

DIFFERENTIABLE NEURAL ARCHITECTURE SEARCH FOR OPTIMAL CHANNEL SELECTION IN DERMOSCOPIC SKIN LESION SEGMENTATION**Nazarov Jasurbek Ilhomjon o'g'li**

Kokand University

jasur.knu.ai@gmail.com**Kuziev Shaxobiddin Sobirovich**

Kokand University

<https://orcid.org/0000-0002-4117-4440>**Doi:** <https://doi.org/10.5281/zenodo.20235411>

Abstract. Accurate skin lesion segmentation from dermoscopic images is a fundamental prerequisite for early-stage melanoma detection. The diagnostic richness of intrinsic, shading-attenuated, and retinex-decomposed representations is overlooked by current deep learning techniques, which mostly rely on fixed RGB input channels. In this work, we offer a Differentiable Neural Architecture Search (DARTS) framework for a modified Transformer–CNN hybrid segmentation network (TMU-Net) that automatically selects the best input channel combinations based on gradients. The suggested method relaxes the channel-selection space to a continuous domain, allowing end-to-end joint optimization of both the segmentation network parameters and the channel weights, in contrast to discrete combinatorial search techniques like Genetic Algorithms. Tests on the ISIC 2016, ISIC 2017, and ISIC 2018 benchmark datasets show that DARTS-guided channel selection consistently performs better than GA-guided selection and fixed-channel baselines, with a Dice coefficient of up to 0.912 on ISIC 2018. These findings prove that differentiable architectural search is a scalable and efficient method for choosing input modalities in medical image segmentation.

Keywords. Skin lesion segmentation, dermoscopy, neural architecture search, channel selection, transformer-CNN, medical image analysis.

Аннотация. Точная сегментация поражений кожи на дерматоскопических изображениях является фундаментальным условием для обнаружения меланомы на ранней стадии. Современные методы глубокого обучения, в основном использующие фиксированные входные каналы RGB, игнорируют диагностическое богатство внутренних, ослабленных затенением и разложенных с помощью ретинекса представлений. В данной работе мы предлагаем структуру поиска дифференцируемой нейронной архитектуры (DARTS) для модифицированной гибридной сегментационной сети Transformer–CNN (TMU-Net), которая автоматически выбирает наилучшие комбинации входных каналов на основе градиентов. Предложенный метод сводит пространство выбора каналов к непрерывной области, позволяя осуществлять сквозную совместную оптимизацию как параметров сегментационной сети, так и весов каналов, в отличие от дискретных методов комбинаторного поиска, таких как генетические алгоритмы. Тесты на эталонных наборах данных ISIC 2016, ISIC 2017 и ISIC 2018 показывают, что выбор каналов с помощью DARTS стабильно превосходит выбор каналов с помощью генетического алгоритма и базовые методы с фиксированным количеством каналов, с коэффициентом Дайса до 0,912 на ISIC 2018. Эти результаты доказывают, что дифференцируемый архитектурный поиск является масштабируемым и эффективным методом выбора входных модальностей в сегментации медицинских изображений.

Ключевые слова: сегментация поражений кожи, дерматоскопия, поиск нейронной архитектуры, выбор каналов, трансформер-CNN, анализ медицинских изображений.

INTRODUCTION

One of the cancers that is growing the fastest in the globe is skin cancer, especially melanoma, and the most important factor in determining patient survival is early identification [1, 2]. Dermoscopy, which produces enlarged, polarized-light images that reveal subdermal morphological characteristics unseen to the unaided eye, has become the clinical gold standard for non-invasive skin lesion evaluation [3]. However, due to high intra-class variability, low contrast at lesion margins, the presence of visual artifacts like hair and air bubbles, and the inherent subjectivity of manual expert annotation, accurately defining lesion boundaries in dermoscopic images continues to be a difficult task [4, 5].

