MiHATP:A Multi-Hybrid Attention Super-Resolution Network for Pathological Image Based on Transformation Pool Contrastive Learning*

Zhufeng Xu[1,2], Jiaxin Qin[1,2], Chenhao Li[1,2], Dechao Bu[1,*], Yi Zhao[1,*]

[1] Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
{xuzhufeng22s,lichenhao22s1,budechao,biozy}@ict.ac.cn
[2] University of Chinese Academy of Sciences, Beijing, China
qinjiaxin23@mails.ucas.ac.cn

Abstract. Digital pathology slides can serve medical practitioners or aid in computer-assisted diagnosis and treatment. Collection personnel typically employ hyperspectral microscopes to scan pathology slides into Whole Slide Images (WSI) with pixel counts reaching the million level. However, this process incurs significant acquisition time and data storage costs. Utilizing super-resolution imaging techniques to enhance low-resolution pathological images enables downstream analysis of pathological tissue slice data under low-resource and cost-effective medical conditions. Nevertheless, existing super-resolution methods cannot integrate attention information containing variable receptive fields and effective means to handle distortions and artifacts in the output data. This leads to differences between super-resolution images and authentic images depicting cell contours and tissue morphology. We propose a method named **MiHATP**: A Multi(**Mi**)-Hybrid(**H**) Attention(**A**) Network Based on Transformation(**T**) Pool(**P**) Contrastive Learning to address these challenges. By constructing contrastive losses through reversible image transformation and irreversible low-quality image transformation, MiHATP effectively reduces distortion in super-resolution pathological images. Within MiHATP, we also design a Multi-Hybrid Attention structure to ensure strong modeling capability for long-distance and short-distance information. This ensures that the super-resolution network can obtain richer image information. The experimental results show that MiHATP achieves the best performance in both the super-image reconstruction and downstream cell segmentation and phenotypes tasks. The implementation code will be available at https://github.com/rabberk/MiHATP.git.

Keywords: Super-resolution · Contrastive Learning · Hybrid Attention.

1 Introduction

Precision medical technology involves customizing personalized treatment plans for patients to achieve optimal therapeutic outcomes and success. Digital histopatho-

* Corresponding authors: Y. Zhao and D. Bu.

logical slides offer high-resolution tissue information, allowing physicians to comprehend the histopathological details of patients and formulate effective precision medical strategies [3, 19]. In recent years, with the rapid advancement of computer-aided diagnostic technologies, a series of deep learning diagnostic algorithms based on Whole Slide Image (WSI) pathology slides have been developed [2, 11, 14, 23]. These algorithms rely on the rich information provided by high-resolution pathology slide images. Therefore, the inevitable results of handling large volumes of high-resolution images prolong the data transmission time and increase data storage costs. Meanwhile, scanning equipment for high-magnification digital pathological slides is expensive [17] that not all hospitals can support it. Therefore, the methods are urgently needed to obtain high-quality pathological images at low cost.

Many previous works have demonstrated that super-resolution networks could perform well on medical images [1, 15, 20, 26] such as CT, MRI, pathological images, etc. On the other hand, since the success of Vision Transformer (ViT) in high-level vision tasks, many ViT methods have recently been applied to various low-level tasks [7, 10, 12, 26]. Among them, works such as SwinIR [16] and HAT [5] have emerged successively in super-resolution research. Additionally, with the rise of unsupervised learning, contrastive learning has achieved remarkable success [4, 21]. Some works also make progress by combining contrastive learning with super-resolution tasks. Gang Wu *et al.* [24] attempt to integrate a contrastive learning scheme into super-resolution work, constructing an effective and task-specific data augmentation strategy to generate multiple informative positive and challenging negative samples, demonstrating impressive results. However, these works are unsuitable for transferring to pathological digital pathology images. This is because using sharpening methods to construct positive examples for contrastive learning may potentially distort the contrast and color of the super-resolution images, leading to artifacts in some tissue areas [13], which can affect the performance of algorithms in downstream digital pathology image analysis tasks. To address these issues, our study carefully integrates contrastive learning into super-resolution networks, providing a new approach to improving the analysis of low-magnification digital pathology images. The contributions of our work can be summarized as follows:

