Pacific Graphics (2023)
R. Chaine, Z. Deng, and M. H. Kim (Editors)

*Poster*

# Combining Transformer and CNN for Super-Resolution of Animal Fiber Microscopy Images

Jiagen Li[1], Yatu Ji[†1], Min Lu[1], Li Wang[2], Lingjie Dai[2], Xuanxuan Xu[1], Nier Wu[1] and Na Liu[1].

[1]Inner Mongolia University of Technology, China
[2]National Testing Center of Wool and Cashmere Quality, China

**Abstract**

*The images of cashmere and wool fibers used for scientific research in the textile field are mostly acquired manually under an optical microscope. However, due to the interference of microscope quality, shooting environment, focal length selection, acquisition techniques and other factors, the quality of the obtained photographs tends to have a low resolution, and it is difficult to display the fine fiber texture structure and scale details. To address the above problems, a lightweight super-resolution reconstruction algorithm with multi-scale hierarchical screening is proposed. Specifically, firstly, a hybrid module incorporating SwinTransformer and enhanced channel attention is proposed to extract the global features and obtain the important localization among them, in addition, a multi-scale hierarchical screening filtering module is proposed based on the residual model, which amplifies the feature information focusing on high-frequency regions by splitting the channels to allow the model to adaptively weight the features according to the feature weights and amplifies the feature information focusing on high-frequency regions. Finally, the global average pooling attention module integrates and weights the high-frequency features again to enhance details such as edges and textures. A large number of experiments show that compared with other state-of-the-art algorithms, the proposed method significantly improves the image quality on the fiber dataset, and at the same time proves the effectiveness of the proposed method at all scales in five public datasets, occupies less memory parameters than SwinIR, and not only improves the PSNR and SSIM, but also reduces the parameters compared with the light-weight ESRT.*

**CCS Concepts**

• ***Computing methodologies*** → ***Computer graphics;*** *Image processing;*

## 1. Introduction

With the rapid development of deep learning, many hyper-segmentation models based on convolutional architectures such as DCRN [KLL16]and CARN [AKS18] have been proposed, but their feature extraction can only be performed locally, ignoring long-range dependencies, which is unacceptable for fibrous images with large cross-sectional length comparisons. Attempts to use the Transformer architecture gradually became the current hotspot, and SwinIR [LCS*21] achieved the best performance at that time, but the large computational and memory footprint made it difficult to deploy on mobile terminals.ESRT [LLL*22]was the first hypersegmentation model to combine Transformer and CNN and achieved a trade-off between performance and parameter footprint, but its focus on small targets was not sufficient adequately. In order to solve the situation faced and for the characteristics of fiber images, We try to use the Enconder part of Swin Transformer to extract long distance dependencies of local features, and then design the convolution module to integrate and filter these features.

## 2. Network Design

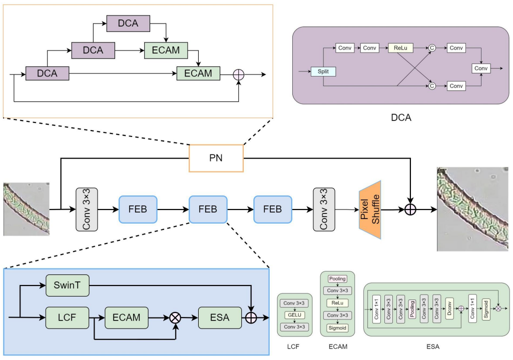

**Figure 1:** *Lightweight super-resolution reconstruction model for multiscale hierarchical screening*

The overall model is composed of 3 stacked FEB modules and the residual PN module, the internal structure of each module can be found in in Figure1, where the PN and DCA modules perform

[†] Corresponding author, email adress: mljyt@imut.edu.cn

DOI: 10.2312/pg.20231283

**Table 1:** *Quantitative comparison on five public benchmark datasets*

| Scale | Method | Params | Set5 | | Set14 | | B100 | | Urban100 | | Manga109 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | [k] | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| ×2 | **Ours** | 468 | 38.01 | 0.9611 | 33.75 | 0.9206 | 32.22 | 0.9010 | 32.29 | 0.9299 | 38.90 | 0.9777 |
| ×3 | **Ours** | 475 | 34.50 | 0.9284 | 30.46 | 0.8448 | 29.16 | 0.8082 | 28.50 | 0.8590 | 33.98 | 0.9466 |
| ×4 | **Ours** | 484 | 32.28 | 0.8964 | 28.74 | 0.7849 | 27.69 | 0.7407 | 26.57 | 0.7995 | 31.02 | 0.9152 |

the extraction of edge and detail texture features at different scales and further fusion to filter the useful features, and the combination of the SwinT and ESA modules pays better attention to both the global and the local, and the overall process has been explained in the abstract.

