

# Lightweight Image Pre-processing Model for Improving Segmentation of Infrared Human Eye Images

Aleksei Samarin, Aleksei Toropov,  
Egor Kotenko, Artem Nazarenko,  
Elena Mikhailova, Valentin Malykh  
ITMO University

St. Petersburg, Russia

avsamarin@itmo.ru, toropov.ag@hotmail.com,  
kotenkoed@gmail.com, aanazarenko@itmo.ru,  
e.mikhailova@itmo.ru, valentin.malykh@phystech.edu

Alexander Savelev, Alexander Motyko  
St. Petersburg Electrotechnical University "LETI"  
St. Petersburg, Russia  
algsavelev@gmail.com, aamotyko@etu.ru

**Abstract**—This study focuses on improving iris and pupil segmentation in infrared images, a crucial task for gaze-tracking systems. We propose a lightweight image preprocessing approach, which ensures controlled data transformation, effectively eliminating parasitic reflections, noise, and contrast inconsistencies without introducing unwanted artifacts. Our approach employs analytically defined transformations, incorporating local and global filtering techniques to enhance the quality of input images while preserving physiologically relevant features. We compared deep learning-based segmentation models with and without preprocessing, demonstrating that the proposed approach significantly improves performance. Experimental results show notable gains in the mIoU metric, confirming increased robustness to illumination and image quality variations.

## I. INTRODUCTION

In recent years, medical research has increasingly gravitated toward automation, with a pronounced emphasis on leveraging statistical data analysis techniques and machine learning methodologies [1]–[4]. Among the diverse array of image processing challenges [5]–[7], the analysis of ocular structures stands out due to its inherent complexity and distinctive nature. The optical properties of the human eye exhibit a high degree of heterogeneity: the iris, pupil, and cornea possess varying coefficients of reflection and absorption, complicating the interpretation of visual data. In particular, infrared-based ocular imaging has emerged as an indispensable tool for assessing physiological parameters, underscoring the pivotal role of precisely segmenting eye structures in such images.

The accurate delineation of the iris and pupil boundaries in infrared images is of paramount importance across multiple domains, including healthcare, scientific research, and industrial applications. For instance, the efficacy of the psychophysiological assessment platform developed by NPP VIDEOMIX [8] is directly contingent upon the precision of eye structure segmentation within the infrared spectrum.

Ocular image segmentation methodologies can be broadly categorized into three principal groups. The first encompasses

classical approaches, which partition an image into regions based on disparities in intensity, texture, or spectral characteristics [9]–[15]. The second comprises solutions that harness the capabilities of deep neural networks, which exhibit remarkable efficacy in segmentation tasks by capturing intricate spatial dependencies [16]–[18]. The third category includes specialized algorithms explicitly designed for biomedical image analysis, employing models that have been fine-tuned to accommodate the idiosyncrasies of medical datasets [19]–[23].

Contemporary gaze-tracking systems infer gaze direction by analyzing the positional relationship between specular highlights and ocular anatomical features. However, ensuring high precision in such estimations necessitates robust segmentation techniques. Even marginal inaccuracies in boundary detection can introduce substantial deviations in gaze estimation, particularly in high-frame-rate tracking systems that demand precise temporal and spatial resolution.

One of the most pressing challenges in segmentation pertains to the presence of artifacts induced by variations in illumination, spurious reflections, and the intrinsic physiological characteristics of the eye. Such artifacts can severely distort the delineation of anatomical structures, thereby compromising the reliability of gaze modeling. This issue is particularly pronounced in infrared imaging, where improper handling of reflections and glare can lead to significant segmentation errors.

These limitations are especially evident in sophisticated systems such as MIX GT-19, where segmentation algorithms must demonstrate heightened adaptability and precision. Addressing these constraints necessitates the development of novel, task-specific segmentation strategies capable of accommodating the intricate conditions of infrared imaging. In this work, we introduce enhanced segmentation techniques tailored for iris and pupil detection, specifically designed to mitigate existing shortcomings and improve segmentation fidelity. The proposed advancements are aimed at augmenting the reliability of gaze-

tracking systems, thereby establishing a new benchmark for accuracy in this domain.

