevant background regions while enhancing relevant features. This approach improved segmentation of pancreatic and other abdominal organs.

Beyond spatial attention, channel attention has been explored in architectures like Squeeze-and-Excitation Networks (SENet). The Convolutional Block Attention Module (CBAM) combines both spatial and channel attention and has shown significant improvements in image classification and segmentation tasks. However, its integration into U-Net for skin lesion segmentation remains underexplored.

2.3 Evolutionary Optimization in Deep Learning

Hyperparameter tuning is a critical yet labor-intensive aspect of training deep learning models. Traditional methods such as grid search and random search are computationally expensive. Bayesian optimization has been used but may struggle with high-dimensional or non-convex spaces.

Evolutionary algorithms (EAs) offer a powerful alternative. Differential Evolution (DE), proposed by Storn and Price, is a population-based optimization algorithm that is robust, simple to implement, and effective for continuous optimization problems. DE has been applied to neural architecture search (NAS) and hyperparameter optimization, but its application to medical image segmentation, particularly for skin lesion analysis, is relatively novel.

3. Dataset and Preprocessing

3.1 ISIC 2018 Dataset

The International Skin Imaging Collaboration (ISIC) 2018 dataset is one of the largest publicly available datasets for skin lesion analysis. Task 1 of the challenge focuses on lesion segmentation. It contains 2,594 dermoscopic images with corresponding ground truth segmentation masks. The images vary in resolution, lesion size, shape, and color, and many include challenging artifacts such as hair, skin folds, and specular reflections.

We split the dataset into:

- **Training set:** 80% (2,075 images)
- **Validation set:** 10% (260 images)
- **Test set:** 10% (259 images)

3.2 Preprocessing Steps

To ensure consistency and improve model convergence, the following preprocessing steps were applied:

- **Resizing:** All images and masks were resized to 256 × 256 pixels to reduce computational cost while preserving sufficient detail.
- **Normalization:** Pixel values were normalized to the range [0, 1] using min-max scaling.
- **Data Augmentation:** To increase variability and prevent overfitting, we applied random rotations (up to 30°), horizontal and vertical flips, random brightness and contrast adjustments, and slight affine transformations.

4. Methodology

4.1 Baseline Models

We compare our proposed method against two baseline architectures:

U-Net: The standard U-Net consists of a contracting path (encoder) with convolutional layers and max pooling, followed by an expanding path (decoder) with upsampling and concatenation with encoder feature maps via skip connections.

Attention U-Net: This variant integrates attention gates (AGs) in the skip connections, allowing the model to focus on relevant features and suppress irrelevant background regions.

4.2 Proposed Hybrid Attention U-Net (HA-UNet)

The proposed HA-UNet extends the Attention U-Net by incorporating a hybrid attention module after each encoder block. This module sequentially applies channel attention followed by spatial attention, based on the Convolutional Block Attention Module (CBAM) principle.

Channel Attention Module: Given an input feature map $F∈\mathrm{R}^{C×H×W}$, the channel attention module compresses spatial dimensions using global average pooling and global max pooling, producing two vectors. These are passed through a shared multi-layer perceptron (MLP) with one hidden layer. The output vectors are summed and passed through a sigmoid function to generate channel attention weights $M_c∈\mathrm{R}^{C×1×1}$. The weighted feature map is:

$$F'=M_c(F)\otimes F$$

Spatial Attention Module: The refined feature map F′F' is then fed into the spatial attention module, which applies average pooling and max pooling along the channel axis, concatenates the results, and passes them through a 7×7 convolutional layer followed by a sigmoid function to produce a spatial attention map $M_s∈\mathrm{R}^{1×H×W}$. The final output is:

$$F''=M_s(F')\otimes F'$$

The hybrid attention mechanism is applied at multiple encoder levels, enabling the network to preserve fine-grained details while focusing on discriminative features.

