

Training Set Design for Uneven Illumination Correction in High-Resolution Whole Slide Images

Sama Nemati (MSc)¹, Hasti Shabani (PhD)^{1*}, Ahmad Mahmoudi-Aznavah (PhD)²

¹Institute of Medical Science and Technology, Shahid Beheshti University, Tehran, Iran

²Cyberspace Research Institute, Shahid Beheshti University, Tehran, Iran

ABSTRACT

Uneven illumination correction is considered a critical pre-processing step in creating digital images from optical microscopes, particularly in whole-slide imaging (WSI). While deep learning-based methods have suggested new possibilities, they often struggle with generalizing to unseen images and require substantial computational resources. The most common approach for training deep neural networks in this field relies on patch-based processing, which may overlook the global illumination distribution, leading to inconsistencies in correction. This study aimed to identify a key limitation in deep learning models for uneven illumination correction, highlighting the importance of preserving the original image resolution and incorporating a global view of illumination patterns to enhance generalization. To address this, we proposed a new training set design strategy that optimizes neural network performance while utilizing computational resources effectively. Our approach ensures a more uniform correction across entire WSI slides, reducing artifacts and improving image consistency. The proposed strategy enhances model robustness and scalability, making deep learning-based illumination correction more practical for clinical and research applications.

Citation: Nemati S, Shabani H, Mahmoudi-Aznavah A. Training Set Design for Uneven Illumination Correction in High-Resolution Whole Slide Images. *J Biomed Phys Eng.* 2025;15(3):291-298. doi: 10.31661/jbpe.v0i0.2502-1890.

Keyword

Uneven Illumination Correction; Artifacts; Pathology; Deep Learning; Training Set

Introduction

Whole slide imaging (WSI) provides a digital high-resolution image from a sample covering the whole field of view (FOV) of the entire slide [1]. This technique eliminates manually scanning biological samples and offers specialists advanced image-based analysis and interpretation. In this regard, different image processing techniques, such as illumination correction and stitching, are required to put the captured images (tiles) next to each other to form a reliable image of the entire tissue [2]. Optical microscope modalities, such as bright-field, dark-field, and fluorescence often exhibit uneven illumination, known as vignetting, which occurs due to non-uniform illumination distribution, lens imperfections, or variations in sample thickness [3]. Uneven illumination results in an image with a falloff in intensity from the center of the radiation to the borders, which introduces a black plaid pattern on the virtual slide and affects the accuracy and reliability of subsequent analysis tasks [1]. Therefore, correcting uneven illumination is a crucial pre-processing step to enhance the quality and consistency of images captured by a WSI scanner [4]. Deep learning methods, which have significantly advanced

*Corresponding author:
Hasti Shabani
Institute of Medical Science and Technology,
Shahid Beheshti University, Tehran, Iran
E-mail:
ha_shabani@sbu.ac.ir

Received: 19 February 2025
Accepted: 12 May 2025

various domains of image processing, have also shown promising results in addressing uneven illumination effects. Compared to analytical approaches, an effective deep neural network offers advantages such as the processing of color images without the need to process each channel separately, automated feature extraction, and real-time correction. However, brightness enhancement has been extensively studied in computer vision, the approaches are not directly applicable to WSI due to the following reasons: 1) computational complexity: images from the WSI technique are large and high-resolution, often consisting of gigapixel-sized images, and many computer vision algorithms may not be scalable or efficient enough to handle the computational demands of WSI, 2) homogeneity preservation: inconsistent illumination or appearance among tiles introduces errors in the stitching step and further analysis, and 3) robustness and generalization: computer vision algorithms for uneven illumination correction are often developed and evaluated on specific datasets and imaging conditions, however, the WSI encompasses a wide range of uneven illumination patterns [4,5].

