Revolutionizing Space Health (Swin-FSR): Advancing Super-Resolution of Fundus Images for SANS Visual Assessment Technology

Khondker Fariha Hossain[1(✉)], Sharif Amit Kamran[1], Joshua Ong[2], Andrew G. Lee[3], and Alireza Tavakkoli[1]

[1] Dept. of Computer Science & Engineering, University of Nevada, Reno, USA
khondkerfarihah@nevada.unr.edu
[2] Michigan Medicine, University of Michigan, Ann Arbor, USA
[3] Blanton Eye Institute, Houston Methodist Hospital, Houston, USA

Abstract. The rapid accessibility of portable and affordable retinal imaging devices has made early differential diagnosis easier. For example, color funduscopy imaging is readily available in remote villages, which can help to identify diseases like age-related macular degeneration (AMD), glaucoma, or pathological myopia (PM). On the other hand, astronauts at the International Space Station utilize this camera for identifying spaceflight-associated neuro-ocular syndrome (SANS). However, due to the unavailability of experts in these locations, the data has to be transferred to an urban healthcare facility (AMD and glaucoma) or a terrestrial station (e.g., SANS) for more precise disease identification. Moreover, due to low bandwidth limits, the imaging data has to be compressed for transfer between these two places. Different super-resolution algorithms have been proposed throughout the years to address this. Furthermore, with the advent of deep learning, the field has advanced so much that $\times 2$ and $\times 4$ compressed images can be decompressed to their original form without losing spatial information. In this paper, we introduce a novel model called Swin-FSR that utilizes Swin Transformer with spatial and depth-wise attention for fundus image super-resolution. Our architecture achieves Peak signal-to-noise-ratio (PSNR) of 47.89, 49.00 and 45.32 on three public datasets, namely iChallenge-AMD, iChallenge-PM, and G1020. Additionally, we tested the model's effectiveness on a privately held dataset for SANS and achieved comparable results against previous architectures.

1 Introduction

Color fundus imaging can detect and monitor various ocular diseases, including age-related macular degeneration (AMD), glaucoma, and pathological myopia (PM) [29]. In remote and under-developed areas, color funduscopy imaging has become increasingly accessible, allowing healthcare professionals to identify and manage these ocular diseases before they progress to irreversible stages. However, interpreting these images can be challenging for inexperienced or untrained personnel, necessitating the transfer

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43990-2_65.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14226, pp. 693–703, 2023.
https://doi.org/10.1007/978-3-031-43990-2_65

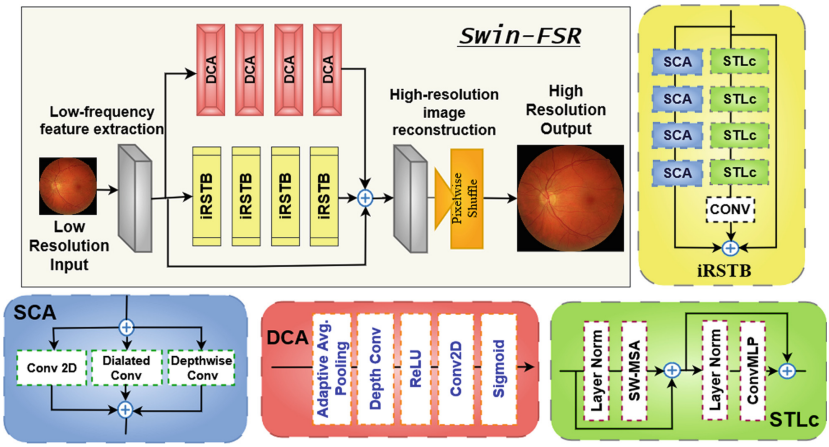

Fig. 1. An overview of the proposed Swin-FSR consisting of a low-frequency feature extraction module, improved Residual Swin-transformer BLock (iRSTB), Depth-wise Channel Attention (DCA) Block, and High-resolution image reconstruction block. Furthermore, iRSTB block consists of three parallel branches, i) Swin-Transformer with ConvMLP block (STLc), ii) Spatial and Channel Attention (SCA) Block, and iii) an identity mapping of the input, which is added together.

of data to urban healthcare facilities where specialists can make more accurate diagnoses [3,21]. Compressing and decompressing the data without losing spatial information can be utilized in this scenario by the super-resolution algorithm.