## II. RELATED WORK

Conventional methodologies for eye image segmentation predominantly rely on clustering-based algorithms, which facilitate the delineation of anatomical structures such as the iris and pupil by aggregating pixels with similar intensity, texture, and color properties [11]–[14]. Among these techniques, k-means clustering remains one of the most widely adopted due to its computational simplicity and efficiency. This method partitions an image into a predefined number of clusters, assigning pixels based on their Euclidean distance to randomly initialized centroids [10]–[12]. While this approach is computationally lightweight and well-suited for processing large datasets, its inherent rigidity—stemming from the requirement to specify a fixed number of clusters—reduces its adaptability when dealing with complex ocular structures.

To enhance the accuracy of k-means, preprocessing techniques such as Gabor filters, graph-based segmentation, and morphological operations are commonly integrated. However, this method remains highly sensitive to variations in image quality and structural heterogeneity. As an alternative, the mean-shift clustering algorithm [15] employs a density-based strategy, dynamically identifying regions of interest without the need to predefine the number of clusters. Despite its adaptability, the high computational burden associated with mean-shift often renders it impractical for real-time applications, particularly in resource-constrained environments.

The shortcomings of traditional segmentation techniques become particularly evident when processing images under non-uniform illumination, varying object scales, and diverse orientations. To address these challenges, deep learning-based approaches—particularly convolutional neural networks (CNNs)—have gained prominence, owing to their ability to extract complex spatial dependencies [16]–[18]. Early breakthroughs in CNN-driven eye segmentation were marked by models such as PupilNet [24], [25] and DeepIris [26], which leveraged convolutional layers for feature extraction and fully connected layers for segmentation map generation.

Recent advancements have led to the refinement of CNN architectures, with UNet [19], [20], [27], [28] emerging as a leading solution in biomedical image segmentation. By integrating an encoder-decoder framework alongside upsampling and downsampling operations, UNet achieves superior delineation of anatomical boundaries. Another noteworthy development is ENet [29], which prioritizes computational efficiency, making it particularly suitable for mobile and embedded applications. Further refinements of these architectures, such as MinENet, EyeNet [30], and EyeMMS [31], have demonstrated exceptional performance (mIoU over 0.92) on benchmark datasets like OpenEDS [32].

Despite these advancements, existing CNN-based models continue to face significant limitations when applied to infrared eye image segmentation. The primary obstacles include specular reflections, variations in corneal reflectance, and

the physiological uniqueness of each individual’s eye. These factors pose considerable challenges in high-precision gaze-tracking systems, such as MIX GT-19, where segmentation accuracy is paramount to system reliability.

An in-depth analysis of current solutions reveals that mainstream CNN architectures struggle to generalize to infrared imaging conditions, as confirmed by evaluations on InP [33], a dataset developed in collaboration with NPP VIDEOMIX. Existing methodologies, including those proposed in [34], fail to achieve the required segmentation fidelity, underscoring the necessity for more domain-specific algorithmic adaptations.

One of the most promising avenues for improvement lies in infrared image preprocessing, incorporating localized contrast enhancement, suppression of parasitic reflections, and stabilization of anatomical contours. However, given the computational constraints of medical hardware, these preprocessing techniques must remain computationally efficient while minimizing the burden on processing resources.

Furthermore, the application of end-to-end generative neural networks in this domain introduces substantial risks. Since such models operate within tensor-based latent representations, they inherently lack strict control over generated outputs, potentially introducing artificial artifacts that do not correspond to the actual physiological structure of the eye. This issue is particularly problematic for gaze-tracking applications, where data fidelity is a critical requirement.

Thus, achieving high-accuracy segmentation for infrared eye imaging necessitates the development of custom-tailored preprocessing techniques that effectively mitigate artifacts, precisely delineate the iris and pupil, and retain computational efficiency. Optimizing such algorithms is pivotal for enhancing the robustness of gaze-tracking systems, ensuring stable performance even in challenging infrared imaging conditions. In this study, we introduce an adapted preprocessing approach specifically designed to address these challenges, offering an optimal solution for infrared eye image segmentation.

## III. PROPOSED SOLUTION

We introduce an adapted approach to eye image preprocessing based on the UniFi methodology [35], [36], which ensures strict control over the data transformation process and guarantees the absence of unwanted artifacts. Unlike generative neural network-based methods, the UniFi approach employs analytically defined transformations that combine local and global filtering techniques. These filters perform contrast correction, suppress noise, and eliminate parasitic reflections while fully preserving the physical integrity of the original image. Due to the deterministic nature of these operations, the risk of uncontrolled modifications in the image structure is completely mitigated, making this method particularly effective for iris and pupil segmentation in the infrared spectrum.