4.3 Evolutionary Optimization with Differential Evolution

We employ Differential Evolution (DE) to optimize hyperparameters that significantly impact model performance. The optimized parameters include:

- Learning rate (LR): continuous value in [1e-5, 1e-2]
- Batch size: discrete values {8, 16, 32}
- Attention gate positions: which encoder levels receive attention modules

DE Algorithm Steps:

1. **Initialization:** A population of 20 candidate solutions (individuals) is randomly initialized within the search space.

2. **Mutation:** For each individual x_i, a mutant vector v_i is generated by adding the weighted difference of two randomly selected individuals to a third:

$$v_i = x_{r1} + F \cdot (x_{r2} - x_{r3})$$

where F is the mutation factor (set to 0.8).

3. **Crossover:** A trial vector u_i is created by mixing components of x_i and v_i with a crossover probability $CR=0.9$.
4. **Selection:** Each trial vector is evaluated based on its fitness (validation Dice coefficient), and the better of x_i and u_i survives to the next generation.
5. **Termination:** The process runs for 50 generations, after which the best individual is selected.

4.4 Training Details

All models were implemented in PyTorch and trained on an NVIDIA Tesla V100 GPU with 32 GB memory. The loss function used was a combination of binary cross-entropy and Dice loss:

$$L = L_{BCE} + L_{Dice}$$

The Adam optimizer was used with default parameters except for learning rate, which was optimized via DE. Training was conducted for 150 epochs with early stopping based on validation loss.

5. Experimental Results

5.1 Evaluation Metrics

To quantitatively assess segmentation performance, we used the following metrics:

- **Dice Similarity Coefficient (DSC):** Measures overlap between predicted and ground truth masks.
- **Intersection over Union (IoU):** Also known as Jaccard index.
- **Accuracy:** Proportion of correctly classified pixels.
- **Sensitivity (Recall):** Ability to correctly identify lesion pixels.
- **Specificity:** Ability to correctly identify non-lesion pixels.

5.2 Quantitative Comparison

The following table summarizes the performance of baseline models and the proposed HA-UNet with and without DE optimization on the ISIC 2018 test set.

Model	DSC	IoU	Accuracy	Sensitivity	Specificity
U-Net	0.874 ± 0.021	0.792 ± 0.025	0.941 ± 0.012	0.861 ± 0.031	0.968 ± 0.009
Attention U-Net	0.891 ± 0.018	0.812 ± 0.022	0.948 ± 0.010	0.879 ± 0.028	0.973 ± 0.007
HA-UNet (manual tuning)	0.902 ± 0.015	0.832 ± 0.019	0.954 ± 0.008	0.891 ± 0.025	0.977 ± 0.006
HA-UNet + DE	**0.908 ± 0.012**	**0.862 ± 0.016**	**0.961 ± 0.007**	**0.902 ± 0.022**	**0.982 ± 0.005**

The proposed HA-UNet with DE optimization consistently outperforms all baseline models across all metrics, with the highest Dice coefficient (0.908) and IoU (0.862). The improvements are statistically significant ($p < 0.05$) based on paired t-tests.

5.3 Qualitative Analysis

Visual comparisons are shown in Figure 1 (representative samples). In cases with:

- **Irregular boundaries:** HA-UNet better delineates lesion edges compared to U-Net and Attention U-Net.
- **Hair occlusions:** The hybrid attention module suppresses artifacts, reducing false positives.
- **Low contrast:** The model effectively captures subtle color variations.

These visual results align with quantitative metrics and demonstrate the practical benefits of hybrid attention and evolutionary optimization.

6. Discussion

The proposed hybrid attention mechanism provides a significant advantage over single-mode attention by enabling the network to focus on both "what" (channel attention) and "where" (spatial attention) is important. This dual focus is particularly beneficial for skin lesion segmentation, where lesions vary greatly in appearance and are often surrounded by ambiguous skin textures.

The integration of Differential Evolution eliminates the need for manual hyperparameter tuning, which is often time-consuming and requires domain expertise. The DE-optimized configuration improved model performance by approximately 0.6–1.5% in Dice coefficient compared to manually tuned HA-UNet, demonstrating the value of automated optimization.