In a similar manner, color funduscopy imaging has found applications beyond the confines of the planet. For example, astronauts onboard the International Space Station (ISS) utilize this imaging to identify spaceflight-associated neuro-ocular syndrome (SANS) [10]. This condition can occur due to prolonged exposure to microgravity. It affects astronauts during long-duration spaceflight missions and can present with asymmetric/unilateral/bilateral optic disc edema and choroidal folds, which are easily identifiable by Color fundus images [11]. With the low bandwidth communication between ISS and terrestrial station [19], it becomes harder for experts to conduct early diagnosis and take preventive measures. So, super-resolution techniques can be vital in this adverse scenario. Medical experts can visualize and analyze these changes and act accordingly [18].

Image super-resolution and compression using different upsampling filters [6,20,23], has been a staple in image processing for a long time. Yet, those conventional approaches required manually designing convolving filters, which couldn't adapt to learning spatial and depth features and had less artifact removal ability. With the advent of deep learning, convolutional neural network-based super-resolution has paved the way for fast and low-computation image reconstructions with less error [4,25,26]. A few years back, attention-based architectures [17,22,27,28] were the state-of-the-art for any image super-resolution tasks. However, with the introduction of shifted window-based transformer models [8,13], the accuracy of these models superseded attention-based

ones with regard to different reconstruction metrics. Although swin-transformer-based models are good at extracting features of local patches, they lose the overall global spatial and depth context while upsampling with window-based patch merging operations.

Our Contributions: Considering all the relevant factors, we introduce the novel Swin-FSR architecture. It incorporates low-frequency feature extraction, deep feature extraction, and high-quality image reconstruction modules. The low-frequency feature extraction module employs a convolution layer to extract low-level features and is then directly passed to the reconstruction module to preserve low-frequency information. Our novelty is introduced in the deep feature extraction where we incorporated Depth-wise Channel Attention block (DCA), improved Residual Swin-transformer Block - (iRSTB), and Spatial and Channel Attention block (SCA). To validate our work, we compare three different SR architectures for four Fundus datasets: iChallenge-AMD [5], iChallenge-PALM [7], G1020 [1] and SANS. From Fig. 2 and Fig. 3 , it is apparent that our architecture reconstructs images with high PSNR and SSIM.

2 Methodology

2.1 Overall Architecture

As shown in Fig. 1, SwinFSR consists of four modules: low-frequency feature extraction, deep patch-level feature extraction, deep depth-wise channel attention, and high-quality (HQ) image reconstruction modules. Given a low-resolution (LR) image $I_{LR} \epsilon \mathbb{R}^{H \times W \times C}$. Here, H = height, W = width, c = channel of the image. For low-frequency feature extraction, we utilize a 3×3 convolution with stride= 1 and is denoted as H_{LF}, and it extracts a feature $F_{LF} \epsilon \mathbb{R}^{H \times W \times C_o ut}$ with which illustrated in Eq. 1.

$$F_{LF} = H_{LF}(I_{LR}) \tag{1}$$

It has been reported that the convolution layer helps with better spatial feature extraction and early visual processing, guiding to more steady optimization in transformers [24]. Next, we have two parallel branches of outputs as denoted in Eq. 2. Here, H_{DCA} and H_{iRSTB} are two new blocks that we propose in this study, and they both take the low-feature vector as F_{LF} input and generate two new features namely, F_{SF} and F_{PLF}. We elaborate this two blocks in Subsect. 2.2 and 2.3

$$\begin{aligned} F_{SF} &= H_{DCA}(F_{LF}) \\ F_{PLF} &= H_{iRSTB}(F_{LF}) \end{aligned} \tag{2}$$

Finally, we combine all three features from our previous modules, namely, F_{LF}, F_{SF}, and F_{PLF} and apply a final high-quality (HQ) image reconstruction module to generate a high-quality image I_{HQ} as given in Eq. 3.

$$I_{HQ} = H_{REC}(F_{LF} + F_{SF} + F_{PLF}) \tag{3}$$

where H_{REC} is the function of the reconstruction block. Our low-frequency block extracts shallow features, whereas the two parallel depth-wise channel-attention and

improved residual swin-transformer blocks extract spatially and channel-wise dense features extracting lost high-frequencies. With these three parallel residual connections, SwinFSR can propagate and combine the high and low-frequency information to the reconstruction module for better super-resolution results. It should be noted that the reconstruction module consists of a 1×1 convolution followed by a Pixel-shuffle layer to upsample the features.