Beyond presenting a lightweight infrared image preprocessing model, this study conducts an in-depth evaluation of its effectiveness when integrated into existing eye segmentation solutions. Specifically, a comparative analysis was carried out,

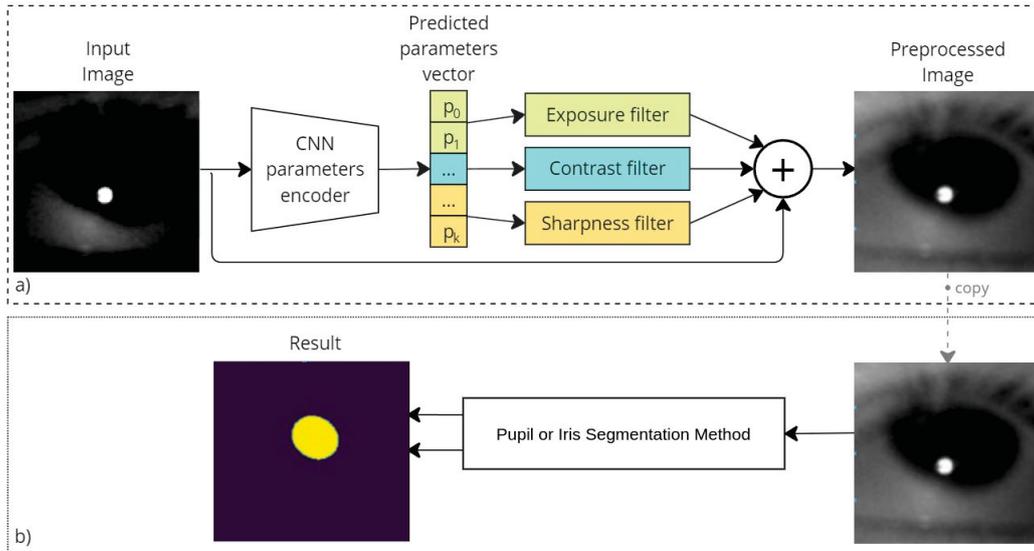


Fig. 1. Overall scheme of the proposed approach. a) Our proposed image pre-processing module; b) Segmentation model part of the pipeline.

where segmentation algorithms were tested in two scenarios: in their original form and with the addition of UniFi-based preprocessing. The obtained results indicate that incorporating the proposed module significantly enhances segmentation accuracy, reduces artifact interference, and improves the robustness of algorithms under varying illumination conditions and noise levels.

Further details on the experimental setup, evaluation methodology, and dataset characteristics are thoroughly discussed in the Evaluation section.

#### A. Method Overview

The proposed approach is structured as a multi-stage composite algorithm, integrating an image preprocessing module and one of the segmentation models for ocular structures. This modular architecture provides adaptability across various infrared image processing scenarios, ensuring both flexibility and scalability.

A fundamental component of the system is the preprocessing module, which is responsible for noise filtration, suppression of parasitic reflections, and localized contrast normalization. This stage is critical for enhancing the quality of input data by mitigating artifacts that could otherwise deteriorate the accuracy of subsequent segmentation. A detailed breakdown of the employed filtering techniques and their application principles is provided in Section III-B.

The second core element of the approach is the segmentation model, designed to precisely delineate the anatomical structures of the eye, including the iris and pupil, in infrared imagery. This selection allowed for a comprehensive assessment of the robustness of our approach and its adaptability across diverse architectural paradigms. The baseline models utilized for experimental validation are thoroughly described in Section IV-B.

The complete processing pipeline is illustrated in Figure 1, which outlines the sequential data transformation flow, starting from image preprocessing and culminating in the segmentation stage.

#### B. Image Pre-processing Model

The pre-processing framework outlined in [36] served as the foundation for our method, with several targeted modifications introduced to enhance its applicability. The structural representation of the corrective transformation process is formalized as follows:

$$I_e = I_o + \sum_{i=1}^n f_i(I_o, h_i(I_{so})).$$

This formulation encapsulates a modular configuration, where the total number of computational blocks corresponds to the number of employed filters. Each block  $i$  processes a scaled version of the source image  $I_{so}$  through a parameter generator  $h_i$ , which extracts a parameter set  $p_i$  specific to the corresponding filter function  $f_i$ . The resulting transformation outputs are then applied independently to the original image  $I_o$ , and the final enhanced image is obtained by summing the original image with the cumulative filter outputs.