- We construct a framework based on reversible and irreversible transformation pools for contrastive learning on low-magnification tissue images in pathology. This framework conducts contrastive learning simultaneously in both image space and feature space, which effectively builds super-resolution pathological images with more apparent cell contours and feature representations with more robust local characterization.
- Our method incorporates a multi-hybrid attention mechanism, allowing for the blending of multiple attention strategies. This enables our method to acquire the most suitable receptive field information adaptively.
- We validate the comparison between different super-resolution methods and MiHATP at multiple scales and verify the downstream cell segmentation and

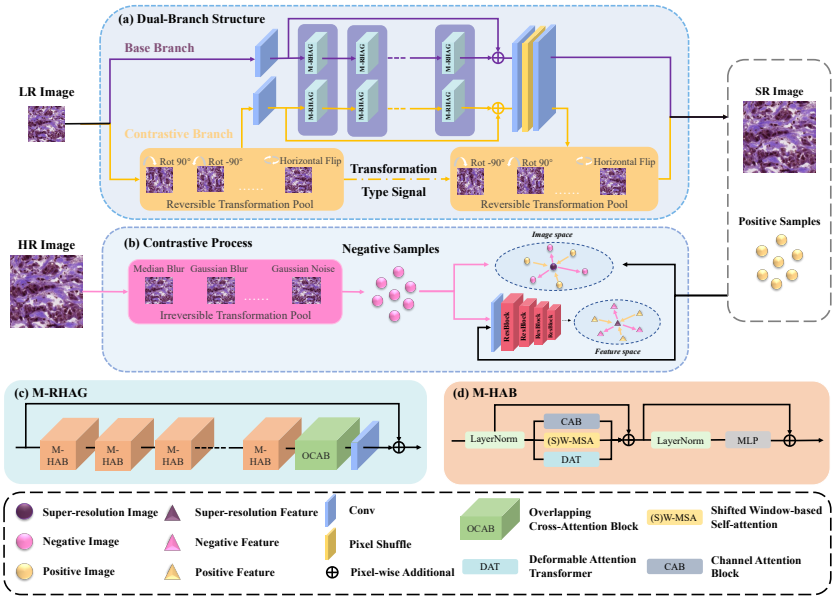

Fig. 1: Overview architecture of our proposed MiHATP model.

phenotypes tasks of super-resolution images under different methods. The results show that our method achieves state-of-the-art performance.

2 Method

The entire pipeline of the MiHAPT is shown in Fig. 1. It consists of a Dual-Branch Structure and a Contrastive Process, the framework operates as follows: The Dual-Branch Structure utilizes two separate branches to generate the super-resolution (SR) image and positive samples processed through a pairwise reversible transformation pool. The Contrastive Process generates negative samples through an irreversible transformation pool. The SR image, positive samples, and low-quality negative samples are utilized for supervised contrastive learning in both the image space and feature representation space.

2.1 Reversible and Irreversible Transformation Pool

As mentioned above, to avoid distortion of cell contours and artifacts, we did not utilize the methods in existing work to construct contrastive loss for super-resolution tasks. We construct a contrastive learning based on transformation pools. The transformation pools consist of a pair of reversible transformation pool for constructing positive examples and an irreversible transformation pool for constructing the negatives. The structure is illustrated in Fig. 1, and we can

generate the following contrastive learning set:

$$I_i^{SR} = F_{base}\ (I_i^{LR}; \theta_{base}\ ) \tag{1}$$

$$X_i^P = \{P_j | P_j = \ T_{inv}(F_{con}(T_{for}(I_i^{LR}); \theta_{con}))\ \}_{j=1}^{K_p} \tag{2}$$

$$X_i^N = \{N_j | N_j = T_{irr}(I_i^{HR})\}_{j=1}^{K_n} \tag{3}$$

Among them, K_p and K_n represent the sample quantities of positive and negative examples generated through the transformation pool. I_i^{SR} represents the i-th super-resolution images in training set produced by MiHATP, while X_i^P and X_i^N denote the positive and negative samples of I_i^{SR}. F_{base} and F_{con} represent the two Super-Resolution Branch pipeline functions in the Dual-Branch structure, and θ_{base} and θ_{con} represent the approximate parameters in the corresponding networks. T_{for}, T_{inv}, and T_{irr} respectively denote random transformation functions in the reversible positive transformation, reversible inverse transformation, and irreversible transformation pools.

