Spatial-Frequency Mutual Learning for Face Super-resolution

THU-PM-167

Chenyang Wang, Junjun Jiang*, Zhiwei Zhong and Xianming Liu.

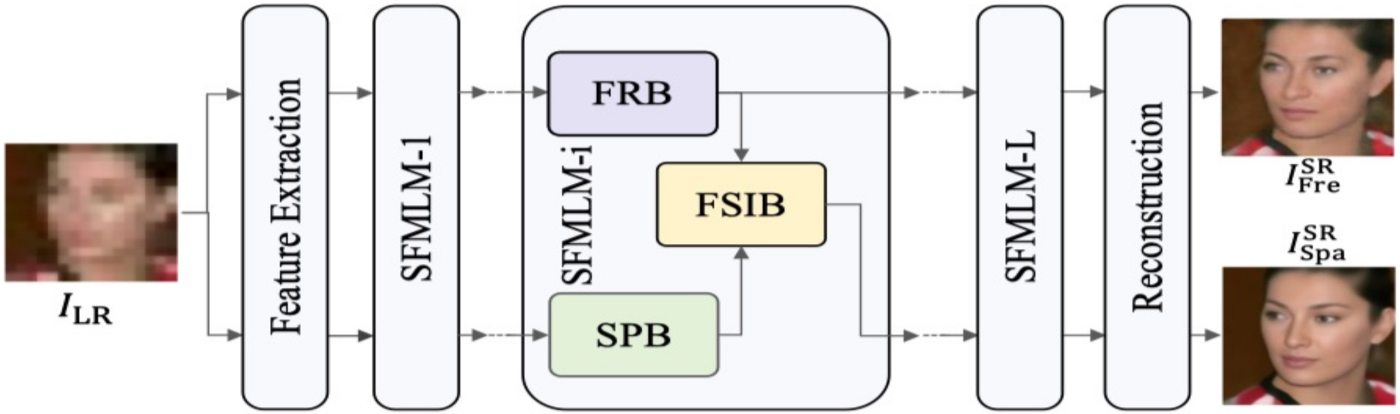
School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