The parameter generation module adheres to the structural principles of LFIEM [36], adopting a lightweight convolutional architecture. This generator comprises two primary stages: the first stage consists of three convolutional layers with a stride of 2, followed by batch normalization and a LeakyReLU activation function (except for the initial layer). The number of feature maps is progressively increased, with 16, 32, and 128 filters for the first, second, and third layers, respectively. The second stage incorporates two fully connected layers, separated by ReLU activation, which refine the extracted feature representations. At the final stage, an

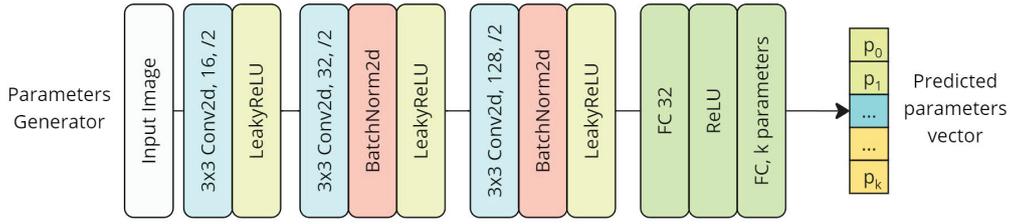


Fig. 2. Parameters generator architecture.

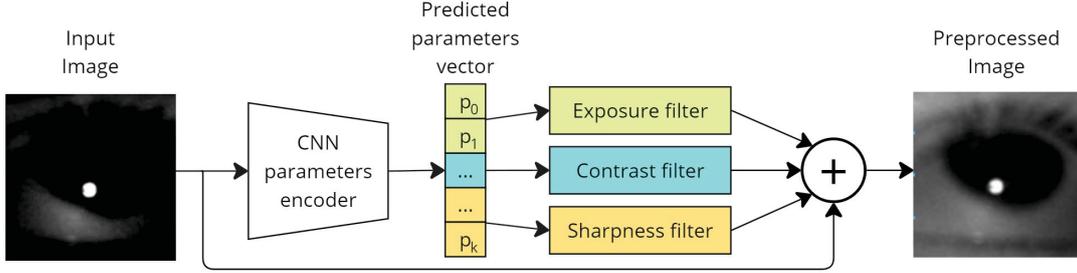


Fig. 3. Image pre-processing module structure.

activation function is applied based on the required output constraints: a sigmoid function for values within the range  $[0,1]$ , a hyperbolic tangent function for values in  $[-1,1]$ , and no activation when constraint-free filtering is needed. The complete architecture of the parameter generator is visualized in Fig. 2. Notably, despite its advanced design, the module remains computationally lightweight, comprising only 47,000 trainable parameters.

Addressing prevalent challenges in microscopic scene images, such as blur artifacts, insufficient contrast, and low sharpness, necessitated the implementation of corrective transformation filters.

One of the essential transformations is the *sharp* filter, mathematically formulated as follows:

$$I_{out} = I_{in} \otimes \frac{1}{\nu} (K + M \cdot q),$$

where  $K$  denotes the convolution kernel,  $M$  represents a corresponding map matrix of identical dimensions, and  $\nu$  is a normalization coefficient computed as the summation of all elements within  $(K + M \cdot q)$ . This transformation is applied independently to the red, green, and blue channels, each parameterized by a distinct trainable variable. The filter matrix configurations are explicitly defined as follows:

$$K = \begin{pmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & -476 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{pmatrix},$$

$$M = \begin{pmatrix} 0.8 & 0.8 & 0.8 & 0.8 & 0.8 \\ 0.8 & 0.9 & 0.9 & 0.9 & 0.8 \\ 0.8 & 0.9 & 1 & 0.9 & 0.8 \\ 0.8 & 0.9 & 0.9 & 0.9 & 0.8 \\ 0.8 & 0.8 & 0.8 & 0.8 & 0.8 \end{pmatrix}.$$

Another fundamental transformation is automatic *contrast* correction, dynamically adjusted by the parameter  $p \in [-1, 1]$ . This parameter dictates the modification applied to each pixel of the input image, leading to the following computational transformation:

$$I_{out}[x, y] = \begin{cases} (I_{in}[x, y] - 0.5) \cdot \frac{1}{1-r}, & \text{if } r > 0 \\ (I_{in}[x, y] - 0.5) \cdot (1-r), & \text{otherwise;} \end{cases}$$

Given the substantial variation in lighting conditions commonly encountered in microscopic imaging, exposure correction is a critical step in the preprocessing pipeline. The corresponding automatic *exposure* adjustment transformation is given by:

$$I_{out}[x, y] = I_{in}[x, y] \cdot 2^t.$$

To optimize computational efficiency and minimize redundant operations, we consolidated all predictor functions for the parameter estimation process into a unified neural network encoder. This architectural refinement significantly reduces processing overhead while maintaining the integrity of the corrective transformations.