In 2016, Mei et al. proposed a fully convolutional network to eliminate the uneven illumination of dermoscopy tiles [6]. They used convolutional layers, ReLU activation, and pooling layers to extract information at different scales, merging them with skip connections and using Euclidean distance as a loss function. The proposed method utilized 1000 high-quality dermoscopy tiles without uneven illumination patterns as ground truth images and a simulation method to provide more data based on 4 synthetic uneven illumination patterns. Although the illumination pattern was relatively well corrected in the output images of this network, a severe color shift was also observed. In 2021, Wang et al. proposed another fully convolutional network, composed of a feature encoder, feature decoder, and detail supplement module to first estimate the distribution of illumination information in the input image and then compensate for the corresponding uneven illumination [4]. The results were reported based on a private

dataset containing 300 female reproductive tract pathological cells (FRTPC) tiles with the same illumination pattern, in which 250 and 50 images have been allocated to the training and test sets, respectively. Additionally, they employed tiles from 49-01 and 53-03 samples from the public dataset [1] as train and test sets, respectively. The loss function is a weighted sum of Euclidean distance and the structural similarity index (SSIM). The results reveal that the network has been successful in correcting data similar to the training set. In 2024, Nemati et al. proposed a model based on pix2pix, a subset of a generative adversarial network (GAN), with modifications tailored for uneven illumination correction [5]. Their study demonstrated that increasing the bottleneck size enhances uniformity in the corrected images. However, they also observed that a larger bottleneck reduces the level of detail in the reconstructed output. This finding highlights a fundamental trade-off between achieving more uniform illumination correction and preserving fine image details.

In this study, we proposed a straightforward yet effective training set design that improves the generalization capability of tile-based deep-learning methods for correcting uneven illumination in WSI. The proposed method is focused on the efficiency of deep learning algorithms and their adaptability to previously unseen textures and illumination patterns.

Material and Methods

This technical study investigates a novel training set design strategy for deep learning-based uneven illumination correction in WSI.

Whole Slide Imaging dataset

In the present study, the public WSI dataset of bright-field optical microscopic images was used as the Tak dataset [1]. This dataset is acknowledged as an appropriate dataset for uneven illumination correction applications because the tiles are available in their raw form with no pre-processing, and the corresponding ground truth tiles are obtained from a golden standard Empty-Zero algorithm [5]. This dataset contains ten samples of different cells with

distinctive texture and illumination patterns, each identified by a unique digital identifier (ID). Each sample, except for one, includes 100 tiles (sample with ID 194-01-70 has 94 tiles). Images are all 8-bit, three-channel RGB color with a size of 1719×2304 pixels.

Proposed Approach

The supervised tile-based deep learning structures are fully automatic without any human intervention for parameter settings, as mentioned in the previous section. They can handle various types of uneven illumination in real time; however, their performance is highly dependent on the training set. Regardless of network strength, they struggle with generalizing to new, unseen images with different textures and illumination patterns. To address this limitation, we focused on the training phase to enhance adaptability to variations in illumination patterns and texture details, while concurrently maintaining the original resolution of the images captured by the WSI technique, utilizing the available computational resources.

Our strategy aimed to preserve the entire uneven illumination patterns as global features critical to effective learning for the networks. We designed three distinct training sets utilizing a modified version of the network proposed by Wang et al. [4], in which we do not require patch-based training strategy. It is worth mentioning that the Wang network uses small patches of size 384×384 pixels, randomly extracted from the input image due to resource constraints associated with processing large, high-resolution whole slide images. However, this approach has a significant drawback: the entire uneven illumination pattern, which occurs smoothly as a global feature across the captured images by the WSI technique, may not be adequately captured by the model. This limitation results in a lack of diversity in the network's understanding of such patterns, potentially impairing its generalizability to new and varied datasets.

