Super-resolution for Whole Slide Histological Images

Zhongao Sun[1], Alexander Khvostikov[1], Andrey Krylov[1] and Nikolai Krainiukov[2]

[1]*Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Leninskie Gory, 1, building 52, Moscow, 119991, Russia*

[2]*Shenzhen MSU-BIT University 1 International University Park Road, Dayun New Town, Longgang District, Shenzhen, Guangdong Province, P.R. China*

Abstract

Histological images serve as crucial tools in the diagnosis and treatment of diverse afflictions. However, the acquisition of images exhibiting exceptional resolution whole slide images, WSIs, capturing intricate textures and vital nuances, can present a formidable challenge, primarily due to the requirement of expensive and intricate apparatus, proficient personnel, and considerable time commitments. To tackle this predicament, it is important that we conceive an effective and precise framework to increase the resolution of whole slide histological images. There are several algorithms used for super-resolution, including interpolation-based, deep learning based and bayes based algorithms. After scrutinizing and dissecting the available super-resolution models and algorithms, we arrived at the conclusion that the most suitable approach for histological WSIs would be to fine-tune the already trained Real-ESRGAN model to reconstruct histological images and apply it in a patch-based way. For histological WSIs, it is typical to have a lagre number of empty areas that do not contain tissue. To circumvent the impact of these empty areas on the model's efficiency, we filter the dataset using Shannon's information entropy. Furthermore, we have modified the structure of the loss function to optimize the reconstruction of low-level details of histological images. In this study, we fine-tune the pre-trained Real-ESRGAN model using the histological image dataset PATH-DT-MSU. It enabled us to outperform all preexisting models in terms of reconstructing low-level details in WSI histological images. Moreover, without retraining the model, we have tested it on additional histological image datasets, thereby proving its high generalization ability.

Keywords

Histological images, Whole slide images, Super-resolution, Convolutional neural networks, generative neural networks

1. Introduction

Medical images are an important tool in the diagnosis and treatment of various diseases. However, obtaining high-quality images with fine textures and important details can be difficult due to complex and expensive equipment, as well as the need for trained personnel and a significant amount of time. In addition, in the case of histological whole slide images (WSIs), the typical resolution of an image can be around 100000×100000 pixels and the image size exceeds several gigabytes. Since these images take up a lot of space, they are difficult to transfer between

histological labs over the network. Therefore, the developed resolution enhancement method can be considered as a kind of compressing codec, so that images will take up less space. Therefore, there is a need for methods that increase the resolution of histological images.

One such method is the use of super-resolution medical imaging models to reconstruct high-quality images from low-quality ones. By improving image quality through post-processing, these models can provide a valuable tool for medical professionals to more accurately diagnose and analyze diseases.

Currently, generative adversarial networks (GANs) [1] are widely used in the field of super-resolution. Although general super-resolution models based on GANs, work well on natural images, medical images have a different structure and require more reconstruction details. Therefore, models trained on common datasets may not show optimal performance in the field of medical super-resolution imaging.

In this regard, the Real-ESRGAN [2] model is an efficient super-resolution model that improves the quality and resolution of general images. It should be noted that this model generates more realistic images with improved details and increase resolution compared to its predecessors. It is worth noting that conventional metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) exhibit fundamental disparities with the subjective assessment of human observers [3]. While non-reference measures such as Ma's score [4] and Natural Image Quality Evaluator (NIQE) [5] are used for evaluating perceptual quality, they are not suitable for medical images. In this study, we fine-tune the pre-trained Real-ESRGAN model using the histological image datasets PATH-DT-MSU to enhance the resolution and quality of such images.

Our aim was to develop a model that not only performs well with our specific dataset but also constructs a model that performs effectively with general histological images. Therefore, we conducted tests on other histological image datasets while keeping the model parameters frozen, and the results demonstrated that our model exhibits a commendable generalization ability.