The structural organization of the pre-processing module is comprehensively illustrated in Fig. 3.

#### IV. EVALUATION

In this study, we leveraged the InP dataset [33] to assess the outcomes of our experiments systematically. A compre-

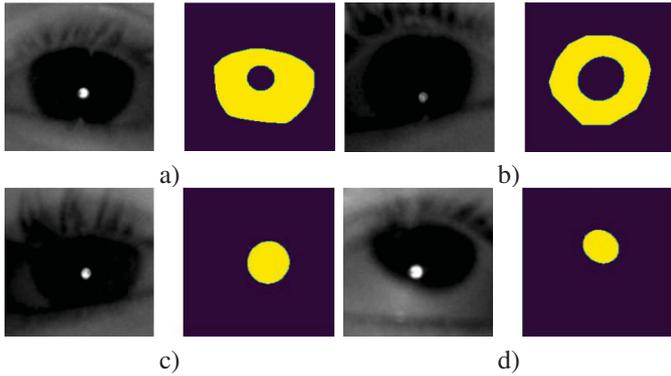


Fig. 4. InP dataset [33] illustration: a) iris and mask; b) iris and mask; c) pupil and mask; d) pupil and mask.

hensive comparative analysis was conducted, enabling a rigorous evaluation of the proposed approaches utilizing different UniFi filters, both relative to each other and against existing solutions that do not incorporate the lightweight preprocessing model. This examination facilitated a stringent validation of the proposed methodologies, providing empirical evidence of their advantages over alternative techniques. By capitalizing on the diversity and representativeness of the dataset, we ensured a holistic performance assessment, reinforcing the practical applicability and real-world feasibility of our developed techniques.

#### A. Dataset

We utilized the InP dataset [33], a publicly available collection of annotated infrared eye images, developed in collaboration with NPP VIDEOMIX. The dataset comprises structured pairs of images, where each instance consists of an infrared capture of a human eye alongside its corresponding binary segmentation mask, delineating key anatomical components—namely, the iris and pupil. Representative examples of these labeled data pairs are provided in Fig. 4.

The dataset is structured into two distinct subsets, each tailored to a specific segmentation task. The first subset comprises 1,758 images dedicated to iris segmentation. These images were obtained using an infrared imaging system, capturing data from 8 individual subjects, with an approximate mean of 439 samples per participant. Every image in this subset is paired with a corresponding ground-truth segmentation mask, ensuring precise delineation of the iris region.

The second subset focuses on pupil segmentation and contains a total of 7,343 images. This portion of the dataset was collected from 21 participants, with an average of 699 images per subject. Analogous to the iris segmentation dataset, each image in this section is accompanied by a manually annotated segmentation mask, allowing for accurate model training and evaluation.

For rigorous performance assessment and to enhance the generalizability of the models, we employed a systematic partitioning strategy. The dataset was split into training and testing subsets using a 5:1 ratio, ensuring an optimal balance between model learning capacity and validation robustness.

This methodical division facilitates effective training while retaining a sufficient volume of unseen data for benchmarking and performance validation.

#### B. Baselines

We conducted a comprehensive evaluation of the performance of standard image segmentation methods by integrating our lightweight image preprocessing model. To achieve this, we selected representative models from each primary category: specialized models and general biomedical segmentation models. This comparative analysis allows us to assess the impact of our preprocessing module on the segmentation accuracy of infrared images of the human eye, as well as to identify the strengths and limitations of our approach when applied alongside various segmentation baselines.

For the first category of models, we employed DeepIris. For the second category, we utilized the Unet and Unet++ models.

All the aforementioned models have demonstrated robust performance in image segmentation tasks. To ensure a fair comparison in terms of the number of trainable parameters, we adopted a backbone feature extractor for both UNet-based segmentation models and keypoint-based segmentation architectures. This strategy guarantees a balanced evaluation of our preprocessing model's capabilities, mitigating potential disparities arising from differences in feature extraction complexity.