## 3. Experiments

In this work, we use a homemade fiber image dataset with a total of 1000 images in the training set including cashmere and wool, a test set of 500 images and no crossover, an optimizer of Adam, an initial learning rate of 0.001, and a total of 1 million iterations. Nvidia 3080 and Pytorch were used for training and testing.

**Table 2:** *Quantitative comparison of different algorithms on fiber image datasets at different scales*

| Scale | Method | Params | Wool | | Cashmere | |
|---|---|---|---|---|---|---|
| | | [k] | PSNR | SSIM | PSNR | SSIM |
| ×3 | FSRCNN | 13 | 33.10 | 0.8875 | 33.16 | 0.9095 |
| | ESRT | 770 | 34.73 | 0.9122 | 34.79 | 0.9138 |
| | OURS | 475 | 34.90 | 0.9224 | 34.85 | 0.9244 |
| ×4 | FSRCNN | 13 | 30.73 | 0.8541 | 30.74 | 0.8543 |
| | ESRT | 751 | 32.74 | 0.8847 | 32.75 | 0.8848 |
| | OURS | 484 | 32.81 | 0.8867 | 32.82 | 0.8872 |

Table 2 demonstrates the superiority of our algorithm on the fiber dataset, which further reduces the parameters compared to the lightweight ESRT and improves the performance marginally, where cashmere gets relatively better results compared to wool due to the regularity of its texture. Meanwhile, in order to verify the generality of the model, we use the DIV2K dataset to train the model and test the model on five public datasets, as shown in Table 1, compared with other existing models, the proposed model has a better performance at the ×2 ×3 ×4 scale.

As can be seen from Figure 2, compared to other reconstruction algorithms, the skeleton and texture parts of the image obtained by the proposed algorithm are more obvious, and the small target features are better recovered.

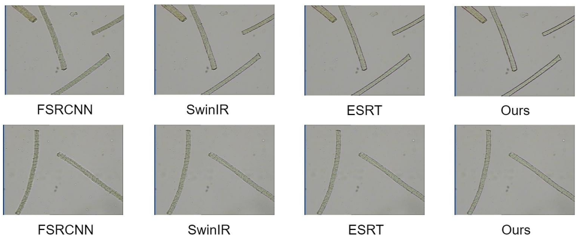

**Figure 2:** *Comparison of reconstruction effects of different models at Turkish cashmere and small-tailed frigid sheep hair × 4 scale*

## 4. Conclusion

The experiments demonstrate the effectiveness of the proposed model on both fiber and public datasets, producing a better balance between parameters and performance. The next step is to optimize the model and add new modules to further improve the quality of the reconstructed images and use them for subsequent advanced vision tasks.

## 5. Acknowledgements

This study is supported by the National Natural Science Foundation of China (62206138), Education Department Science Research Foundation of Inner Mongolia Autonomous Region (JY20220186, NJZZ23081, RZ2300001743), Support Program for Young Scientific and Technological Talents in Inner Mongolia Colleges and Universities (NJYT23059), Inner Mongolia Science and Technology Program Project (2020GG0104), Special Foundation for the Introduce Talents of Inner Mongolia Autonomous Region, China (DC2300001440, DC2300001441), Inner Mongolia Natural Science Foundation (2022LHMS06004)

## References

[AKS18] Ahn N., Kang B., Sohn K.-A.: Fast, accurate, and lightweight super-resolution with cascading residual network. In *Proceedings of the European conference on computer vision (ECCV)* (2018), pp. 252–268. doi:10.1007/978-3-030-01249-6_16. 1

[KLL16] Kim J., Lee J. K., Lee K. M.: Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 1637–1645. doi:10.1109/cvpr.2016.181. 1

[LCS*21] 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* (2021), pp. 1833–1844. doi:10.1109/iccvw54120.2021.00210. 1

[LLL*22] Lu Z., Li J., Liu H., Huang C., Zhang L., Zeng T.: Transformer for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2022), pp. 457–466. doi:10.1049/ipr2.12833. 1

Proceedings published by Eurographics - The European Association for Computer Graphics.