2. Used data

2.1. PATH-DT-MSU

In this work we use PATH-DT-MSU dataset[6][1] as a main dataset for training and testing the proposed super-resolution technique for WSIs. Authentic biopsy and surgical specimens derived from various segments of the human gastrointestinal tract, including the colon, stomach, and esophagus, were employed in the preparation of paraffin blocks. In order to obtain histological images of superior resolution, the Leica DM4000B/DFC495 microscope and the Leica SCN400 scanner (manufactured by Leica Microsystems, Germany) were employed.

The PATH-DT-MSU dataset has been created with the purpose of amalgamating high-quality histological images from diverse segments of the human gastrointestinal tract. It comprises several subsets, encompassing distinct image types and annotations tailored for various targeted objectives[7][8]. In this work we use WSS1 and WSS2 subset of PATH-DT-MSU, containing whole slide images of colon polyps and gastric cancer with annotations.
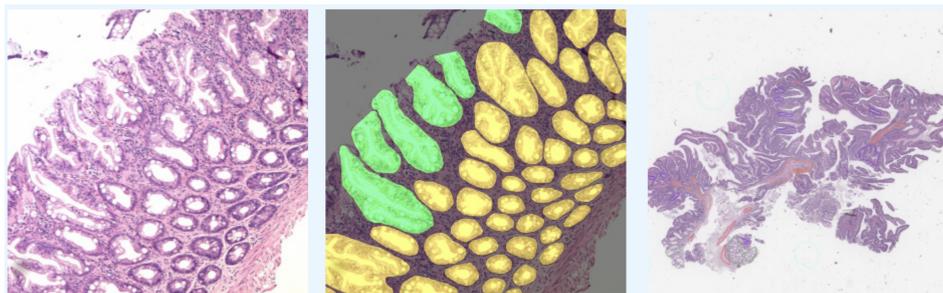
---

[1]https://imaging.cs.msu.ru/en/research/histology

Figure 1: Some examples from PATH-DT-MSU, the image on the left is a partial glandular cell map, the middle is the segmentation result, and the image on the right is an overview of the WSI.

2.2. Deep learning digital pathology dataset

To additionally verify the capabilities of our model, we used Deep learning digital pathology dataset[9][2] for additional testing.

We hvae mainly chosen H&E stained estrogen receptor positive (ER+) breast cancer images, this is because their large variances in appearance as compared to other organs. For example, the area of a breast cancer nucleus can vary by over 200% and have notable differences in texture, morphology, and stain absorption. This may bring challenge for our model to reconstruct the low-level detail. This dataset consist of 143 images ER+ BCa images scanned at 40x. Each image is 2000×2000.

3. Network architecture

3.1. Generator and discriminator of Real-ESRGAN

The generator architecture comprises of 23 RRDB (Residual-in-Residual Dense Block) units, with each unit consisting of three RDB (Residual-Dense Block) blocks [10]. Each Residual-Dense Block consists of five convolutional layers, functioning with the Leaky-ReLU activation function. Following the RRDB blocks, there are a dimensionality expansion block and two additional convolutional layers. It is important to note that the dimensionality expansion block applies the pixel shuffle operation for up-sampling [11].
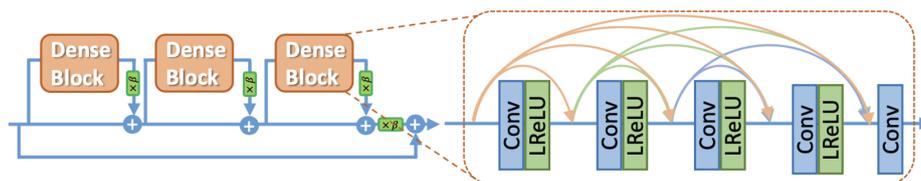

Figure 2: The RRDB of Real-RESGAN block is utilized in our deeper model, and β represents the residual scaling parameter.