#### C. Training Details

We trained all models on  $1 \times$  NVIDIA RTX 3090 GPU for up to 100 epochs. The AdamW optimizer [?] was used with  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ , and a weight decay rate of 0.05. The batch size was set to 8. The initial learning rate was  $10^{-8}$ . Each model underwent 10 training cycles to ensure the robustness and consistency of the testing results.

#### D. Experimental results

To evaluate the effectiveness of our models, we employed a set of metrics widely used in image segmentation tasks. Specifically, we utilized the mean Intersection over Union (mIoU) metric, computed separately for pupil and iris segmentation, which provides an objective and comprehensive assessment of segmentation accuracy. The complete experimental results are presented in Table I.

For baseline comparisons, we selected neural network architectures that are compatible with integration into the MIX GT-19 system. Our analysis included various UNet-based models, encompassing both general architectural enhancements and task-specific adaptations tailored for eye segmentation.

Among the generalized improvements, we evaluated UNet++ [20], which incorporates additional multi-scale feature fusion layers, as well as UNet variants employing different encoder backbones, such as ResNets [37] and EfficientNets [38]. Furthermore, we assessed specialized pupil segmentation techniques optimized for seamless integration into the MIX GT-19 processing pipeline.

The obtained results validate the effectiveness of the proposed approaches, which outperformed all existing solutions

TABLE I. COMPARATIVE ANALYSIS

Filters configuration	Segmentation model	mIoU Iris	mIoU Pupil
w/o preprocessing model	DeepIris	0.799	0.793
w/o preprocessing model	ENet	0.828	0.820
w/o preprocessing model	MinENet	0.837	0.829
w/o preprocessing model	EyeNet	0.855	0.859
w/o preprocessing model	EyeMMS	0.880	0.887
w/o preprocessing model	Unet ResNet-18	0.805	0.798
w/o preprocessing model	Unet ResNet-50	0.820	0.815
w/o preprocessing model	Unet EfficientNet-b4	0.839	0.823
w/o preprocessing model	Unet++ EfficientNet-b4	0.867	0.861
Exposure + Contrast	DeepIris	0.801	0.795
Exposure + Contrast	ENet	0.830	0.825
Exposure + Contrast	MinENet	0.841	0.832
Exposure + Contrast	EyeNet	0.859	0.864
Exposure + Contrast	EyeMMS	0.884	0.891
Exposure + Contrast	Unet ResNet-18	0.807	0.802
Exposure + Contrast	Unet ResNet-50	0.823	0.819
Exposure + Contrast	Unet EfficientNet-b4	0.837	0.826
Exposure + Contrast	Unet++ EfficientNet-b4	0.870	0.865
Exposure + Sharpness	DeepIris	0.810	0.809
Exposure + Sharpness	ENet	0.840	0.831
Exposure + Sharpness	MinENet	0.848	0.839
Exposure + Sharpness	EyeNet	0.866	0.875
Exposure + Sharpness	EyeMMS	0.893	0.900
Exposure + Sharpness	Unet ResNet-18	0.812	0.805
Exposure + Sharpness	Unet ResNet-50	0.828	0.822
Exposure + Sharpness	Unet EfficientNet-b4	0.845	0.832
Exposure + Sharpness	Unet++ EfficientNet-b4	0.879	0.872

examined in our experiments. Notably, while the differences in metric values may appear incremental at first glance, they have a substantial impact in the context of gaze-tracking systems. For systems processing high-frequency image streams from multiple cameras, even marginal improvements in segmentation precision contribute to significantly enhanced stability and reliability. These findings underscore the practical significance of the proposed methods and demonstrate their applicability in real-world infrared eye-tracking scenarios.

## V. CONCLUSION

This study addressed the segmentation of the iris and pupil in infrared images, a critical component in gaze-tracking systems. We introduced an adapted image preprocessing approach based on UniFi, ensuring controlled data transformations, eliminating undesirable artifacts, and preserving essential image features. Unlike generative neural network-based approaches, the proposed method employs analytically defined transformations, incorporating both local and global filtering techniques to perform contrast correction, noise suppression, and parasitic reflection removal. Furthermore, our approach guarantees the absence of newly generated artifacts, common in generative models.

A comprehensive evaluation, including comparative testing of segmentation models with and without UniFi preprocessing, demonstrated that the proposed approach significantly enhances segmentation quality. Specifically, the integration of UniFi preprocessing led to higher segmentation accuracy, reduced sensitivity to noise, and improved robustness to illumination variations. The effectiveness of this approach

is confirmed by mIoU metrics, which indicate a consistent improvement in performance when incorporating the UniFi module into the infrared eye image processing pipeline.