The data transformed by reversible transformation pool maintains consistent image quality with the original pathological slice patches, ensuring quality preservation. Conversely, patches obtained through irreversible transformations serve as negative examples in contrastive learning. We computed the cosine similarity between super-resolution images and positive examples, as well as between super-resolution images and negative examples, and thereby constructed a contrastive loss. This helps moving super-resolution images away from negative examples in both image and feature spaces, guiding them towards a more dispersed position in the sample space.

2.2 Multi-Hybrid Attention

In conventional ViT architectures, to mitigate the computational cost and slow convergence caused by an excessive number of keys participating in each query patch, manual attention region restriction is commonly employed to limit the receptive field. However, this approach often leads to the neglect of potentially relevant long-range attention information. In MiHATP, the Dual-Branch framework incorporating M-RHAG (dual-residual hybrid attention block groups) is introduced to handle data from SR images and outputs of reversible positive transformation pools.

The Dual-Branch structure consists of two pathways, a Base Branch and a Contrastive Branch. Each comprises a set of M-RHAG blocks. Unlike the original RHAG modules proposed in HAT[5], M-RHAG integrates multiple M-HAB (multi-hybrid attention block) modules before the OCAB (Overlapping Cross-attention Block) to enhance the network's adaptability in selecting attention region specific to each input with the DAT (Deformable Attention Transformer) [25]. As illustrated in Fig. 1, each M-RHAG is connected by residual links to several M-HAB blocks. Within the M-HAB modules, the following operations

are applied to the input tensor X:

$$\begin{aligned} X_{Norm} &= \text{Norm}(X), \\ X_{Med} &= \text{SW-MSA}(X_{Norm}) + \alpha\text{CAB}(X_{Norm}) + \beta\text{DAT}(X_{Norm}), \\ X_{M\text{-}HAB} &= X_{Norm} + \text{MLP}(\text{Norm}(X_{Med} + X)) + X_{Med} + X. \end{aligned} \tag{4}$$

Where X_{Norm} and X_{Med} are the intermediate features, $X_{M\text{-}HAB}$ is the final output of the M-HAB module. Norm is the layer normalization operation, and MLP denotes a multi-layer perceptron. SW-MSA and CAB stand for Shifted Window-based Self-attention and Channel Attention Block. α and β are the weights set to prevent the possible conflict.

2.3 Loss Function

In MiHATP, the total loss of the network is defined as follows:

$$Loss_{total} = Loss_{L1}^{Base} + Loss_{L1}^{Con} + \lambda Loss_{CL} \tag{5}$$

Where λ is the weight of the contrastive loss $Loss_{CL}$, $Loss_{L1}^{Base}$ and $Loss_{L1}^{Con}$ are the L1 losses for Base Branch and Contrastive Branch, respectively. For an output image with dimensions $C \times H \times W$, they are defined as follows:

$$Loss_{L1}^{Base} = \frac{1}{N}\sum_{i=1}^{N} |I_i^{HR} - F_{base}(I_i^{LR}; \theta_{base})| \tag{6}$$

$$Loss_{L1}^{Con} = \frac{1}{N}\sum_{i=1}^{N} |T_{for}(I_i^{HR}) - F_{con}(T_{for}(I_i^{LR}); \theta_{con})| \tag{7}$$

Here, N represents the number of training images. We employ supervised loss to ensure effective training of the super-resolution networks for both the Base Branch and Contrastive Branch.