---

[2]http://www.andrewjanowczyk.com/deep-learning/

In order to enhance the performance of the generator in producing high-quality images, it is crucial for the discriminator to possess strong discriminative power (see Figure 3). One of the key improvements in the Real-ESRGAN model is the utilization of a U-Net discriminator [12]. The relativistic approach [13] employed in this discriminator enables the prediction of relative realism rather than absolute values. This facilitates precise pixel-level feedback and the generation of realism values for each pixel [14]. Furthermore, it is important to note that in the domain of super-resolution of medical images, our model must preserve important details and avoid the creation of artifacts. The application of this technique proves particularly beneficial for enhancing the resolution of medical images.
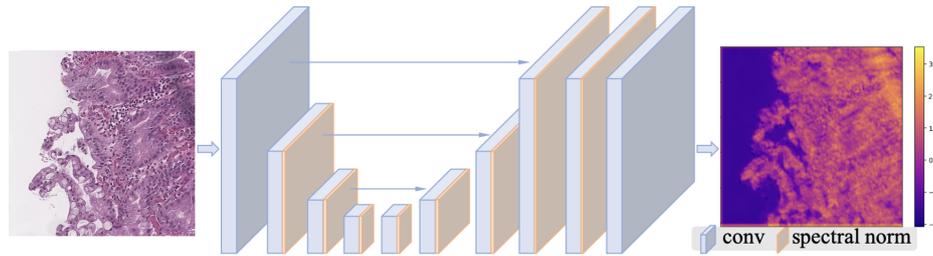

Figure 3: The U-Net discriminator with spectral normalization is employed in Real-ESRGAN to increase the discriminator's capability and stabilize the training dynamics.

3.2. Loss fuction

One of the primary objectives of training a neural network is to minimize the loss function, which is crucial for achieving high model performance. The Real-ESRGAN model employs a three-component loss function, which includes:

$$L = \lambda_{pix}L_{pix} + \lambda_{per}L_{per} + \lambda_{gan}L_{gan} \tag{1}$$

Here, L_{pix} represents the pixel-wise loss[15], L_{per} denotes the perceptual loss[16], and L_{gan} represents the adversarial loss[17]. The coefficients λ_{pix}, λ_{per}, and λ_{gan} control the importance of each component in the overall loss function. A more detailed descruption of used loss functions is provided below:

- **Pixel Loss**: Pixel loss refers to the average absolute difference between the model's output data, denoted as I_{sr}, and the true high-resolution image, denoted as I_{gt}.

$$L_{pix} = \frac{\sum_{x=1}^{W} \sum_{y=1}^{H} \left| I_{sr} - I_{gt} \right|}{W \cdot H}, \text{ where the image dimensions are } W \cdot H \tag{2}$$

- **Perceptual loss** [16]: As we said before, although models based on metrics such as PSNR and SSIM can achieve good results in terms of metrics, while achieving particularly high PSNR, solutions of optimization problems often lack high-frequency content which results in perceptually unsatisfying solutions with overly smooth textures Therefore, based on the

perspective of perceptual similarity, the Real-ESRGAN model uses perceptual loss. From the conceptual point of view, perceptual loss refers to letting many people evaluate the quality of images with the naked eye. Translated to the actual model training process, we use a pre-trained VGG model that has been through millions of memorized images and can extract the features in the images well. The exact formula is as follows:

$$L_{per} = \frac{\sum_{i=1}^{K} \sum_{j=1}^{N} \lambda_i \left| \phi_{ij}(I_{sr}) - \phi_{ij}\left(I_{gt}\right| \right)}{W \cdot H \cdot N \cdot K}, \tag{3}$$

where i denotes the feature map after the i-th convolutional layer, j represents the index of the feature map, λ_i denotes the coefficient of the corresponding convolutional layer, ϕ represents the feature extraction operation, and W, H refer to the image dimensions. N and K represent the number of corresponding convolutional layers and the number of feature maps, respectively. I_{sr} denotes the image after super-resolution, and I_{gt} represents the original high-resolution image. In our code, we utilize the first five convolutional layers of the VGG19 network[18].