To generate the training sets, ground truth tiles of each sample, originally sized at 1719×2304 pixels, were cropped into tiles of size 512×512 pixels. These cropped tiles were then multiplied

by a resized uneven illumination pattern derived either from the same sample or from other samples. Resizing the illumination pattern does not degrade its essential characteristics because its variation is very gradual and consistent. Unlike the random patch extraction approach, which could confuse the network by providing patches from different parts of the unevenly illuminated image, our method combines cropped tiles with the complete illumination pattern. This allows the network to perceive the pattern holistically, which is crucial for accurately learning and correcting uneven illumination. This strategy not only preserves the resolution of whole slide images but also respects computational constraints. The illumination pattern was computed as the average over the extracted illumination patterns of all tiles within a given sample. By maximizing data diversity and providing a comprehensive representation of the illumination variations, this approach enhances the network's generalization performance. Three distinct training sets, as presented in Table 1, were implemented based on the Tak dataset as follows:

Training Set 1 (TS1): With a focus on dataset expansion, this set significantly increased the original 100 images from sample 49-01 to 1200 cropped images. Each of these cropped images was multiplied by a single uneven illumination pattern, thereby enhancing the dataset size and providing a robust foundation for the model's learning capabilities.

Training Set 2 (TS2): To further enrich the illumination pattern variety, the second training set incorporated 3264 cropped images. Within this set, 408 cropped images from sample 49-01 were multiplied by 8 distinct uneven illumination patterns. This deliberate infusion of diverse illumination patterns served to emphasize the adaptability of the model to a broader range of scenarios.

Training Set 3 (TS3): Combining variations in both illumination patterns and textures, TS3 comprised 408 cropped images from 8 distinct samples. Each of these images was multiplied by its corresponding uneven illumination pattern, yielding a total of 3264 images. This

Table 1: The details of the first training set (TS1), second training set (TS2), and third training set (TS3), and cross-validation designs providing the ID of samples and uneven illumination patterns from the Tak dataset used for training and testing.

TS1 1 sample 1 shading pattern Augmentation of number of images	TS2 1 sample 8 shading patterns Augmentation of shading pattern	TS3 8 distinct samples 8 distinct shading patterns Augmentation of shading pattern & texture				
		Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Train						
#Tiles: 1200	#Tiles: 3264	#Tiles: 3264	#Tiles: 3264	#Tiles: 3264	#Tiles: 3264	#Tiles: 3264
Sample: • 49-01	Sample: • 49-01	Samples & patterns:	Samples & patterns:	Samples & patterns:	Samples & patterns:	Samples & patterns:
Pattern: • 49-01	Patterns: • 026-0191 • 051-04-80 • 156-01-86 • 234-01-67 • 31-01 • 33-03 • 36-01 • 49-01	• 051-04-80 • 156-01-86 • 026-01-91 • 234-01-67 • 31-01 • 33-03 • 36-01 • 49-01	• 051-04-80 • 156-01-86 • 026-01-91 • 234-01-67 • 194-01-70 • 31-01 • 31-01 • 53-03 • 49-01	• 156-01-86 • 026-01-91 • 234-01-67 • 194-01-70 • 31-01 • 36-01 • 53-03 • 33-03	• 051-04-80 • 156-01-86 • 234-01-67 • 194-01-70 • 33-03 • 36-01 • 53-03 • 49-01	• 051-04-80 • 026-01-91 • 194-01-70 • 31-01 • 36-01 • 53-03 • 49-01
Test						
	Samples: • 53-03 • 194-01-70		Samples: • 33-03 • 36-01	Samples: • 051-04-80 • 49-01	Samples: • 026-01-91 • 31-01	Samples: • 156-01-86 • 234-01-67

TS: Training Set
#stands for “number of”

comprehensive approach aimed to bolster the model’s ability to generalize to new textures and illumination conditions.

To assess the performance and generalization capability of the trained network, we employed a 5-fold cross-validation strategy for TS3, as depicted in Table 1. In each fold, eight datasets were selected for training while the remaining two datasets were used for testing. This process ensured that every dataset participated in both training and validation exactly once across the 5 folds. This iterative process allowed us to evaluate the network’s consistency and effectiveness in managing variations in illumination patterns and texture details across different subsets of the data.