2.2 Depth-Wise Channel Attention Block

For super-resolution architectures, channel-attention [2,15,28] is an essential robust feature extraction module that helps these architectures achieve high accuracy and more visually realistic results. In contrast, recent shifted-window-based transformers architecture [12,13] for super-resolution do not incorporate this module. However, a recent work [8] utilized a cross-attention module after the repetitive swin-transformer layers. One of the most significant drawbacks of the transformer layer is it works on patch-level tokens where the spatial dimensions are transformed into a linear feature. To retain the spatial information intact and learning dense features effectively, we propose depth-wise channel attention given in Eq. 4.

$$\begin{aligned} x &= AdaptiveAvgPool(x_{in}) \\ x &= \delta(Depthwise\_Conv(x)) \\ x_{out} &= \phi(Conv(x)) \end{aligned} \tag{4}$$

Here, δ is ReLU activation, and ϕ is Sigmoid activation functions. The regular channel-attention utilizes a 2D Conv with $1 \times 1 \times C$ weight vector C times to create output features $1 \times 1 \times C$. Given that adaptive average pooling already transforms the dimension to 1×1, utilizing a spatial convolution is redundant and shoots up computation time. To make it more efficient, we utilize depth-wise attention, with 1×1 weight vector applied on each of the C features separately, and then the output is concatenated to get our final output $1 \times 1 \times C$. We use four **DCA** blocks in our architecture as given in Fig. 1.

2.3 Improved Residual Swin-Transformer Block

Swin Transformer [16] incorporates shifted windows self-attention (SW-MSA), which builds hierarchical regional feature maps and has linear computation complexity compared to vision transformers with quadratic computation complexity. Recently, Swin-IR [13] adopted a modified swin-transformer block for different image enhancement tasks such as super-resolution, JPEG compression, and denoising while achieving high PSNR and SSIM scores. The most significant disadvantage of this block is the Multi-layer perceptron module (MLP) after the post-normalization layer, which has two linear (dense) layers. As a result, it becomes computationally more expensive than a traditional 1D convolution layer. For example, a linear feature output from a swin-transformer layer having depth D, and input channel, X_{in} and output channel, X_{out} will have a total number of parameters, $D \times X_{in} \times X_{out}$. Contrastly, a 1D convolution with kernel size, $K = 1$, with the same input and output will have less number of parameters,

$1 \times X_{in} \times X_{out}$. Here, we assign bias, $b = 0$. So, the proposed swin-transformer block can be defined as Eq. 5 and is illustrated in Fig. 1 as **STLc** block.

$$\begin{aligned} x^1 &= SW\text{-}MSA(\sigma(d^l)) + x \\ x^2 &= ConvMLP(\sigma(x^1)) + x^1 \end{aligned} \tag{5}$$

Here, σ is Layer-normalization and ConvMLP has two 1D convolution followed by GELU activation. To capture spatial local contexts for patch-level features we utilize a patch-unmerging layer in parallel path and incorporate **SCA** (spatial and channel attention) block. The block consists of a convolution ($k = 1, s = 1$), a dilated convolution ($k = 3, d = 2, s = 1$) and a depth-wise convolution ($k = 1, s = 1$) layer. Here, k= kernel, d =dilation and s =stride. Moreover all these features are combined to get the final ouptut. By combining repetitive SCA, $STLc$ blocks and a identity mapping we create our improved reisdual swin-transformer block (**iRSTB**) illustrated in Fig. 1. In Swin-FSR, we incorporate four iRSTB blocks.