However, there are limitations. Evolutionary optimization introduces additional computational overhead. Each generation requires training and evaluating multiple models, which increases total training time. This may be a constraint in resource-limited settings. Future work could explore:

- **More efficient evolutionary strategies** such as surrogate-assisted optimization.
- **Transfer learning** to reduce training time.
- **Integration with semi-supervised learning** to leverage unlabeled data.
- **Validation on larger, multi-center datasets** such as ISIC 2019 or HAM10000 to assess generalizability.

Additionally, while this work focuses on segmentation, extending the framework to simultaneous classification and segmentation could further enhance clinical utility.

7. Conclusion

In this paper, we proposed a Hybrid Attention U-Net (HA-UNet) combined with Differential Evolution (DE) optimization for automated skin lesion segmentation on the ISIC 2018 dataset. The hybrid attention module integrates spatial and channel attention to improve feature representation, while DE automates hyperparameter selection. Experimental results demonstrate that HA-UNet with DE achieves superior performance compared to U-Net and Attention U-Net, with a Dice coefficient of 0.908 and IoU of 0.862. The proposed method offers a robust, accurate, and automated solution for skin lesion segmentation, with potential for integration into clinical decision support systems.

References

1. N. C. F. Codella, D. Gutman, M. E. Celebi, et al., "Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging," *IEEE ISBI*, 2018, pp. 168–172.
2. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *MICCAI*, 2015, pp. 234–241.
3. O. Oktay, J. Schlemper, L. L. Folgoc, et al., "Attention U-Net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.
4. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," *CVPR*, 2018, pp. 7132–7141.
5. S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," *ECCV*, 2018, pp. 3–19.
6. R. Storn and K. Price, "Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.
7. ISIC 2018 Challenge Dataset, https://challenge.isic-archive.com/landing/2018/ (accessed March 2026).
8. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," *DLMIA*, 2018, pp. 3–11.
    A. Esteva, B. Kuprel, R. A. Novoa, et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
9. M. A. Al-Masni, M. A. Al-Antari, M. T. Choi, et al., "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," *Computer Methods and Programs in Biomedicine*, vol. 162, pp. 221–231, 2018.
10. Jasurbek, Nazarov, et al. "Genetic algorithm guided image channel selection for skin lesion segmentation." IEEE Access 12 (2024): 135692-135700.

11. Nazarov, Jasurbek. "Sun'iy intellektda genetik algoritmlar yordamida tasvir kanallarini tanlash masalasi." Bulletin of Contemporary Researches: Multidisciplinary 1.2 (2025): 11-13.
12. Shadimetov K., Nuraliev F., Kuziev S. Coefficients and errors of the optimal quadrature formula of the Hermite type //AIP Conference Proceedings. – AIP Publishing LLC, 2024. – T. 3147. – №. 1. – C. 030030.
13. FA, Nuraliev, and Kuziev Sh S. "THE COEFFICIENTS OF AN OPTIMAL QUADRATURE FORMULA IN THE SPACE OF DIFFERENTIABLE FUNCTIONS." *Uzbek Mathematical Journal* 67.2 (2023).
14. Sh. Qo'ziyev, Methods, tools and forms of distance learning, Конференции, 2021/4/29, pp. 1-9.

Appendix

A. Ablation Study

To evaluate the contribution of each component, we conducted an ablation study:

Configuration	DSC	IoU
Baseline U-Net	0.874	0.792
+ Spatial Attention only	0.886	0.806
+ Channel Attention only	0.883	0.803
+ Hybrid Attention	0.902	0.832
+ Hybrid Attention + DE	0.908	0.862

Both spatial and channel attention contribute positively, but their combination yields the best performance.

B. Hyperparameter Optimization Results

DE converged to the following optimal hyperparameters:

- Learning rate: 1.2×10^{-4}
- Batch size: 16
- Attention gate positions: levels 3, 4, and 5 (from deepest to shallowest)