- **GAN loss** [17]: Real-ESRGAN incorporates a conditional GAN, where the discriminator evaluates not only the realism of the images but also their fidelity to the target content. To achieve this, the GAN loss function consists of two components: the discriminator loss and the generator loss. The discriminator loss measures the discriminator's ability to differentiate real images from generated ones, while the generator loss assesses how well the generated images can deceive the discriminator by appearing real. The objective of both losses is to strike a balance between realism and fidelity to the target content, resulting in the generation of high-quality images while preserving important details,

$$L_{gan} = -\frac{1}{n} \sum_{i=1}^{n} \log D_{\theta_D}(l_{sr}) \tag{4}$$

Here, D_{θ_D} represents the discriminator with parameters. To ensure better gradient behavior, we minimize $-\log D_{\theta_D}(l_{sr})$ instead of $(1 - \log D_{\theta_D}(l_{sr}))$ [19].

4. Experiment and results

4.1. Proposed method in transfer learning

First and foremost, it is indeed a crucial consideration that the resolution of a single whole slide histological image is typically on the order of $100,000 \times 100,000$ pixels, making it impractical to directly input the entire image into the model. To address this issue, we adopt a patch-based approach,where the large WSI is divided into smaller patches. In this case, we utilize 1024×1024 as reference patch, and downsample it to 256×256 as input patch.

Second, it is imperative to acknowledge that the used PATH-DT-MSU suffers from a pronounced data imbalance, whereby a notable proportion of blank background areas is present. Incorporating such areas into the training set would greatly hinder the model's capacity to effectively gauge the quality of the reconstructed high-resolution images, as well as leading to color imbalance in the resultant reconstructions.

In our endeavor to tackle this issue, we have proposed the implementation of Shannon's information entropy as a mechanism to achieve data balance, thereby eliminating the background patches[20]. The formula can be expressed as follows:

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{5}$$

Here the p_i represents the probability of a particular pixel intensity value occurring. In particular, we convert the image to grayscale and compute the probability distribution of pixel intensities in the image. Count the frequency of each unique intensity value and divide it by the total number of pixels in the image to obtain the probability of each intensity value. The resulting value of Shannon's information entropy represents the average amount of information carried by each pixel in the image. Higher entropy values indicate greater uncertainty and information content, while lower entropy values suggest more predictability or redundancy in the image. Therefore, we can employ Shannon's information entropy to filter the bacground image.

By a random selection process, we have chosen a subset of 100 images that exhibit a richer content, after which we computed the average Shannon's information entropy of these images. Consequently, we have determined that images with Shannon's information entropy exceeding 7 align with our data requirements, and thus, we have incorporated them into the training and testing datasets.

Subsequently, we posited that PSNR (Peak Signal-to-Noise Ratio)[3] holds significant importance in capturing the intricate details of histological images. Hence, we amplified the weight L_{pix} to a value of 2 and diligently monitored the fluctuations in PSNR throughout the entire training phase, as illustrated in Figure 4.
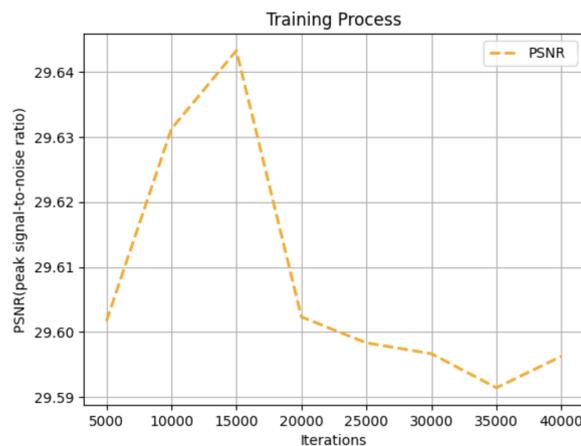

Figure 4: PSNR on test set during the transfer learning, the y-axis is the value of PSNR, and the x-axis is the number of iteration, each epoch consists of 100 iteration.