The field of medical image analysis has completely changed as a result of deep learning. In many segmentation benchmarks, Convolutional Neural Networks (CNNs), particularly the U-Net family of encoder-decoder architectures, have demonstrated state-of-the-art performance [6, 7]. More recently, the use of hybrid Transformer-CNN architectures and Vision Transformers (ViTs) has made it possible to simultaneously collect fine-grained local spatial detail and long-range global context, which is essential for lesion boundary delineation [8, 9]. This dual-pipeline concept is demonstrated by the Contextual Attention Network, or TMU-Net, which routes input data through CNN and parallel transformer branches before combining their complementary representations [10].

The choice of input image channels is a relatively unexplored aspect of dermoscopic segmentation. Intrinsic image decomposition isolates reflectance from shading components, resulting in representations that are more closely linked with genuine tissue chromophore distributions than standard RGB representation, which retains luminance-confounded color information [11, 12]. Previous research showed that segmentation accuracy is significantly improved by Genetic Algorithm (GA)-guided combinatorial search over a six-channel feature space, which includes raw RGB, grayscale, normalized retinex (R_n), intrinsic reflectance, and shading-attenuated pictures [13, 23, 24]. Nevertheless, GA-based search necessitates significant architecture re-training for each candidate chromosomal evaluation, is computationally costly, and is not differentiable.

By reformulating the discrete architectural search as a continuous bilevel optimization problem, Differentiable architectural Search (DARTS) overcomes these constraints and permits gradient-based joint optimization of network weights and architecture parameters in a single training loop [14, 15]. Although DARTS has been widely used in cell-level macro-architecture search, its application to input-channel modality selection in medical imaging is new and has not received much attention.

In this paper, we present a redesigned TMU-Net backbone that incorporates a DARTS-based channel selection mechanism. This work's main contributions are: (i) a bilevel optimization strategy that simultaneously trains channel importance weights and network segmentation parameters; (ii) a differentiable, softmax-parameterized channel mixing operator that learns optimal convex combinations of six dermoscopic input representations; and (iii) thorough evaluation across three ISIC benchmark datasets showing consistent superiority over both fixed RGB inputs and previous GA-guided selection.

MATERIAL AND METHODS

A. Datasets

Three publicly accessible dermoscopic image standards from the International Skin Imaging Collaboration (ISIC) were used for the experiments: ISIC 2016 (900 training/379 test pictures),

ISIC 2017 (2,000 training/600 test images), and ISIC 2018 (2,594 training/1,000 test images) [16, 17]. Every image was standardized to zero mean and unit variance after being shrunk to 224×224 pixels. To reduce overfitting, the training split was subjected to standard data augmentation techniques, such as random horizontal and vertical flipping, rotation ($\pm 30^\circ$), color jitter, and elastic deformation.

B. Multi-Channel Input Representations

Six distinct image representations were derived from each raw dermoscopic image and assembled as a multi-channel input tensor of dimension $6 \times H \times W$. These channels include: (1) RGB — the standard three-channel representation collapsed to a single luminance channel via average pooling; (2) Grayscale — the perceptual grayscale conversion using ITU-R BT.601 coefficients; (3) Normalized Retinex (R_n) — computed as the log-ratio between the image and a Gaussian-smoothed illumination estimate; (4) Intrinsic Reflectance — extracted via a single-image intrinsic decomposition algorithm that separates surface reflectance from illumination [11]; (5) Shading-attenuated — obtained by subtracting the estimated shading layer from the original image; and (6) V_c (Color Contrast) — a local color contrast map derived from the CIELAB color space. All channels were independently normalized to $[0, 1]$ prior to concatenation.

C. Modified TMU-Net Backbone

With modifications to accommodate variable-channel inputs, the segmentation backbone is based on the dual-pipeline TMU-Net design [10]. A bespoke multi-channel convolutional block whose first-layer kernel depth corresponds to the number of active input channels replaces the original ResNet-50 pre-trained stem. Gradient flow via the channel selection parameters is made possible by this adjustment, which is unrestricted by RGB-pretrained weight initialization. The dual pipeline uses a CNN branch that uses depthwise separable convolutions to extract local features and a transformer branch that uses multi-head self-attention to capture global spatial dependencies. Prior to being decoded into a binary segmentation mask, outputs from both branches are combined using learnt cross-attention gating.