2.4 Loss Function

For image super resolution task, we utilize the L1 loss function given in Eq. 6. Here, I_{RHQ} is the reconstructed output of SwinFSR and I_{HQ} is the original high-quality image.

$$L = \|I_{RHQ} - I_{HQ}\| \tag{6}$$

3 Experiments

3.1 Dataset

To assess the performance of our super-resolution models, we employ three distinct public fundus datasets: AMD [5], G1020 [1], PALM [7,14], and one private dataset: SANS. The datasets comprise .jpg, .tif, and .png formats with a high resolution. For training purposes, we used Bicubic Interpolation to resize the images into (512×512) and converted all the images into .png format. The AMD, G1020, PALM, and SANS datasets yield 400, 1020, 400, and 276 images, respectively. We split every dataset in 80% train and 20% test set, so we end up having 320 and 80 images for AMD, 816 and 204 images for G1020, 320 and 80 images for PALM, and 220 and 56 images for SANS. We use 5-fold cross-validation to train our networks.

Data use Declaration and Acknowledgment: The AMD and PALM dataset were released as part of REFUGE Challenge, PALM Challenge. The G1020 was published as technical report and benchmark [1]. The authors instructed to cite their work [5,7,14] for usage. The SANS data is privately held and is provided by the National Aeronautics and Space Administration(NASA) with Data use agreement 80NSSC20K1831.

3.2 Hyper-parameter

We utilized L1 loss for training our models for the super-resolution task. For optimizer, we used Adam [9], with learning rate $\alpha = 0.0002$, $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The batch size was $b = 2$, and we trained for 200 epochs for 8 h with NVIDIA A30 GPU. We utilize PyTorch and MONAI library monai.io for data transformation, training and testing our model. The code repository is provided in this link.

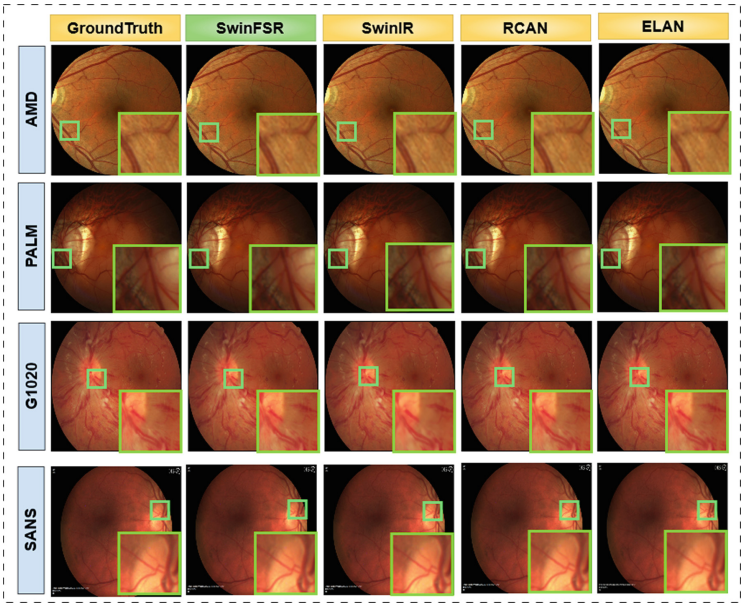

Fig. 2. Qualitative comparison of ($\times 2$) image reconstruction using different SR methods on AMD, PALM, G1020 and SANS dataset. The green rectangle is the zoomed-in region. The rows are for the AMD, PALM and SANS datasets. Whereas, the column is for each different models: SwinFSR, SwinIR, RCAN and ELAN. (Color figure online)

3.3 Qualitative Evaluation

We compared our architecture with some best-performing CNN and Transformer based SR models, including RCAN [28], ELAN [27], and SwinIR [13] as illustrated in Fig. 2 and Fig. 3. We trained and evaluated all four architectures using their publicly available source code on the four datasets. SwinIR utilizes residual swin-transformer blocks with identity mapping for dense feature extractions. In contrast, the RCAN utilizes repetitive channel-attention blocks for depth-wise dense feature retrieval. Similarly, ELAN combines multi-scale and long-range attention with convolution filters to extract spatial and depth features. In Fig. 2, we illustrate $\times 2$ reconstruction results for all four architectures. By observing, we can see that our model's vessel reconstruction is more realistic for $\times 2$ factor samples than other methods. Specifically for AMD and SANS, the

degeneration is noticeable. In contrast, ELAN and RCAN fail to accurately reconstruct thinner and smaller vessels.