{wangchy02,jiangjunjun,zhwzhong,csxm}@hit.edu.cn

哈爾濱工業大學 HARBIN INSTITUTE OF TECHNOLOGY

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

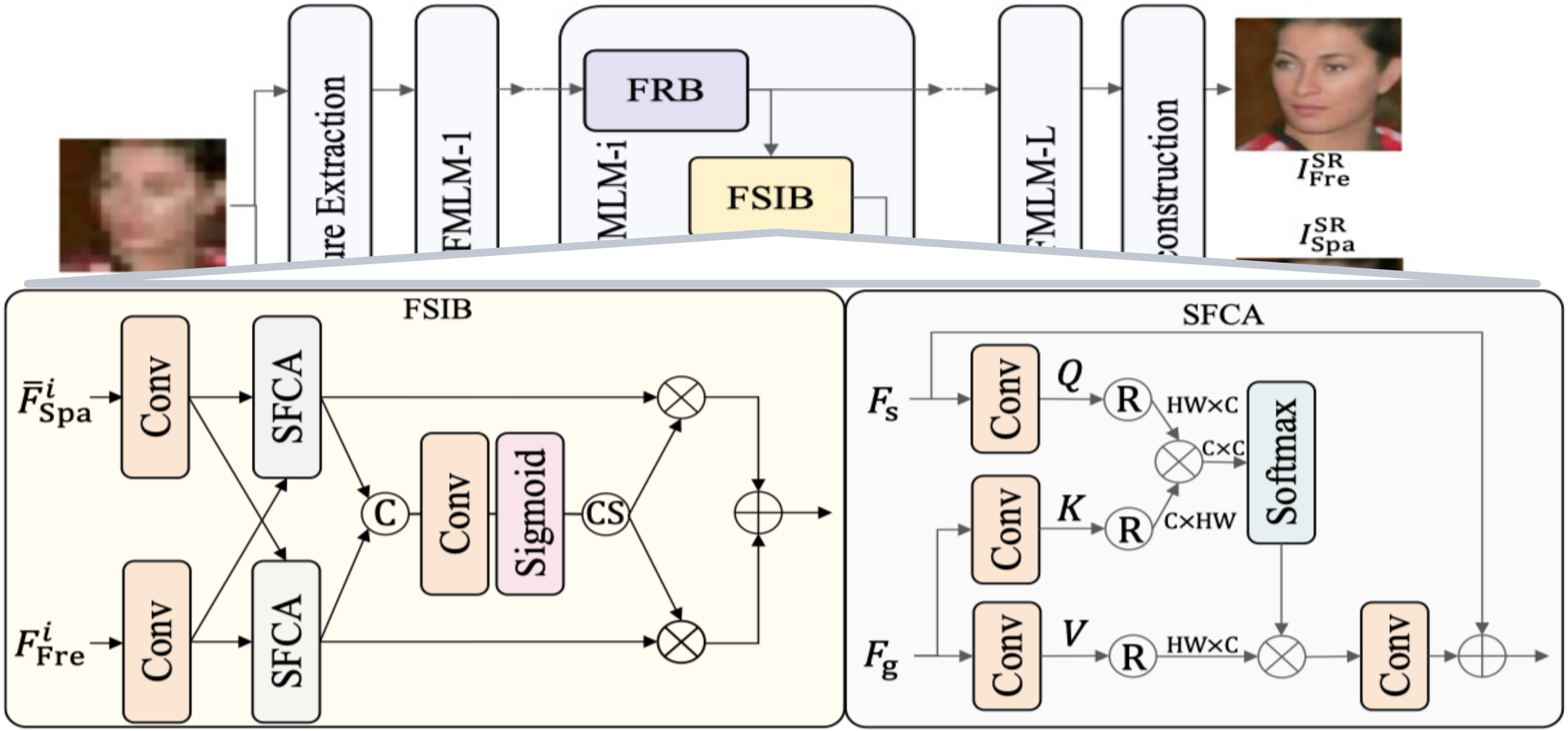
Overview



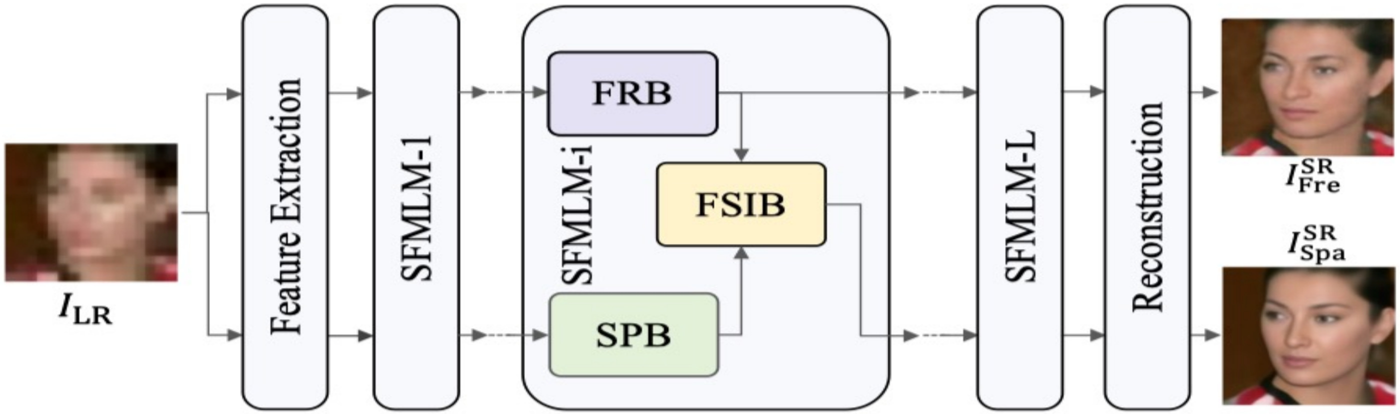
Overview of the proposed SFMNet.