D. Differentiable Channel Selection via DARTS

The proposed DARTS-based channel selector parameterizes the importance of each input channel with a learnable architecture weight vector $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_6] \in \mathbb{R}^6$. During the forward pass, the effective input to the backbone is computed as a soft weighted mixture:

$$\tilde{x} = \sum_i \text{softmax}(\alpha_i) \cdot x_i \quad \text{for } i \in \{\text{RGB, Grayscale, } R_n, \text{Intrinsic, Shading, } V_c\}$$

Architecture parameters α are optimized on the validation set:

$$\min_{\alpha} L_{\text{val}}(w^*(\alpha), \alpha)$$

Network weights w are optimized on the training set:

$$\min_w L_{\text{train}}(w, \alpha)$$

Bilevel optimization via alternating gradient descent (step size η_{α} , η_w)

This bilevel optimization is solved efficiently using the first-order DARTS approximation [14], wherein validation and training gradients are computed in alternating steps without unrolling the inner optimization. After convergence, a discrete channel mask is obtained by applying a threshold $\tau = 0.15$ to the final softmax-normalized architecture weights, retaining only channels with weight exceeding τ .

E. Loss Function and Training Protocol

The segmentation network is trained using a combined loss of binary cross-entropy and Dice loss with equal weighting: $L = 0.5 \cdot L_{\text{BCE}} + 0.5 \cdot L_{\text{Dice}}$. Training was performed for 200 epochs using the Adam optimizer with an initial learning rate of 1×10^{-4} for network weights and 3×10^{-4}

for architecture parameters, with cosine annealing scheduling. All experiments were implemented in PyTorch 2.1 and conducted on a single NVIDIA A100 GPU (40 GB).

Table 1. Quantitative comparison of channel selection strategies on three ISIC benchmarks.

Method	Dataset	Dice (%)	IoU (%)	Accuracy (%)	Sensitivity (%)
TMU-Net (RGB only)	ISIC 2016	87.4	79.3	95.1	84.6
GA-guided [13]	ISIC 2016	89.1	81.7	95.8	86.3
DARTS (Proposed)	ISIC 2016	90.6	83.2	96.4	88.1
TMU-Net (RGB only)	ISIC 2017	85.9	77.1	94.7	83.0
GA-guided [13]	ISIC 2017	87.8	79.5	95.3	85.2
DARTS (Proposed)	ISIC 2017	89.4	81.3	96.0	87.0
TMU-Net (RGB only)	ISIC 2018	88.2	80.1	95.5	85.7
GA-guided [13]	ISIC 2018	90.0	82.6	96.1	87.4
DARTS (Proposed)	ISIC 2018	91.2	84.3	96.9	89.2
Bold values indicate best performance per dataset. IoU = Intersection over Union.					

RESULTS

The regular TMU-Net with fixed RGB inputs and the previously published GA-guided channel selection method [13] served as the two baselines against which the suggested DARTS-based channel selection framework was assessed. Table 1 summarizes the quantitative results for each of the three ISIC benchmarks. The suggested approach outperformed the RGB-only baseline by 3.0%, 4.2%, and 1.4 percentage points on ISIC 2018, the biggest and most difficult benchmark, with a Dice coefficient of 91.2%, an IoU of 84.3%, and an overall accuracy of 96.9%. DARTS demonstrated the superiority of continuous gradient-based search over discrete evolutionary

optimization by achieving an extra gain of 1.2% Dice and 1.7% IoU when compared to GA-guided selection.

