PG23

Daejeon, Korea

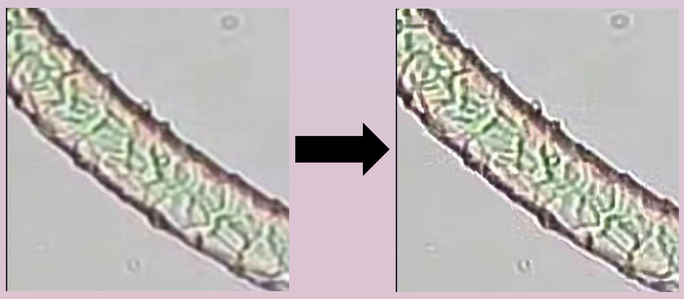
Combining Transformer and CNN for Super-Resolution of Animal Fiber Microscopy Images

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

PROBLEM

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 show the fine fiber texture structure and scale details. Therefore, super-resolution reconstruction of fiber images is of great practical significance.



RELATED WORK

With the rapid development of deep learning, many hyper-segmentation models based on convolutional architectures such as DCRN[1] and CARN[2] 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[3] achieved the best performance at that time, but the large computational and memory footprint made it difficult to deploy on mobile terminals.ESRT[4] 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.Therefore, we design a super-resolution reconstruction model for fiber dataset to address the above problems.

OVERVIEW

First, a hybrid module integrating SwinTransformer and enhanced channel and spatial attention is proposed to extract the global features and obtain the important localization among them, in addition, a multi-scale hierarchical screening filtering module based on the residual model is proposed to amplify the feature information focusing on high-frequency regions by splitting the channel to let the model adaptively weight according to the feature weights. 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.

METHODOLOGY

As shown in *Figure 1*, the FEB module: uses a shift window to model long-term dependencies and acquire global features based on the characteristics of SwinT, while the LCF module filters and extracts shallow local features, and the ECAM module extracts high-frequency information using weighted channel attention, discards low-frequency information and reduces the parameters, which is achieved by multiplying the input features with the attention weights in order to improve the model's perception of important features.The ESA module performs dimensionality reduction by 1×1 convolutional kernel and captures a larger range of contextual information in space using maximum pooling operation, where the convolutional layer and activation function process the features to capture and emphasize the important features so that the model focuses on the important regions in the input features. Finally the global features extracted by SwinT are fused;

the PN module is mainly used to extract the edges and detailed texture features of the fiber image, and the features are fused and extracted at different scales, and the DCA module splits the channel features into A and B, and the A channel's features have a larger sensory field to see more pixel information, and the B channel causes more shallow rough features to propagate to the deeper layers to be preserved, fusing AB and using attention to further filter useful features.This compensates for critical information that may be lost in the stacking process of the FEB module.

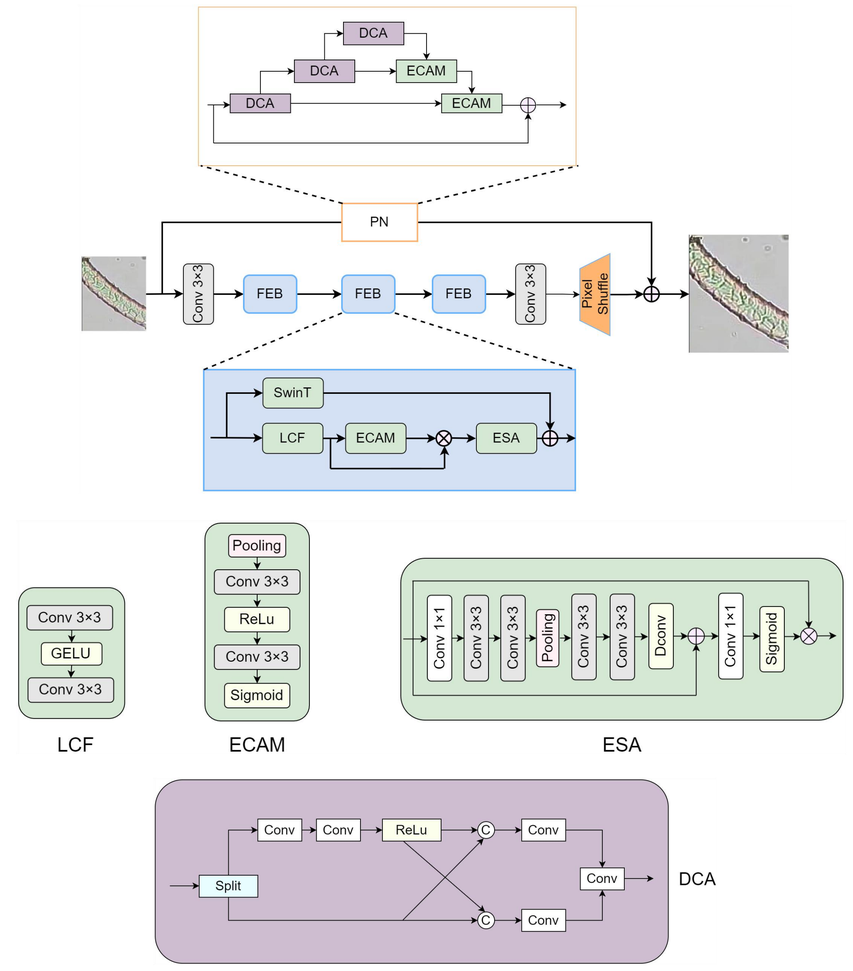


Figure 1:Lightweight super-resolution reconstruction model for multiscale hierarchical screening

RESULTS

Ablation experiments

The results in the *Table 1* show that the PN module using residual jump connections contributes significantly to the performance and parameters, whereas in the absence of the PN module, the number of parameters increases dramatically and the performance decreases. This suggests that the PN module not only maximizes the extraction of important features but also reduces the parameter footprint.Better performance can also be achieved using only the SwinT model, but its parameters are twice as large as those of our model.