Evaluation metrics

Different evaluation criteria provide a more comprehensive assessment of the quality and effectiveness of algorithms. Image processing algorithms are evaluated using reference-based

and non-reference metrics. The reference-based criteria are typically well-defined mathematical calculations, particularly evaluating pixel differences between the processed image and the ground truth. The non-reference metrics, on the other hand, rely on the inherent characteristics of the image itself, eliminating the need for ground truths [7]. In the present study, algorithms are assessed using common criteria of both reference-based and non-reference perspectives. The former includes Mean Squared Error (MSE), Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), Correction Score (CS) [3], and DeltaE2000 [8], and the latter contains Entropy, NIQE [4], PIQE [9], and BRISQUE [10]. Algorithms that perform well across a range of criteria are more likely to be effective in various scenarios and datasets, indicating higher reliability and versatility.

Results

The modified version of Wang network was

executed on Google Colab Pro, a cloud-based platform that provides access to GPU resources with 2 Intel Xeon CPUs @ 2.20 GHz, 13 GB of RAM, and a NVIDIA v100 GPU with 12 GB of VRAM.

First, to show the effect of the patch size (PS) on the network's global view, we evaluate the stitched result of the corrected tiles instead of a single tile. We employed the Fast and Robust Microscopic Image Stitching (FRMIS) algorithm [2] for stitching. Stitched image reveals residual patterns from the incomplete correction of uneven illumination that may be overlooked in isolated tiles. Wang's original training set consisted of 100 tiles from only sample 49-01 of the Tak dataset, with data augmentation, performed using random rotations of 90 and 180 degrees [4]. Notably, their training strategy employed a patch size of 384×384 pixels. We modified this parameter and increased the PS to 1024, as shown in the stitched results of three samples from the Tak dataset in Figure 1. The left column is the stitched results of ground truth tiles while the middle and right columns depict the stitched results of tiles corrected by the original Wang network trained with two different PSs. The qualitative inspection indicates smaller PS results in incomplete uneven illumination correction and leads to visible seams in the stitched image. On the contrary, training

the network utilizing patches of greater size resulted in seamless stitched images. However, the results with a higher PS particularly in test samples 234-01-67 and 156-01-86 are more colorless than those obtained from a smaller PS.

To study the impact of the designed training sets, Figure 2 presents a comprehensive quantitative evaluation of uneven illumination correction using the modified Wang network trained on different training sets described in the Method section. The assessment relies on common pixel-wise metrics (MSE, SSIM, PSNR, CS, and DeltaE) as well as non-reference criteria (Entropy, NIQE, PIQE, and BRISQUE) applied to test sets (53-03 and 194-01-70). The results unveil substantial differences in median values of MSE, SSIM, PSNR, and CS. Notably, Wang's model trained on the original dataset (PS=384) exhibits higher errors, while performance gradually improves with the training sets TS1 to TS3. The interquartile range (IQR) of the evaluation metrics follows a similar trend: both the original Wang model and TS1 show greater variability, with TS1 even exceeding Wang's original model in some cases. In contrast, TS2 and TS3 present reduced IQR values, with TS2 achieving the smallest spread in both MSE and SSIM, indicating more consistent performance. DeltaE scores revealed TS1, which significantly augmented the original dataset, resulted