In the second experiment, we show results for $\times 4$ reconstruction for all SR models in Fig. 3. It is apparent from the figure that our model's reconstruction is more realist than other transformer and CNN-based architectures, and the vessel boundary is sharp, containing more degeneration than SwinIR, ELAN and RCAN.. Especially for AMD , G1020 and PALM, the vessel edges are finer and sharper making it easily differentiable. In contrast, ELAN and RCAN generate pseudo vessels whereas SwinIR fails to generate some smaller ones. For SANS images, the reconstruction is much noticable for the $\times 4$ than $\times 2$.

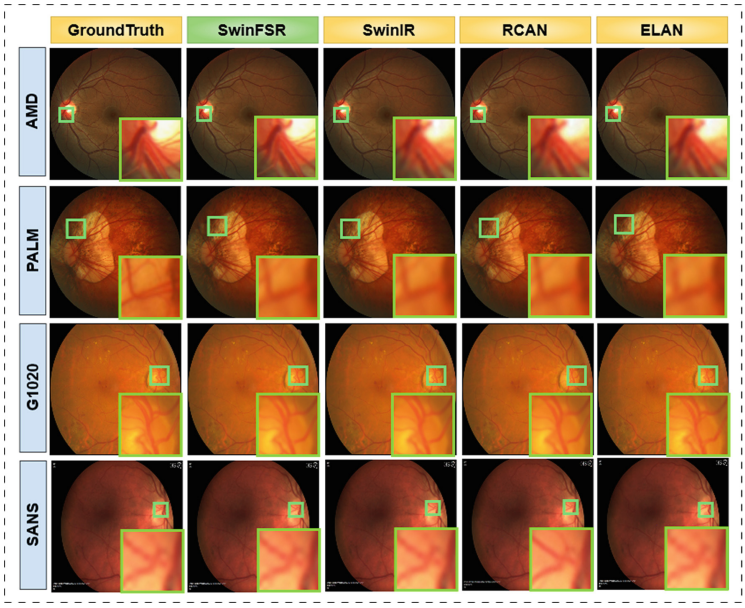

Fig. 3. Qualitative comparison of ($\times 4$) image reconstruction using different SR methods on AMD, PALM, G1020 and SANS dataset. The green rectangle is the zoomed-in region. The rows are for the AMD, PALM and SANS datasets. Whereas, the column is for each different models: SwinFSR, SwinIR, RCAN and ELAN. (Color figure online)

3.4 Quantitative Bench-Marking

For quantitative evaluation, we utilize Peak Signal-to-Noise-Ratio (PSNR) and Structural Similarity Index Metric (SSIM), which has been previously employed for measuring similarity between original and reconstructed images in super-resolution tasks [13,27,28]. We illustrate quantitative performance in Table. 1 between SwinFSR and other state-of-the-art methods: SwinIR [13], RCAN [28], and ELAN [27]. Table. 1 shows that SwinFSR's overall SSIM and PSNR are superior to other transformer and

CNN-based approaches. For $\times 2$ scale reconstruction, SwinIR achieves the second-best performance. Contrastly, for $\times 4$ scale reconstruction, RCAN outperforms SwinIR while scoring lower than our SwinFSR model for PSNR and SSIM.

Table 1. Quantitative comparison on AMD [5], PALM [7,14], G1020 [1], & SANS.

Dataset		AMD		PALM		G1020		SANS	
2X									
Model	Year	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
SwinFSR	2023	**98.70**	**47.89**	**99.11**	**49.00**	**98.65**	**49.11**	**97.93**	**45.32**
SwinIR	2022	98.68	47.78	99.03	48.73	98.59	48.94	97.92	45.17
ELAN	2022	98.18	44.21	98.80	46.49	98.37	47.48	96.89	36.84
RCAN	2018	98.62	47.76	99.04	48.83	98.53	48.29	97.91	45.29
4X									
Model	Year	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
SwinFSR	2023	**96.51**	**43.28**	**97.34**	**43.27**	**97.13**	**44.67**	**95.82**	**39.14**
SwinIR	2022	96.40	42.98	97.27	43.07	97.06	44.44	95.80	39.02
ELAN	2022	94.76	39.12	97.03	42.47	97.11	44.39	95.16	35.84
RCAN	2018	96.20	42.62	97.38	43.18	97.05	44.37	95.73	38.92

3.5 Clinical Assessment

We carried out a diagnostic assessment with two expert ophthalmologists and test samples of 80 fundus images (20 fundus images per disease classes: AMD, Glaucoma, Pathological Myopia and SANS for both original x2 and x4 images, and super-resolution enhanced images). Half of the 20 fundus images were control patients without disease pathologies; the other half contained disease pathologies. The clinical experts were not provided any prior pathology information regarding the images. And each of the experts was given 10 images with equally distributed control and diseased images for each disease category.