We also utilize a novel contrastive learning in both image space and feature space. This method aims to bring the original super-resolution images obtained from the Base Branch closer and the positive images obtained after the inverse transformation of the Contrastive Branch while pushing away the negative samples obtained from the irreversible transformation. The defined $Loss_{CL}$ is as follows:

$$Loss_{CL} = \frac{1}{L+1}\sum_{l=0}^{L} Loss_{CL}^{l} \tag{8}$$

The contrastive loss $Loss_{CL}^{l}$ for each layer is defined as:

$$Loss_{CL}^{l} = \frac{-1}{NK_p}\sum_{i=1}^{N}\sum_{k=1}^{K_p} log\frac{exp(Sim(r_i^l, p_i^{l,k})/\tau))}{exp(Sim(r_i^l, p_i^{l,k})/\tau) + \sum_{j=1}^{K_n} exp(Sim(r_i^l, n_i^{l,j})/\tau)} \tag{9}$$

Where τ is the temperature parameter, L is the number of layers in the feature extraction network, K_p and K_n are the numbers of positive and negative samples obtained through transformation pooling sampling, respectively. Sim is the similarity calculation function, for which we utilize cosine similarity. In our algorithm, contrastive learning is also applied to the image space information of the SR output layer. $\{r_i^l\}_{i=1}^N$ represents the feature representation of the i-th super-resolution image in the training set at the l-th layer of the feature extraction network. $\{p_i^{l,k}\}_{k=1}^{K_p}$ and $\{n_i^{l,j}\}_{l=1}^{K_n}$ denote the positive and negative feature representations of the i-th image in the training set, respectively, at the l-th layer of the feature extraction network. It is noteworthy that when $l = 0$, it signifies the output of the image by the super-resolution network's output layer.

3 Experiment & Result discussion

3.1 Dataset and Implementation

We use Breast invasive carcinoma (BRCA) and Colon adenocarcinoma (COAD) datasets from The Cancer Genome Atlas Program (TCGA) to validate the robustness and generalization capability of our method. We follow the CLAM [18] methodology to segment foreground tissue regions at 40x magnification. The regions are divided into 192x192 resolution patches, serving as high-resolution images for all experiments. Additionally, patches at 5x, 10x, and 20x magnifications are chosen as low-resolution data to simulate downscaling by factors of 8x, 4x, and 2x, respectively. For all experiments, we randomly select 2000 patches for the training set and 100 patches as test samples. Regarding the structural configuration of MiHATP, it is specified that there are six M-RHAG modules for each branch. Within each branch, six M-HAT modules are incorporated. Within the Multi-HAT module, α and β were set to 0.01. We employed the Adam optimizer with an initial learning rate of 0.0001 divided by ten every 50000 iterations. The weight of the contrastive loss, λ, is set to 0.01. The temperature parameter τ is set to 0.05.

3.2 Experimental Results

The Results of Comparison Experiments The results of super-resolution image under different methods are reported in Table 1. Under each experimental condition, five non-intersecting sets of test images were collected. The final result was the average of the outcomes from five experiments. On the COAD and BRCA datasets, MiHATP exhibits significant improvements compared to state-of-the-art (SOTA) algorithms, particularly at an x8 resolution where MiHATP demonstrates over a 3% increase in PSNR. At x4 and x2 resolutions, MiHATP also shows improvements ranging from 0.90% to 1.38%. Moreover, MiHATP demonstrates superior performance in SSIM, especially noticeable at higher magnification factors, despite the limited dataset. Some Examples are shown in Fig. 2. The results show that MiHATP can effectively depict image details

Table 1: Results on the COAD and BRCA datasets($x\%$ shows improvement over best baseline (SwinIR) and $x\%$ shows the decrease in performance compared to MiHATP).