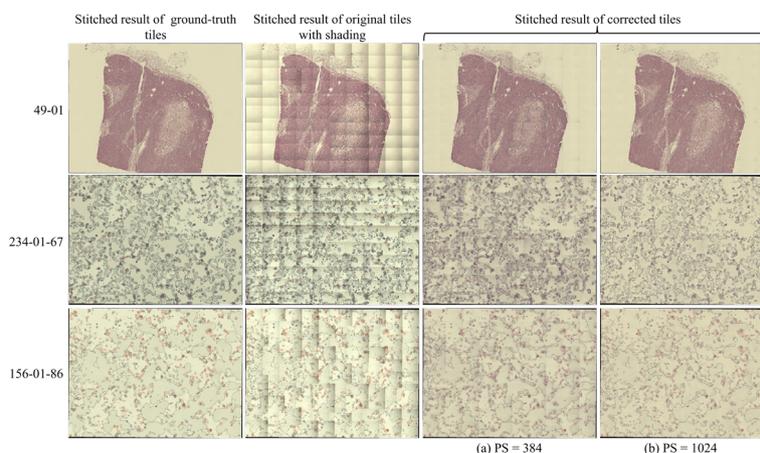


Figure 1: The effect of increasing the global view of the network by increasing the patch size (PS). The stitched results of ground truth tiles, original tiles with shading, and corrected tiles by the Wang network using the original training set including 100 images from only one sample (49-01) of the Tak dataset with (a) PS=384, and (b) PS=1024.

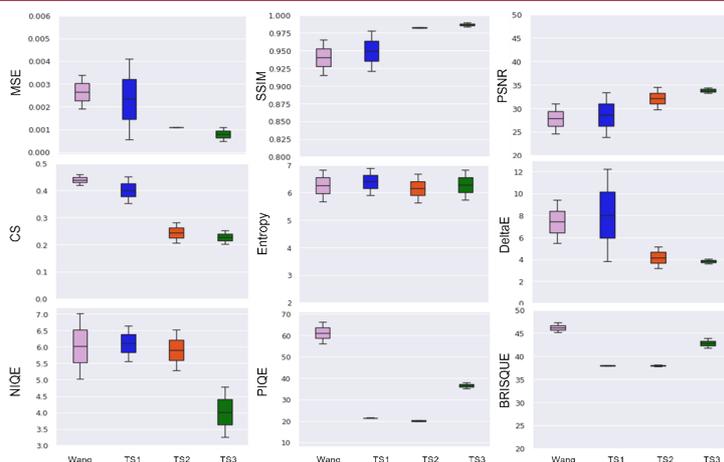


Figure 2: The performance of the network trained with different training sets including the original training set and patch size (PS=384) (labeled as Wang), first training set (TS1), second training set (TS2), and third training set (TS3-Fold1) applied to test sets (53-03 and 194-01-70).

in higher values compared to Wang's original training set. In contrast, TS2 and TS3 achieved lower values than TS1, suggesting better color consistency. For no-reference quality metrics, entropy values remained relatively stable with limited IQR variation. NIQE scores displayed a decreasing trend, with TS3 achieving the lowest median value, indicating superior perceptual quality. Similarly, PIQE and BRISQUE trends categorize Wang's original training set as poor, TS3 as fair, and both TS1 and TS2 as excellent, demonstrating the effectiveness of the proposed training set refinements.

The stitched results obtained from the ground truth images, the original images with uneven illumination, and images corrected by the original Wang, and the modified Wang trained using different configurations in TS1, TS2, and TS3 are shown in Figure 3. The original Wang results exhibit incomplete uneven illumination correction and visible seams in the stitched result of 49-01 training sets, which become more pronounced in the test samples, particularly in 194-01-70, where reconstruction of texture details is poor. The incomplete correction is evident in every single tile. Training with TS1 almost eliminates the illumination pattern in the stitched result of the 49-01 training sets. However, visible patterns persist in the stitching of test samples (53-03 and 194-01-70) and the corresponding single tiles. Notably, TS1 improves color recon-

struction details for 194-01-70 compared to the original Wang. TS2 shows promising results in uneven illumination correction and texture details reconstruction, particularly for test samples (53-03 and 194-01-70). While some inhomogeneities are observed in the results of training sample 49-01 and test sample 53-03, the overall performance is enhanced. TS3 yields seamless results that are most similar to the ground truths in terms of color reconstruction, although again, some inhomogeneities remain due to non-perfect illumination correction in the training set 49-01 and test sample 53-03 results.