4.2. Result on Test dataset of PATH-DT-MSU

We applied the model to the filtered test set by means of super-resolution, reconstructing the 512×512 image to 2048×2048. Consequently, we achieved **PSNR of 29.54** and a **SSIM of 0.62** on the test set. In the realm of super-resolution, a PSNR value close to 30 is often deemed acceptable, while the SSIM value is not as favorable, possibly due to our model not being explicitly optimized for SSIM. Furthermore, we randomly selected and present the results in Figure 5.
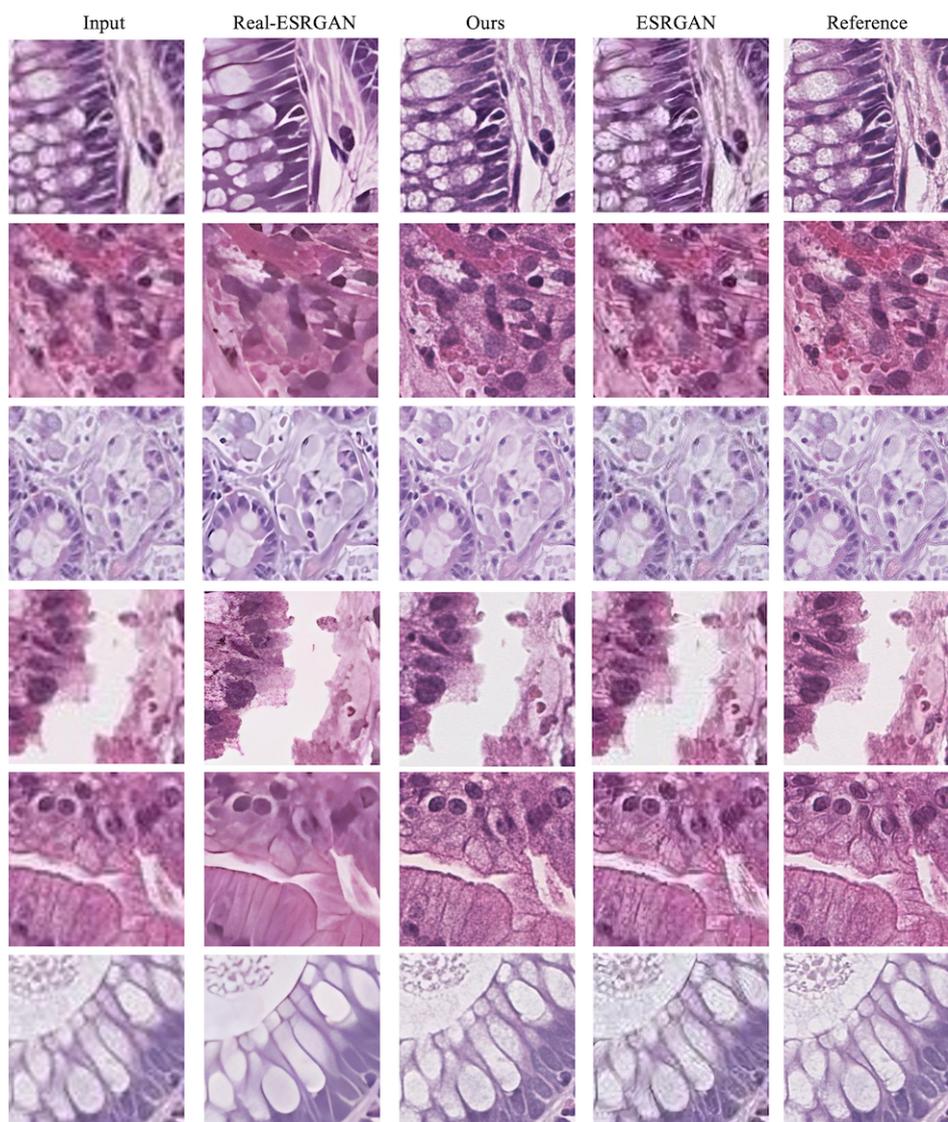

Figure 5: Results of zooming the histological patches 4x in dataset PATH-DT-MSU, the rightmost column is the original 40x patch, the leftmost column is the input 10x patch, and the middle column is our super-resolution result.

As evident from the figure, the high-resolution image reconstructed by our model excels in preserving details and exhibits the least deviation from the reference image. The original Real-ESRGAN model tends to smoothen the edges excessively, leading to a loss of significant details. On the other hand, the ESRGAN model poorly recovers details and introduces artifacts in multiple regions.