Method	COAD Dataset					
	x2		x4		x8	
	PSNR(db)	SSIM	PSNR(db)	SSIM	PSNR(db)	SSIM
SRCNN [6]	29.1065↓ 9.96%	0.8989↓ 4.64%	23.1145↓ 12.98%	0.7356↓ 6.10%	19.0812↓ 12.16%	0.4342↓ 16.42%
LDM [22]	30.4564↓ 5.81%	0.9010↓ 4.41%	23.9633↓ 9.79%	0.7424↓ 5.23%	19.2209↓ 11.52%	0.4391↓ 15.48%
IDM [8]	30.9378↓ 4.32%	0.9102↓ 3.44%	24.1479↓ 9.09%	0.7469↓ 4.66%	19.5464↓ 10.02%	0.4532↓ 12.76%
ResShift [27]	31.4319↓ 2.79%	0.9282↓ 1.53%	24.3835↓ 8.21%	0.7452↓ 4.88%	19.8880↓ 8.45%	0.4826↓ 7.10%
SHISRCNet [26]	31.8732↓ 1.43%	0.9388↓ 0.40%	25.2459↓ 4.96%	0.7636↓ 2.53%	20.6238↓ 5.06%	0.4811↓ 7.39%
SwinIR [16]	31.9037↓ 1.34%	0.9377↓ 0.52%	26.2012↓ 1.36%	0.7745↓ 1.14%	20.8928↓ 3.82%	0.4923↓ 5.24%
MiHATP (ours)	**32.3356**↑ 1.35%	**0.9426**↑ 0.52%	**26.5634**↑ 1.38%	**0.7834**↑ 1.15%	**21.7234**↑ 3.98%	**0.5195**↑ 5.53%

Method	BRCA Dataset					
	x2		x4		x8	
	PSNR(db)	SSIM	PSNR(db)	SSIM	PSNR(db)	SSIM
SRCNN [6]	32.5606↓ 8.09%	0.9264↓ 3.67%	27.2737↓ 9.62%	0.8415↓ 6.53%	21.9137↓ 11.42%	0.6151↓ 17.78%
LDM [22]	33.0968↓ 6.57%	0.9327↓ 3.02%	27.9129↓ 7.51%	0.8431↓ 6.35%	22.1959↓ 10.28%	0.6383↓ 14.68%
IDM [8]	33.6549↓ 5.00%	0.9423↓ 2.02%	28.1479↓ 6.73%	0.8546↓ 5.08%	22.3688↓ 9.58%	0.6874↓ 8.11%
ResShift [27]	34.2976↓ 3.18%	0.9412↓ 2.13%	28.9399↓ 4.10%	0.8574↓ 4.77%	22.5324↓ 8.92%	0.7032↓ 6.00%
SHISRCNet [26]	34.8323↓ 1.67%	0.9528↓ 0.93%	29.7133↓ 1.54%	0.8709↓ 3.27%	23.3348↓ 5.68%	0.7112↓ 4.93%
SwinIR [16]	35.1081↓ 0.90%	0.9600↓ 0.18%	29.8595↓ 1.05%	0.8870↓ 1.47%	23.8881↓ 3.44%	0.7296↓ 2.47%
MiHATP (ours)	**35.4256**↑ 0.90%	**0.9617**↑ 0.18%	**30.1778**↑ 1.05%	**0.9003**↑ 1.50%	**24.7399**↑ 3.60%	**0.7481**↑ 2.54%

Table 2: Results on cell segmentation and phenotypes.

Metric	Scale	Bicubic	SRCNN	LDM	IDM	ResShift	SHISRCNet	SwinIR	MiHATP
Dice	x2	0.8138	0.8150	0.8184	0.8205	0.8234	0.8241	0.8252	**0.8323**↑**0.71%**
	x4	0.8043	0.8050	0.8056	0.8070	0.8072	0.8089	0.8097	**0.8177**↑**0.80%**
	x8	0.7833	0.7867	0.7904	0.7917	0.7932	0.7951	0.8013	**0.8056**↑**0.62%**
	HR	0.8394							
ACC	x2	0.9246	0.9260	0.9276	0.9294	0.9301	0.9311	0.9308	**0.9330**↑**0.19%**
	x4	0.9148	0.9149	0.9160	0.9173	0.9205	0.9217	0.9221	**0.9264**↑**0.43%**
	x8	0.9086	0.9088	0.9096	0.9114	0.9145	0.9181	0.9188	**0.9204**↑**0.16%**
	HR	0.9377							

in super-resolution pathological data and accurately represent tissue structures across various super-resolution magnifications.