Discussion

Our comprehensive evaluation of various training set configurations used in a deep-learning-based model (here the Wang network) provides valuable insights into methodologies of deep learning for the correction of uneven illumination. In particular, Figure 1 demonstrates that increasing the PS leads to more homogeneous outputs, supporting the hypothesis that capturing the full global illumination context is essential for effective training. However, enlarging the PS to 1024 results in less colorful images compared to a smaller PS, attributed to the increase in the bottleneck size of the module from 3 to 8, leading to reduced precision in reconstructing details [5]. In response, we have replaced using PS with cropped images, which

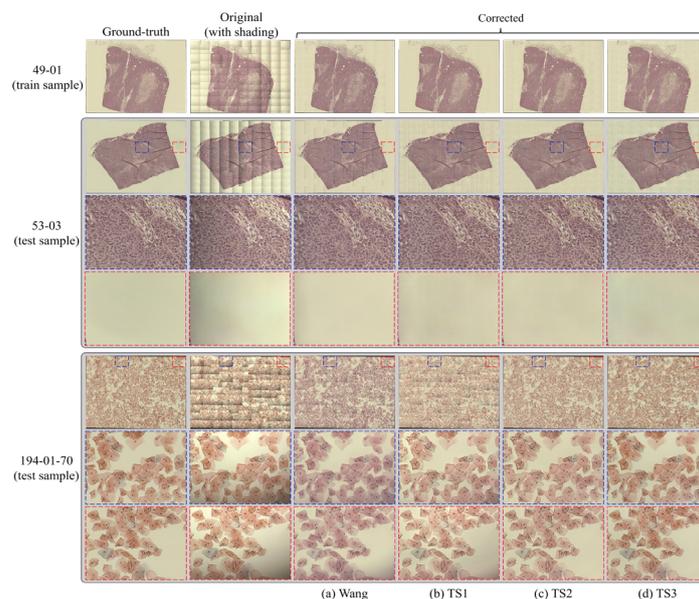


Figure 3: The performance of the network trained with different training sets including the original training set and patch size (PS=384) (labeled as Wang), first training set (TS1), second training set (TS2), and third training set (TS3-Fold1) applied to test sets (53-03 and 194-01-70).

is the procedure described in the Method section in our training set strategies (TS1, TS2, and TS3).

The quantitative assessments of the original Wang are deemed unacceptable, as they exhibit deterioration in most criteria (Figure 2). Limited variation of texture and a small PS contribute to this poor performance, even in terms of the results of the training set. The TS1 demonstrates slight improvements in all criteria, consistent with the findings in Figure 3c. No visible seams in the stitching results of the training samples indicate successful learning and adaptation of the model to the provided data. The homogeneity in the background of test sample 53-03 further suggests the model's effectiveness in generating more balanced and visually appealing outputs compared to the original Wang. Test sample 194-01-70 reveals some challenges in achieving complete uneven illumination correction. Nevertheless, the reconstructed texture color outperforms the original Wang method (PS=384), indicating improvements in the correction process. Results in Figure 2 highlight the advantages of training modified Wang using alternative training sets TS2 and TS3. The smallest IQR of TS2 in criteria, such as MSE, SSIM,

PIQE, and BRISQUE signifies successful generalization, eliminating new patterns effectively. Test results from TS2 consistently demonstrate reliability, suggesting robust performance in handling diverse illumination scenarios. TS3 outperformed the other training configurations in pixel-wise evaluation metrics, generating results more closely aligned with the ground truth. Stitched results further validate this alignment, showcasing texture details resembling ground truths and corrected illumination patterns to an acceptable degree. The model's generalization capabilities, evident in TS2, contribute to its ability to address new patterns, enhancing overall robustness. The reliable test results from both TS2 and TS3 underscore the model's proficiency in producing outputs with improved pixel-wise metrics and alignment with ground truth. However, some inhomogeneity in the results of TS2 and TS3 on 49-01 and 53-03 samples is noted, which we attribute to the procedure for obtaining illumination patterns by averaging patterns extracted from all tiles of a sample. The bold and dense texture colors in these two samples may have influenced the final uneven illumination patterns.