4.3. Testing on other datasets

To assess the generalizability of our model, we applied the trained model to other histological images without any additional adjustments. We obtained additional publicly available histological datasets[9] and evaluated our model accordingly. The results are presented in Figure 6.

The test images presented in the figures demonstrate that our model upscale image without losing intricate details, and in some cases, even exhibits a clearer quality compared to the original images. This can be attributed to the proposed neural network architecture, which enables the generation of high-quality images.

Based on the performance on other datasets, we can conclude that our model exhibits good generalization ability.
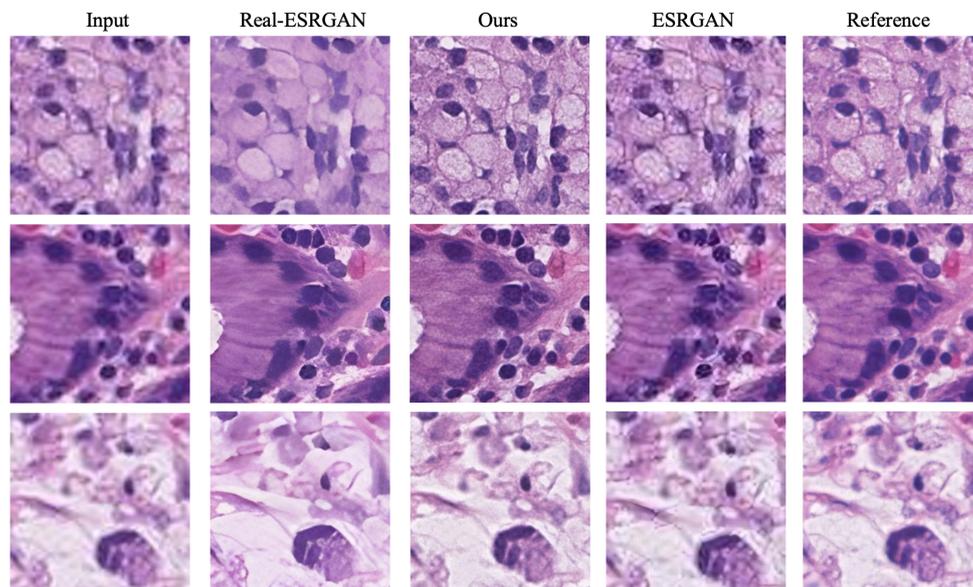

Figure 6: Results of zooming the nucleus patches 4x in Deep learning ditital pathology dataset.[9] The input column is the bilinear interpolation from the 128×128 patches.

5. Conclusion

In this study, we addressed the challenge of handling large histological whole slide images (WSI), which occupy substantial storage space and pose difficulties in network transmission. To tackle this issue, we explored super-resolution models that focus on locally preserving low-resolution

images and transforming them into high-resolution counterparts using a model. We compared different methods and ultimately chose the Real-ESRGAN model, performing transfer learning on a pretrained model.

To address the issue of dependency on details in histological super-resolution images and mitigate the inclusion of background images that cause color imbalance in the reconstructed image, we proposed approaches utilizing Shannon's information entropy for dataset filtering and modifying parts of the loss function. The trained model achieved a PSNR of 29.54 and an SSIM of 0.62 on the PATH-DT-MSU WSS1 and WSS2 test dataset, indicating that our model reconstructs images with more details compared to other models.

To evaluate the model's generalization ability, we also tested it on other datasets, and the results were also promising.

In conclusion, the proposed super-resolution model is able to increase the resolution of histological whole slide images by 4 times, which is equivalent to increasing the effective optical resolution from 10x to 40x and retain the rich details. Thus the size of stored and transmitted through networks WSIs can be decreased by 16 times. This research contributes to the development of efficient and accurate methods for handling large histological images in various applications.

6. Acknowledgements

7. Reference

[1] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. *Generative Adversarial Networks*. arXiv:1406.2661 [stat.ML], 2014.