Thus, this work not only validates the effectiveness of the proposed preprocessing method but also establishes a foundation for further advancements in eye segmentation systems. Future research may focus on further optimization of filtering algorithms as well as adapting the proposed approach to other medical and industrial image analysis tasks, expanding its applicability beyond infrared eye segmentation.

## ACKNOWLEDGMENT

The research was carried out with the financial support of the ITMO University Research Projects in AI Initiative (project No. 640113).

## REFERENCES

- [1] N. Obukhova, A. Motyko, B. Timofeev, and A. Pozdeev, "Method of endoscopic images analysis for automatic bleeding detection and segmentation," in *2019 24th Conference of Open Innovations Association (FRUCT)*. IEEE, 2019, pp. 285–290.
- [2] A. Samarin, A. Savelev, A. Toropov, A. Dzestelova, V. Malykh, E. Mikhailova, and A. Motyko, "One-staged attention-based neoplasms recognition method for single-channel monochrome computer tomography snapshots," *Pattern Recognition and Image Analysis*, vol. 32, no. 3, pp. 645–650, 2022.
- [3] N. A. Obukhova, A. A. Motyko, and A. A. Pozdeev, "Two-stage method for polyps segmentation in endoscopic images," *Computer Vision in Control Systems—6: Advances in Practical Applications*, pp. 93–106, 2020.
- [4] A. Samarin, A. Savelev, and V. Malykh, "Two-staged self-attention based neural model for lung cancer recognition," in *2020 Science and Artificial Intelligence conference (SAI ence)*. IEEE, 2020, pp. 50–53.