The Results of Downstream Task Validation Experiments We apply the super-resolution network trained on the COAD dataset directly to assess its effectiveness in downstream tasks on the colorectal nuclear segmentation and phenotypes (CoNSeP) dataset **without further training**. The pretrained HoverNet [9] serves as our baseline network for downstream testing. We conduct tasks for magnification factors of x2, x4, and x8. To visually demonstrate the potential performance upper and lower bounds of tasks at different scales, we also perform performance tests on high-resolution images and upsampled images directly obtained through Bicubic interpolation. As shown in Table 2, MiHATP obtains the optimal Dice and ACC scores in all experimental settings, demonstrating that

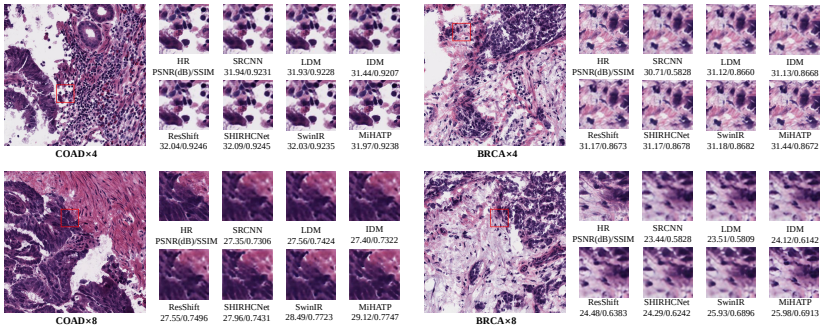

Fig. 2: The x4 and x8 super-resolution results on COAD and BRCA pathological image patches achieved by different methods.

ID	Module				PSNR(dB)	SSIM
	M-HAB	CL				
		IS	ITP	RTP		
1					26.4159	0.7564
2	✔				26.4295	0.7604
3	✔	✔	✔		26.4899	0.7714
4	✔	✔		✔	26.5012	0.7719
6	✔		✔	✔	26.4423	0.7611
7		✔	✔	✔	26.5329	0.7812
8	✔	✔	✔	✔	**26.5634**	**0.7834**

Table 3: The ablation study results for key components of MiHATP.

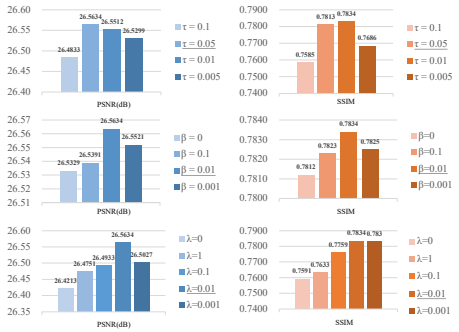

Fig. 3: The ablation study for different hyper-parameters of τ, β, λ.

MiHATP effectively captures cellular morphological information in low-resolution patches.

Ablation Study As shown in Table 3, We conduct ablation studies on the COAD datasets x4 scale super-resolution task. Where IS stands for image space contrastive learning, ITP stands for irreversible transformation pool, and RTP stands for reversible transformation pool. If RTP or ITP is not present, the similarity between the training data and the positive or negative example set is set to 1, respectively. If without IS, contrast learning is performed only in the feature space by default. The experimental results indicate that compared to the baseline (Exp 1), when utilizing the contrastive learning strategy based on transformation pooling (Exp 7) and M-HAB (Exp 2) separately, the PSNR can be improved by 0.117 dB and 0.014 dB respectively, while the SSIM can be increased by 2.48% and 0.40% respectively. When both modules (Exp 8) are applied simultaneously, PSNR and SSIM increase by 0.1475dB and 2.70% compared

to baseline. Additionally, we investigate the impact of crucial hyperparameters in MiHATP, and the results are shown in Fig.3. The hyperparameter settings selected by MiHATP achieve optimal super-resolution performance.

4 Conclusion

This paper introduces MiHATP, a super-resolution approach for pathological images. Our method utilizes a Multi-Hybrid fusion attention strategy, enabling the capture of rich information within variable receptive fields, both short and long distances. Moreover, our approach combines sample constructions using reversible and irreversible transformation pools in both image and feature spaces, forming a contrastive learning framework.