[2] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, and Chen Change Loy. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. *arXiv preprint arXiv:2107.10833*, 2021.

[3] Ledig,C.,Theis,L.,Husza r,F.,Caballero,J.,Cunningham,A.,Acosta,A.,Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super- resolution using a generative adversarial network. In: CVPR. (2017)

[4] Ma, C., Yang, C.Y., Yang, X., Yang, M.H.: Learning a no-reference quality metric for single-image super-resolution. CVIU 158 (2017) 1–16

[5] Mittal, A., Soundararajan, R., Bovik, A.C.: Making a completely blind image quality analyzer. IEEE Signal Process. Lett. 20(3) (2013) 209–212

[6] A. Khvostikov, A. Krylov, I. Mikhailov, O. Kharlova, N. Oleynikova, P. Malkov. "Automatic mucous glands segmentation in histological images" In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences

[7] I. Mikhailov, A. Khvostikov, A. S. Krylov, P. Malkov, N. Oleynikova, O. Kharlova, N. Danilova. "Development of semi-automatic interactive algorithm for annotating histological images of colon epithelial neoplasms" // Virchows Archiv,, Vol. 477, S1, Springer Verlag (Germany), 2020, S27-S27. DOI.

[8] A. Khvostikov, A. S. Krylov, I. Mikhailov, P. Malkov. "CNN assisted hybrid algorithm for medical images segmentation" // In: ICBIP '20: Proceedings of the 2020 5th International Conference on Biomedical Signal and Image Processing. Suzhou, China, 2020, pp. 14 - 19. DOI.

[9] Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. J Pathol Inform. 2016 Jul 26;7:29. doi: 10.4103/2153-3539.186902. PMID: 27563488; PMCID: PMC4977982.

[10] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong and Yun Fu. Residual Dense Network for Image Super-Resolution. *arXiv preprint arXiv:1802.08797*, 2018.

[11] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., ... & Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1874-1883).

[12] Schonfeld, E., Schiele, B., & Khoreva, A. (2020). A u-net based discriminator for generative adversarial networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 8207-8216).

[13] Jolicoeur-Martineau, A. (2018). The relativistic discriminator: a key element missing from standard GAN. arXiv preprint arXiv:1807.00734.

[14] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, Xiaoou Tang, ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks, arXiv:1809.00219

[15] X. He and J. Cheng, "Revisiting L1 Loss in Super-Resolution: A Probabilistic View and

Beyond," arXiv preprint arXiv:2201.10084, 2022.

[16] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In ECCV, 2016.

[17] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In NeurIPS, 2014.

[18] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[19] I.Goodfellow,J.Pouget-Abadie,M.Mirza,B.Xu,D.Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Advances in Neural Information Processing Systems (NIPS), pages 2672–2680, 2014.

[20] Kieran G. Larkin. Reflections on Shannon Information: In search of a natural information-entropy for images.

[21] Gupta, R., Sharma, A., & Kumar, A. (2020). Super-resolution using gans for medical imaging. Procedia Computer Science, 173, 28-35.

[22] Dharejo, F. A., Zawish, M., Zhou, F. D. Y., Dev, K., Khowaja, S. A., & Qureshi, N. M. F. (2021). Multimodal-Boost: Multimodal Medical Image Super-Resolution using Multi-Attention Network with Wavelet Transform. arXiv preprint arXiv:2110.11684.

[23] Ahmad, W., Ali, H., Shah, Z., & Azmat, S. (2022). A new generative adversarial network for medical images super resolution. Scientific Reports, 12.

[24] Dong, C., Loy, C.C., He, K., Tang, X. (2014). Learning a Deep Convolutional Network for Image Super-Resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8692. Springer, Cham.

[25] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.C., & Bengio, Y. (2014). Generative Adversarial Nets. NIPS.

[26] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *CoRR*, abs/1609.04802, 2016.

[27] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang and Wenzhe Shi. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *CVPR*, 2017.

[28] Gatys, L., Ecker, A.S., Bethge, M.: Texture synthesis using convolutional neural networks. In: NIPS. (2015)