Consistent channel preferences were found in the learnt architectural weight distributions, which were averaged across five different runs. The intrinsic reflectance channel had the greatest mean weight ($\alpha = 0.31$), followed by normalized retinex ($\alpha = 0.24$) and shading-attenuated representations ($\alpha = 0.21$). When more decomposed representations are available, the raw grayscale channel consistently obtained the lowest weight ($\alpha = 0.07$), indicating negligible marginal contribution. Crucially, the computational efficiency of DARTS search was confirmed by the fact that it converged in about 35% less GPU-hours than a full GA search run (14.2 hours vs. 22.8 hours on ISIC 2018). Qualitative segmentation maps also show that DARTS-guided inputs result in boundary delineations that are crisper and more anatomically convincing, especially when low-contrast lesion peripheries and high hair artifact contamination are present. These are common failure modes for RGB-only models.

DISCUSSION

The study's findings support the theory that an important but frequently disregarded axis of model optimization in dermoscopic segmentation is input channel composition. The additional representational diversity provided by retinex-decomposed and intrinsic channels encodes clinically relevant lesion-background contrast information not captured by standard color imaging, as demonstrated by the consistent superiority of DARTS-selected multi-channel inputs over fixed RGB across three independent benchmarks. There are two main reasons why DARTS is better than GA-guided selection. First, while GA depends on a surrogate fitness function assessed on discrete, independently trained architectures, which introduces both approximation error and variance, differentiable optimization directly optimizes the segmentation objective with respect to channel importance weights. Second, the study's findings support the theory that an important but frequently disregarded axis of model optimization in dermoscopic segmentation is input channel composition. The additional representational diversity provided by retinex-decomposed and intrinsic channels encodes clinically relevant lesion-background contrast information not captured by standard color imaging, as demonstrated by the consistent superiority of DARTS-selected multi-channel inputs over fixed RGB across three independent benchmarks. There are two main reasons why DARTS is better than GA-guided selection. First, while GA depends on a surrogate fitness function assessed on discrete, independently trained architectures, which introduces both approximation error and variance, differentiable optimization directly optimizes the segmentation objective with respect to channel importance weights.

CONCLUSION

In order to solve the issue of choosing the best input channel for dermoscopic skin lesion segmentation, this research introduced a novel use of Differentiable Neural Architecture Search. The suggested framework automates and enhances the channel discovery process previously handled by Genetic Algorithm-guided combinatorial search by embedding a softmax-parameterized channel mixing operator within a modified TMU-Net backbone and training the entire system end-to-end via bilevel gradient optimization. In comparison to GA-based alternatives, the suggested DARTS-based method reduced search computational cost by about 35% while achieving state-of-the-art Dice coefficients of 90.6%, 89.4%, and 91.2% on the ISIC 2016, 2017, and 2018 benchmarks, respectively. Intrinsic reflectance and normalized retinex decompositions have the maximum diagnostic information density for lesion-background separation, according to the learnt channel weight distributions. These results are in line with the underlying biophysics of skin chromophore absorption. These findings demonstrate that differentiable architectural search is a viable, effective, and repeatable method for optimizing input modality in medical image segmentation. This capability naturally extends beyond dermoscopy to other multi-modal imaging

scenarios. Our goal is to help the medical imaging research community embrace and advance gradient-based architectural search by making the complete implementation code and pre-trained model weights available. Future research will examine cross-domain generalization and channel-conditioned uncertainty estimation for clinical imaging cohorts with more diverse acquisition and demographics.