- [5] A. Samarin, A. Savelev, A. Toropov, A. Dzestelova, V. Malykh, E. Mikhailova, and A. A. Motyko, "One-stage classifiers based on u-net and autoencoder with attention for recognition of neoplasms from single-channel monochrome computed tomography images," *Pattern Recognition and Image Analysis*, vol. 33, pp. 132–138, 2023. [Online]. Available: <https://api.semanticscholar.org/CorpusID:259310593>
- [6] A. Samarin, A. Savelev, A. Toropov, A. Dzestelova, V. Malykh, E. Mikhailova, and A. Motyko, *Prior Segmentation and Attention Based Approach to Neoplasms Recognition by Single-Channel Monochrome Computer Tomography Snapshots*, 08 2023, pp. 561–570.
- [7] N. A. Obukhova, A. A. Motyko, U. Kang, S.-J. Bae, and D.-S. Lee, "Automated image analysis in multispectral system for cervical cancer diagnostic," in *2017 20th conference of open innovations association (FRUCT)*. IEEE, 2017, pp. 345–351.
- [8] "Npp videomix." [Online]. Available: <https://v-mix.ru>
- [9] D. J. Bora and A. K. Gupta, "Clustering approach towards image segmentation: An analytical study," *CoRR*, vol. abs/1407.8121, 2014. [Online]. Available: <http://arxiv.org/abs/1407.8121>
- [10] L. Jin, F. Xiao, and W. Haopeng, "Iris image segmentation based on k-means cluster," in *2010 IEEE International Conference on Intelligent Computing and Intelligent Systems*, vol. 3, 2010, pp. 194–198.
- [11] A. Premana, A. Wijaya, and M. Soeleman, "Image segmentation using gabor filter and k-means clustering method," 10 2017, pp. 95–99.
- [12] H. Shi and W.-L. Lee, "Image segmentation using k-means clustering, gabor filter and moving mesh method," *The Imaging Science Journal*, vol. 69, no. 5-8, pp. 407–416, 2021.
- [13] S. Jardim, J. António, and C. Mora, "Graphical image region extraction with k-means clustering and watershed," *Journal of Imaging*, vol. 8, no. 6, p. 163, 2022.
- [14] M. Y. Choong, W. L. Khong, W. Y. Kow, L. Angeline, and K. T. K. Teo, "Graph-based image segmentation using k-means clustering and normalised cuts," in *2012 Fourth International Conference on Computational Intelligence, Communication Systems and Networks*. IEEE, 2012, pp. 307–312.
- [15] N. J. Gandhi, V. J. Shah, and R. Kshirsagar, "Mean shift technique for image segmentation and modified canny edge detection algorithm for circle detection," in *2014 International Conference on Communication and Signal Processing*, 2014, pp. 246–250.
- [16] W. Gu, S. Bai, and L. Kong, "A review on 2d instance segmentation based on deep neural networks," *Image and Vision Computing*, vol. 120, p. 104401, 2022.
- [17] A. M. Hafiz and G. M. Bhat, "A survey on instance segmentation: state of the art," *International journal of multimedia information retrieval*, vol. 9, no. 3, pp. 171–189, 2020.
- [18] G. Park, C. Han, W. Yoon, and D. Kim, "Mhsan: multi-head self-attention network for visual semantic embedding," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 1518–1526.
- [19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015. [Online]. Available: <http://arxiv.org/abs/1505.04597>
- [20] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," *CoRR*, vol. abs/1807.10165, 2018. [Online]. Available: <http://arxiv.org/abs/1807.10165>
- [21] F. Isensee, J. Petersen, S. A. A. Kohl, P. F. Jäger, and K. H. Maier-Hein, "nnu-net: Breaking the spell on successful medical image segmentation," *CoRR*, vol. abs/1904.08128, 2019. [Online]. Available: <http://arxiv.org/abs/1904.08128>
- [22] S. Pathan and A. Tripathi, "Y-net: Biomedical image segmentation and clustering," 2020.
- [23] F. Milletari, N. Navab, and S. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," *CoRR*, vol. abs/1606.04797, 2016. [Online]. Available: <http://arxiv.org/abs/1606.04797>
- [24] W. Fuhl, T. Santini, G. Kasneci, and E. Kasneci, "Pupilnet: Convolutional neural networks for robust pupil detection," *arXiv preprint arXiv:1601.04902*, 2016.
- [25] W. Fuhl, T. Santini, G. Kasneci, W. Rosenstiel, and E. Kasneci, "Pupilnet v2. 0: Convolutional neural networks for cpu based real time robust pupil detection," *arXiv preprint arXiv:1711.00112*, 2017.
- [26] A. Gangwar and A. Joshi, "Deepirisnet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition," in *2016 IEEE international conference on image processing (ICIP)*. IEEE, 2016, pp. 2301–2305.
- [27] J. Wei, H. Huang, M. Sun, R. He, and Z. Sun, "Toward accurate and reliable iris segmentation using uncertainty learning," *ArXiv*, vol. abs/2110.10334, 2021. [Online]. Available: <https://api.semanticscholar.org/CorpusID:239050050>
- [28] C. Wang, J. Muhammad, Y. Wang, Z. He, and Z. Sun, "Towards complete and accurate iris segmentation using deep multi-task attention network for non-cooperative iris recognition," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2944–2959, 2020.
- [29] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "Enet: A deep neural network architecture for real-time semantic segmentation," *arXiv preprint arXiv:1606.02147*, 2016.
- [30] P. Kansal and S. Devanathan, "Eyenet: Attention based convolutional encoder-decoder network for eye region segmentation," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. IEEE, 2019, pp. 3688–3693.
- [31] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "Eye-mms: Miniature multi-scale segmentation network of key eye-regions in embedded applications," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0.
- [32] S. J. Garbin, Y. Shen, I. Schuetz, R. Cavin, G. Hughes, and S. S. Talathi, "Openeds: Open eye dataset," *arXiv preprint arXiv:1905.03702*, 2019.
- [33] "Iris and pupil dataset (inp)." [Online]. Available: <https://github.com/itmo-cv-lab/eye-tracking-dataset>
- [34] A. Valenzuela, C. Arellano, and J. Tapia Farias, "Towards an efficient segmentation algorithm for near-infrared eyes images," *IEEE Access*, vol. 8, pp. 171 598–171 607, 01 2020.
- [35] A. Samarin, A. Nazarenko, A. Savelev, A. Toropov, A. Dzestelova, E. Mikhailova, A. Motyko, and V. Malykh, "A model based on universal filters for image color correction," *Pattern Recognit. Image Anal.*, vol. 34, no. 3, p. 844–854, Sep. 2024. [Online]. Available: <https://doi.org/10.1134/S1054661824700731>
- [36] O. Tatanov and A. Samarin, "Lfiem: Lightweight filter-based image enhancement model," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 873–878.
- [37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2015. [Online]. Available: <https://api.semanticscholar.org/CorpusID:206594692>
- [38] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *ArXiv*, vol. abs/1905.11946, 2019. [Online]. Available: <https://api.semanticscholar.org/CorpusID:167217261>