Table 1:Impact of different components on model performance in fiber datasets

PN	ESA+ECAM	SwinT	Params	Wool		Cashmere	
			[k]	PSNR	SSIM	PSNR	SSIM
×	×	✓	868	37.90	0.9506	37.65	0.9510
×	✓	✓	705	36.97	0.9442	37.86	0.9488
✓	✓	✓	468	37.82	0.9523	37.95	0.9586

Comparison experiments

Table 2 demonstrates the superiority of our algorithm on the fiber dataset, where a small improvement in performance is obtained while maintaining a small parameter footprint, where cashmere gets relatively better results compared to wool due to the regularity of its texture. Specific comparison images are shown in *Figure 2* and *Figure 3*, where the skeleton and texture parts of the images obtained by our algorithm are more obvious, and the small target features are better recovered. Meanwhile, in order to verify the generality of the model, we use the DIV2K dataset to train the model and test the model in five public datasets, as shown in *Table 3*, compared with other existing models, the proposed model has a strong competitiveness in the ×2 ×3 ×4 scale.

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 3:Quantitative Comparison of Different Algorithms at Different Scales 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	**SRCNN**	8	36.66	0.9542	32.45	0.9067	31.36	0.8879	29.50	0.8946	35.60	0.9663
	FSRCNN	13	37.00	0.9558	32.63	0.9088	31.53	0.8920	29.88	0.9020	36.67	0.9710
	DCRN	1774	37.63	0.9588	33.04	0.9118	31.85	0.8942	30.75	0.9133	37.55	0.9732
	CARN	1592	37.76	0.9590	33.52	0.9166	32.09	0.8978	31.92	0.9256	38.36	0.9765
	SwinIR	878	**38.14**	**0.9611**	**33.86**	**0.9206**	**32.31**	**0.9012**	**32.76**	**0.9340**	**39.12**	**0.9783**
	Ours	468	38.01	**0.9611**	33.75	**0.9206**	32.22	0.9010	32.29	0.9299	38.90	0.9777
×3	**SRCNN**	8	32.75	0.9090	29.30	0.8215	28.41	0.7863	26.24	0.7989	30.48	0.9117
	FSRCNN	13	33.18	0.9140	29.37	0.8240	28.53	0.7910	26.43	0.8080	31.10	0.9210
	DCRN	1774	33.82	0.9226	29.76	0.8311	28.80	0.7963	27.15	0.8276	32.24	0.9343
	CARN	1592	34.29	0.9255	30.29	0.8407	29.06	0.8034	28.06	0.8493	33.50	0.9440
	SwinIR	886	**34.62**	**0.9289**	**30.54**	**0.8463**	**29.20**	**0.8082**	**28.66**	**0.8624**	33.98	**0.9478**
	ESRT	770	34.42	0.9268	30.43	0.8433	29.15	0.8063	28.46	0.8574	33.95	0.9455
	Ours	475	34.50	0.9284	30.46	0.8448	29.16	**0.8082**	28.50	0.8590	33.98	0.9466
×4	**SRCNN**	8	30.48	0.8626	27.50	0.7513	26.90	0.7101	24.52	0.7221	27.58	0.8555
	FSRCNN	13	30.72	0.8660	27.61	0.7550	26.98	0.7150	24.62	0.7280	27.90	0.8610
	DCRN	1774	31.53	0.8854	28.02	0.7670	27.23	0.7233	25.14	0.7510	28.93	0.8854
	CARN	1592	32.13	0.8937	28.60	0.7806	27.58	0.7349	26.07	0.7837	30.47	0.9084
	SwinIR	897	**32.44**	**0.8976**	**28.77**	**0.7858**	**27.69**	0.7406	26.47	0.7980	30.92	0.9151
	ESRT	751	32.19	0.8947	28.69	0.7833	**27.69**	0.7379	26.39	0.7962	30.75	0.9100
	Ours	484	32.28	0.8964	28.74	0.7849	**27.69**	**0.7407**	**26.57**	**0.7995**	**31.02**	**0.9152**

Figure 2:Comparison of reconstruction effects of different models at Turkish cashmere × 4 scale

Figure 3:Comparison of reconstruction effects of different models at small-tailed frigid sheep hair × 4 scale

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

[1] Kim J, Lee J K, Lee K M. Deeply-recursive convolutional network for image super-resolution[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1637-1645.
[2] Ahn N, Kang B, Sohn K A. Fast, accurate, and lightweight super-resolution with cascading residual network[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 252-268.
[3] Liang J, Cao J, Sun G, et al. Swinir: Image restoration using swin transformer[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 1833-1844.
[4] Lu Z, Li J, Liu H, et al. Transformer for single image super-resolution[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 457-466.