REFERENCES

1. Siegel, R.L., Miller, K.D., Fuchs, H.E., and Jemal, A. (2022). Cancer statistics. *CA: A Cancer Journal for Clinicians*, 72(1), 7–33. <https://doi.org/10.3322/caac.21708>
2. Narayanan, D.L., Saladi, R.N., and Fox, J.L. (2010). Ultraviolet radiation and skin cancer. *International Journal of Dermatology*, 49(9), 978–986. <https://doi.org/10.1111/j.1365-4632.2010.04474.x>
3. Argenziano, G., Soyer, H.P., Chimenti, S., et al. (2003). Dermoscopy of pigmented skin lesions: results of a consensus meeting. *Journal of the American Academy of Dermatology*, 48(5), 679–693.
4. Mirikharaji, Z., Abhishek, K., Bissoto, A., et al. (2023). A survey on deep learning for skin lesion segmentation. *Medical Image Analysis*, 88, 102863. <https://doi.org/10.1016/j.media.2023.102863>
5. Hasan, M.K., Ahamad, M.A., Yap, C.H., and Yang, G. (2023). A survey, review, and future trends of skin lesion segmentation and classification. *Computers in Biology and Medicine*, 155, 106624. <https://doi.org/10.1016/j.compbiomed.2023.106624>
6. Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of MICCAI*, Lecture Notes in Computer Science, Vol. 9351, pp. 234–241.
7. Sarwar, N., Irshad, A., Naith, Q.H., Alsufiani, K.D., and Almalki, F.A. (2024). Skin lesion segmentation using deep learning algorithm with ant colony optimization. *BMC Medical Informatics and Decision Making*, 24(1), 265. <https://doi.org/10.1186/s12911-024-02665-4>
8. Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. In *Proceedings of ICLR 2021*.
9. Chen, J., Lu, Y., Yu, Q., et al. (2021). TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*.
10. Azad, R., Heidari, M., Wu, Y., and Merhof, D. (2022). Contextual Attention Network: Transformer meets U-Net. In *Proceedings of MICCAI Workshop on Machine Learning in Medical Imaging (MLMI)*, pp. 1–10. [TMU-Net]
11. Barron, J.T. and Malik, J. (2015). Shape, illumination, and reflectance from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(8), 1670–1687.
12. Hosny, K.M., Elshoura, D., Mohamed, E.R., Vrochidou, E., and Papakostas, G.A. (2023). Deep learning and optimization-based methods for skin lesion segmentation: A review. *IEEE Access*, 11, 85467–85488. <https://doi.org/10.1109/ACCESS.2023.3302581>
13. Jasurbek, N., Ivan, D. F., Ajani, O. S., & Mallipeddi, R. (2024). Genetic algorithm guided image channel selection for skin lesion segmentation. *IEEE Access*, 12, 135692–135700.
14. Liu, H., Simonyan, K., and Yang, Y. (2019). DARTS: Differentiable architecture search. In *Proceedings of the International Conference on Learning Representations (ICLR 2019)*.
15. Zela, A., Elsken, T., Saikia, T., Marrakchi, Y., Brox, T., and Hutter, F. (2020). Understanding and robustifying differentiable architecture search. In *Proceedings of ICLR 2020*.
16. Codella, N.C.F., Gutman, D., Celebi, M.E., et al. (2018). Skin lesion analysis toward melanoma detection: ISIC 2017 challenge. In *Proceedings of IEEE 15th International Symposium on Biomedical Imaging (ISBI)*, pp. 168–172.

17. Tschandl, P., Rosendahl, C., and Kittler, H. (2018). The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5, 180161. <https://doi.org/10.1038/sdata.2018.161>
18. Akram, S., et al. (2023). Segmentation and classification of skin lesions using hybrid deep learning method in the Internet of Medical Things. *Skin Research and Technology*, 29, e13524. <https://doi.org/10.1111/srt.13524>
19. Masood, H., Naseer, A., and Saeed, M. (2024). Optimized skin lesion segmentation: Analysing DeepLabV3+ and ASSP against generative AI-based deep learning approach. *Foundations of Science*. <https://doi.org/10.1007/s10699-024-09957-w>
20. Holland, J.H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor.
21. Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, MA.
22. Eshelman, L.J. and Schaffer, J.D. (1993). Real-coded genetic algorithms and interval schemata. *Foundations of Genetic Algorithms*, 2, 187–202.
23. Nazarov, J. (2025). Sun'iy intellektda genetik algoritmlar yordamida tasvir kanallarini tanlash masalasi. *Bulletin of Contemporary Researches: Multidisciplinary*, 1(2), 11-13.
24. Nazarov, J. (2026). ОПТИМИЗАЦИЯ ГИПЕРПАРАМЕТРОВ НЕЙРОННЫХ СЕТЕЙ С ИСПОЛЬЗОВАНИЕМ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ. *Eurasian Journal of Mathematical Theory and Computer Sciences*, 6(1), 24-32.