The accuracy and F1-score for original x4 images are as follows, 70.0% and 82.3% (AMD), 75% and 85.7% (Glaucoma), 60.0% and 74.9% (Palm), and 55% and 70.9% (SANS). The accuracy and F1-score for original x2 are as follows, 80.0% and 88.8% (AMD), 80% and 88.8% (Glaucoma), 70.0% and 82.1% (Palm), and 65% and 77.4% (SANS). The accuracy and F1-score for our model Swin-FSR's output from $\times 4$ images are as follows, 90.0% and 93.3% (AMD), 90.0% and 93.7% (Glaucoma), 75.0% and 82.7% (Palm), and 75% and 81.4% (SANS). The accuracy and F1-score for Swin-FSR's output from $\times 2$ images are as follows, 90.0% and 93.3% (AMD), 90.0% and 93.7% (Glaucoma), 80.0% and 85.7% (Palm), and 80% and 85.7% (SANS).

We also tested SWIN-IR, ELAN, and RCAN models for diagnostic assessment, out of which SWIN-IR upsampled images got the best results. For x4 images, the model's

accuracy and F-1 score are 80% and 87.5% (AMD), 85.0% and 90.3% (Glaucoma), 70.0% and 80.0% (Palm), and 70% and 76.9% (SANS). For x2 images, the model's accuracy and F-1 score are 80% and 87.5% (AMD), 80% and 88.8% (Glaucoma), 70.0% and 80.0% (Palm), and 75% and 81.4% (SANS). Based on the above observations, our model-generated images achieves the best result.

3.6 Ablation Study

Effects of iRSTB, DCA, and SCA Number: We illustrate the impacts of iRSTB, DCA, and SCA numbers on the model's performance in Supplementary Fig. 1 (a), (b), and (C). We can see that the PSNR and SSIM become saturated with an increase in any of these three hyperparameters. One drawback is that the total number of parameters grows linearly with each additional block. Therefore, we choose four blocks for iRSTB, DCA, and SCA to achieve the optimum performance with low computation cost.

Presence and Absence of iRSTB, DCA, and SCA Blocks: Additionally, we provide a comprehensive benchmark of our model's performance with and without the novel blocks incorporated in Supplementary Table 1. Specifically, we show the performance gains with the usage of an improved residual swin-transformer block (iRSTB) and depth-wise channel attention (DCA). As the results illustrate, by comprising these blocks, the PSNR and SSIM reach higher scores.

4 Conclusion

In this paper, we proposed Swin-FSR by combining novel DCA, iRSTB, and SCA blocks which extract depth and low features, spatial information, and aggregate in image reconstruction. The architecture reconstructs the precise venular structure of the fundus image with high confidence scores for two relevant metrics. As a result, we can efficiently employ this architecture in various ophthalmology applications emphasizing the Space station. This model is well-suited for the analysis of retinal degenerative diseases and for monitoring future prognosis. Our goal is to expand the scope of this work to include other data modalities.

Acknowledgement. Research reported in this publication was supported in part by the National Science Foundation by grant numbers [OAC-2201599],[OIA-2148788] and by NASA grant no 80NSSC20K1831.