We also evaluated the model's performance

across five folds, demonstrating consistency, with measurements remaining within a similar range and no significant variations, confirming the proposed approach's effectiveness. Addressing the observed inhomogeneity could likely be achieved by enriching the training dataset; however, we were unable to explore this further due to limited computational resources. Nevertheless, we anticipate that increasing the training dataset would mitigate these variations, leading to more homogeneous results, as seen in TS1 when trained on 1200 images from sample 49-01.

Conclusion

This study proposed a training set design strategy to enhance deep learning-based methods for uneven illumination correction to reduce their reliance on substantial computational resources and improve generalization to unseen data. Unlike many image processing tasks, where smaller input patches are advantageous, our findings demonstrate that effective illumination correction requires preserving the full global context of illumination patterns. By maintaining the original image resolution and whole illumination patterns in the training sets, the proposed approach improves the performance of conventional deep learning methods on the benchmark dataset. These results highlight the importance of illumination pattern representation and provide a practical, scalable solution suitable for clinical imaging workflows.

Authors' Contribution

H. Shabani and S. Nemati conceived and designed the research framework. S. Nemati conducted the implementations, analyzed the experimental data, and drafted the original manuscript. H. Shabani and A. Mahmoudi-Aznaveh critically revised and edited the manuscript. H. Shabani was responsible for project administration and supervision. All authors have read and approved the final version for publication.

Conflict of Interest

None

Data and Code Availability Statement

The training sets (TS1, TS2, and TS3) described in this study and the code generating the results are publicly available at https://github.com/labCOI/microscopy_illumination_correction.

References

1. Tak YO, Park A, Choi J, Eom J, Kwon HS, Eom JB. Simple Shading Correction Method for Brightfield Whole Slide Imaging. *Sensors (Basel)*. 2020;**20**(11):3084. doi: 10.3390/s20113084. PubMed PMID: 32485985. PubMed PMCID: PMC7308847.
2. Mohammadi FS, Shabani H, Zarei M. Fast and robust feature-based stitching algorithm for microscopic images. *Sci Rep*. 2024;**14**(1):13304. doi: 10.1038/s41598-024-61970-y. PubMed PMID: 38858367. PubMed PMCID: PMC11164998.
3. Peng T, Thorn K, Schroeder T, Wang L, Theis FJ, Marr C, Navab N. A BaSiC tool for background and shading correction of optical microscopy images. *Nat Commun*. 2017;**8**:14836. doi: 10.1038/ncomms14836. PubMed PMID: 28594001. PubMed PMCID: PMC5472168.
4. Wang J, Wang X, Zhang P, Xie S, Fu S, Li Y, Han H. Correction of uneven illumination in color microscopic image based on fully convolutional network. *Opt Express*. 2021;**29**(18):28503-20. doi: 10.1364/OE.433064. PubMed PMID: 34614979.
5. Nemati S, Shabani H. Uneven Illumination Correction in Whole Slide Imaging using Pix2Pix. In 32nd International Conference on Electrical Engineering (ICEE); Tehran, Iran: IEEE; 2024. p. 1-6.
6. Mei XF, Xie FY, Jiang ZG. Uneven illumination removal based on fully convolutional network for dermoscopy images. In 13th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP); Chengdu, China: IEEE; 2016. p. 243-7.
7. Sinha P, Russell R. A perceptually based comparison of image similarity metrics. *Perception*. 2011;**40**(11):1269-81. doi: 10.1068/p7063. PubMed PMID: 22416586.
8. Sharma G, Wu W, Dalal EN. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Res Appl*. 2005;**30**(1):21-30. doi: 10.1002/col.20070.
9. Venkatanath N, Praneeth D, Bh MC, Channappayya SS, Medasani SS. Blind image quality evaluation using perception based features. In Twenty first national conference on communications (NCC); Mumbai, India: IEEE; 2015. p. 1-6.
10. Mittal A, Soundararajan R, Bovik AC. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*. 2012;**20**(3):209-12. doi: 10.1109/LSP.2012.2227726.