- We develop a novel spatial-frequency mutual network (SFMNet) equipped with Fourier transform, which can not only **achieve image-size receptive field** but also **maintain facial structure**.
- This is the first method that explores the potential of both spatial and frequency information for face super-resolution.

Overview



◆ We carefully design a frequency-spatial interaction block to **mutually fuse global frequency information and local spatial information**.



Overview of the proposed SFMNet.

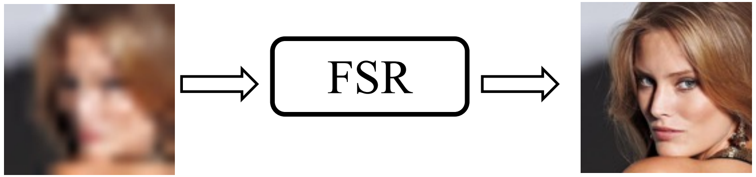
Experimental results demonstrate that our method achieves **the state-of-the-art performance** in terms of visual results and quantitative metrics.

- We carefully design a frequency-spatial interaction block to **mutually fuse global frequency information and local spatial information**.

Face Super-resolution (FSR)

◆ Face super-resolution(FSR)：

recovers high-resolution face image from the given low-resolution one.



LR SR

◆ FSR can：

➢ improve face image quality and provide pleasing visual experience

➢ boost downstream tasks, e.g., face recognition, face analysis, etc.

Method	Bicubic	Ma *et al.*	LapSRN	UR-DGN	SICNN
Identity Similarity	0.2913	0.3823	0.4361	0.3682	**0.5978**
LFW Acc	97.51%	97.58%	97.46%	97.20%	**98.25%**
YTF Acc	93.08%	93.26%	93.10%	92.78%	**93.82%**

Motivation

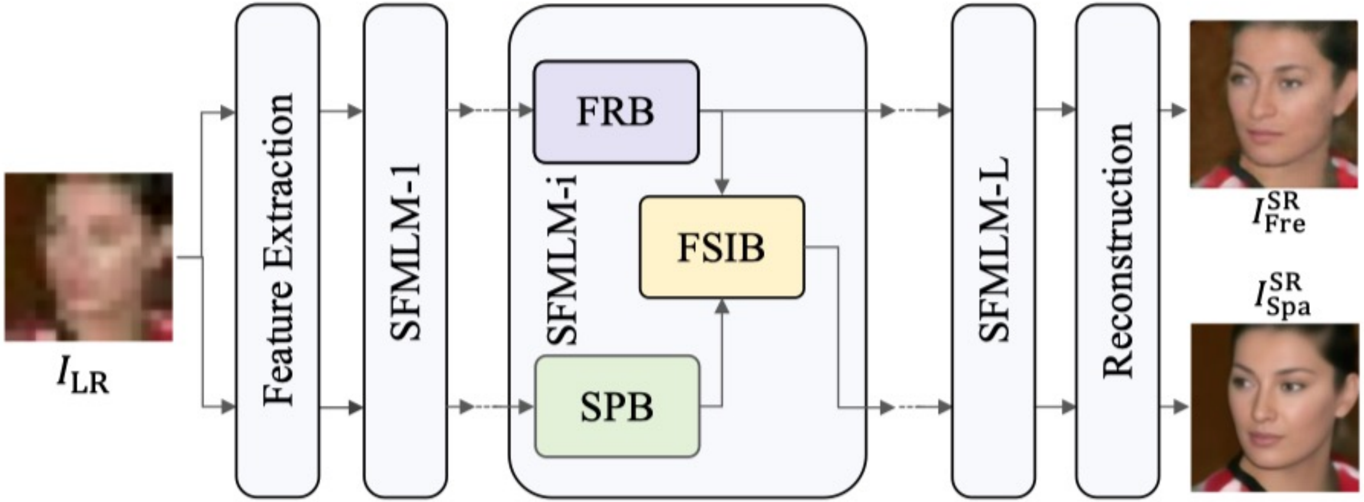
- Challenges of FSR
  - Limited receptive field.
  - Failure to maintain facial structure.
- Observation
  - Fourier transform can achieve image-size receptive field.