Acknowledgments. This work is supported by The National Key R&D Program of China (2022YFF1203303).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ahmad, W., Ali, H., Shah, Z., Azmat, S.: A new generative adversarial network for medical images super resolution. Scientific Reports **12**(1), 9533 (2022)
2. Bai, B., Yang, X., Li, Y., Zhang, Y., Pillar, N., Ozcan, A.: Deep learning-enabled virtual histological staining of biological samples. Light: Science & Applications **12**(1), 57 (2023)
3. Bera, K., Schalper, K., Rimm, D., Velcheti, V., Madabhushi, A.: Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. Nature Reviews Clinical Oncology **16** (08 2019). https://doi.org/10.1038/s41571-019-0252-y
4. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PMLR (2020)
5. Chen, X., Wang, X., Zhou, J., Qiao, Y., Dong, C.: Activating more pixels in image super-resolution transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22367–22377 (2023)
6. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence **38**(2), 295–307 (2015)
7. Fan, H., Xiong, B., Mangalam, K., Li, Y., Yan, Z., Malik, J., Feichtenhofer, C.: Multiscale vision transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6824–6835 (2021)
8. Gao, S., Liu, X., Zeng, B., Xu, S., Li, Y., Luo, X., Liu, J., Zhen, X., Zhang, B.: Implicit diffusion models for continuous super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10021–10030 (2023)
9. Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N.: Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Medical image analysis **58**, 101563 (2019)

10. Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., et al.: A survey on vision transformer. IEEE transactions on pattern analysis and machine intelligence **45**(1), 87–110 (2022)
11. Kang, M., Song, H., Park, S., Yoo, D., Pereira, S.: Benchmarking self-supervised learning on diverse pathology datasets. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3344–3354 (2023)
12. Khan, S., Naseer, M., Hayat, M., Zamir, S.W., Khan, F.S., Shah, M.: Transformers in vision: A survey. ACM computing surveys (CSUR) **54**(10s), 1–41 (2022)
13. Lee, J., Seo, S., Kim, M.: Sipsa-net: Shift-invariant pan sharpening with moving object alignment for satellite imagery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10166–10174 (2021)
14. Li, B., Li, Y., Eliceiri, K.W.: Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14318–14328 (2021)
15. Li, Y., Sixou, B., Peyrin, F.: A review of the deep learning methods for medical images super resolution problems. Irbm **42**(2), 120–133 (2021)
16. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1833–1844 (2021)
17. Liu, J.T., Glaser, A.K., Bera, K., True, L.D., Reder, N.P., Eliceiri, K.W., Madabhushi, A.: Harnessing non-destructive 3d pathology. Nature biomedical engineering **5**(3), 203–218 (2021)
18. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature biomedical engineering **5**(6), 555–570 (2021)
19. Ma, J., Yu, J., Liu, S., Chen, L., Li, X., Feng, J., Chen, Z., Zeng, S., Liu, X., Cheng, S.: Pathsrgan: Multi-supervised super-resolution for cytopathological images using generative adversarial network. IEEE Transactions on Medical Imaging **39**(9), 2920–2930 (2020). https://doi.org/10.1109/TMI.2020.2980839
20. Peng, C., Zhou, S.K., Chellappa, R.: Da-vsr: domain adaptable volumetric super-resolution for medical images. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24. pp. 75–85. Springer (2021)
21. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PMLR (2021)
22. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
23. Song, A.H., Jaume, G., Williamson, D.F., Lu, M.Y., Vaidya, A., Miller, T.R., Mahmood, F.: Artificial intelligence for digital and computational pathology. Nature Reviews Bioengineering **1**(12), 930–949 (2023)
24. Wu, G., Jiang, J., Liu, X.: A practical contrastive learning framework for single-image super-resolution. IEEE Transactions on Neural Networks and Learning Systems (2023)
25. Xia, Z., Pan, X., Song, S., Li, L.E., Huang, G.: Vision transformer with deformable attention. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4794–4803 (2022)

26. Xie, L., Li, C., Wang, Z., Zhang, X., Chen, B., Shen, Q., Wu, Z.: Shisrcnet: Super-resolution and classification network for low-resolution breast cancer histopathology image. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 23–32. Springer Nature Switzerland, Cham (2023)
27. Yue, Z., Wang, J., Loy, C.C.: Resshift: Efficient diffusion model for image super-resolution by residual shifting. Advances in Neural Information Processing Systems **36** (2024)