References

1. Bajwa, M.N., Singh, G.A.P., Neumeier, W., Malik, M.I., Dengel, A., Ahmed, S.: G1020: a benchmark retinal fundus image dataset for computer-aided glaucoma detection. In: 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1–7. IEEE (2020)
2. Dai, T., Cai, J., Zhang, Y., Xia, S.T., Zhang, L.: Second-order attention network for single image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11065–11074 (2019)

3. Das, V., Dandapat, S., Bora, P.K.: A novel diagnostic information based framework for super-resolution of retinal fundus images. Comput. Med. Imaging Graph. **72**, 22–33 (2019)
4. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2015)
5. Fu, H., et al.: Adam: automatic detection challenge on age-related macular degeneration. In: IEEE Dataport (2020)
6. Hardie, R.: A fast image super-resolution algorithm using an adaptive wiener filter. IEEE Trans. Image Process. **16**(12), 2953–2964 (2007)
7. Huazhu, F., Fei, L., José, I.: PALM: pathologic myopia challenge. Comput. Vis. Med. Imaging (2019)
8. Jin, K., et al.: SwiniPASSR: Swin transformer based parallax attention network for stereo image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 920–929 (2022)
9. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
10. Lee, A.G., Mader, T.H., Gibson, C.R., Brunstetter, T.J., Tarver, W.J.: Space flight-associated neuro-ocular syndrome (SANS). Eye **32**(7), 1164–1167 (2018)
11. Lee, A.G., et al.: Spaceflight associated neuro-ocular syndrome (SANS) and the neuro-ophthalmologic effects of microgravity: a review and an update. NPJ Microgravity **6**(1), 7 (2020)
12. Li, B., Li, X., Lu, Y., Liu, S., Feng, R., Chen, Z.: HST: hierarchical swin transformer for compressed image super-resolution. In: Karlinsky, L., Michaeli, T., Nishino, K. (eds.) ECCV 2022. LNCS, vol. 13802, pp. 651–668. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-25063-7_41
13. 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)
14. Lin, F., et al.: Longitudinal changes in macular optical coherence tomography angiography metrics in primary open-angle glaucoma with high myopia: a prospective study. Invest. Ophthalmol. Vis. Sci. **62**(1), 30–30 (2021)
15. Lin, Z., et al.: Revisiting RCAN: improved training for image super-resolution. arXiv preprint arXiv:2201.11279 (2022)
16. Liu, Z., et al.: Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022 (2021)
17. Niu, B., et al.: Single Image Super-Resolution via a Holistic Attention Network. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12357, pp. 191–207. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58610-2_12
18. Ong, J., et al.: Neuro-ophthalmic imaging and visual assessment technology for spaceflight associated neuro-ocular syndrome (SANS). Survey Ophthalmol. **67**, 1443–1466 (2022)
19. Seas, A., Robinson, B., Shih, T., Khatri, F., Brumfield, M.: Optical communications systems for NASA's human space flight missions. In: International Conference on Space Optics-ICSO 2018, vol. 11180, pp. 182–191. SPIE (2019)
20. Sen, P., Darabi, S.: Compressive image super-resolution. In: 2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers, pp. 1235–1242. IEEE (2009)
21. Sengupta, S., Wong, A., Singh, A., Zelek, J., Lakshminarayanan, V.: DeSupGAN: multi-scale feature averaging generative adversarial network for simultaneous de-blurring and super-resolution of retinal fundus images. In: Fu, H., Garvin, M.K., MacGillivray, T., Xu, Y., Zheng, Y. (eds.) OMIA 2020. LNCS, vol. 12069, pp. 32–41. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-63419-3_4

22. Song, X., et al.: Channel attention based iterative residual learning for depth map super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5631–5640 (2020)
23. Van Ouwerkerk, J.: Image super-resolution survey. Image Vis. Comput. **24**(10), 1039–1052 (2006)
24. Xiao, T., Singh, M., Mintun, E., Darrell, T., Dollár, P., Girshick, R.: Early convolutions help transformers see better. In: Advances in Neural Information Processing Systems, vol. 34, pp. 30392–30400 (2021)
25. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Trans. Image Process. **26**(7), 3142–3155 (2017)
26. Zhang, K., Zuo, W., Zhang, L.: Learning a single convolutional super-resolution network for multiple degradations. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3262–3271 (2018)
27. Zhang, X., Zeng, H., Guo, S., Zhang, L.: Efficient long-range attention network for image super-resolution. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) ECCV 2022. LNCS, vol. 13677, pp. 649–667. Springer, Cham (2022)
28. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 286–301 (2018)
29. Zhang, Z., et al.: A survey on computer aided diagnosis for ocular diseases. BMC Med. Inform. Decis. Mak. **14**(1), 1–29 (2014)