$$\mathcal{F}(x)(u,v)=\frac{1}{\sqrt{HW}}\sum_{h=0}^{H-1}\sum_{w=0}^{W-1}x(h,w)e^{-2j\pi\left(\frac{h}{H}u+\frac{w}{W}v\right)}$$

  - The phase component can well characterize facial structure.

Face images and the reconstructed results by phase component.

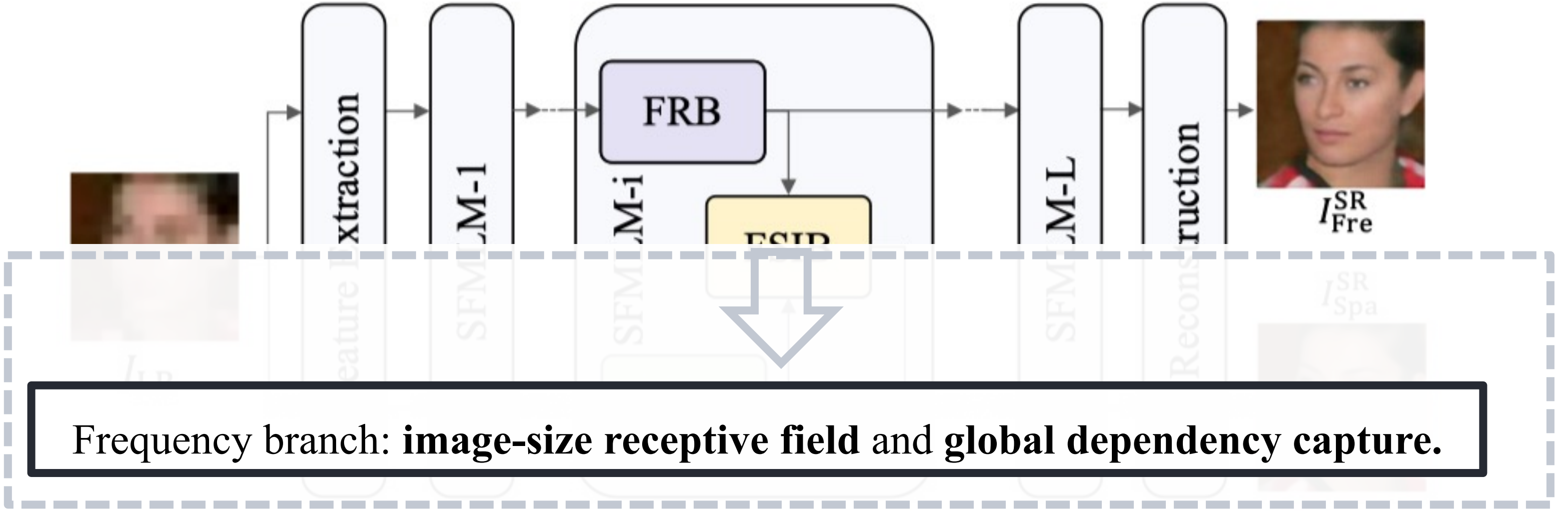
SFMNet



Overview of the proposed SFMNet.

- We develop a spatial-frequency mutual network (SFMNet) equipped with Fourier transform. To the best of our knowledge, this is the first method that explores the potential of both spatial and frequency information for face super-resolution.

SFMNet

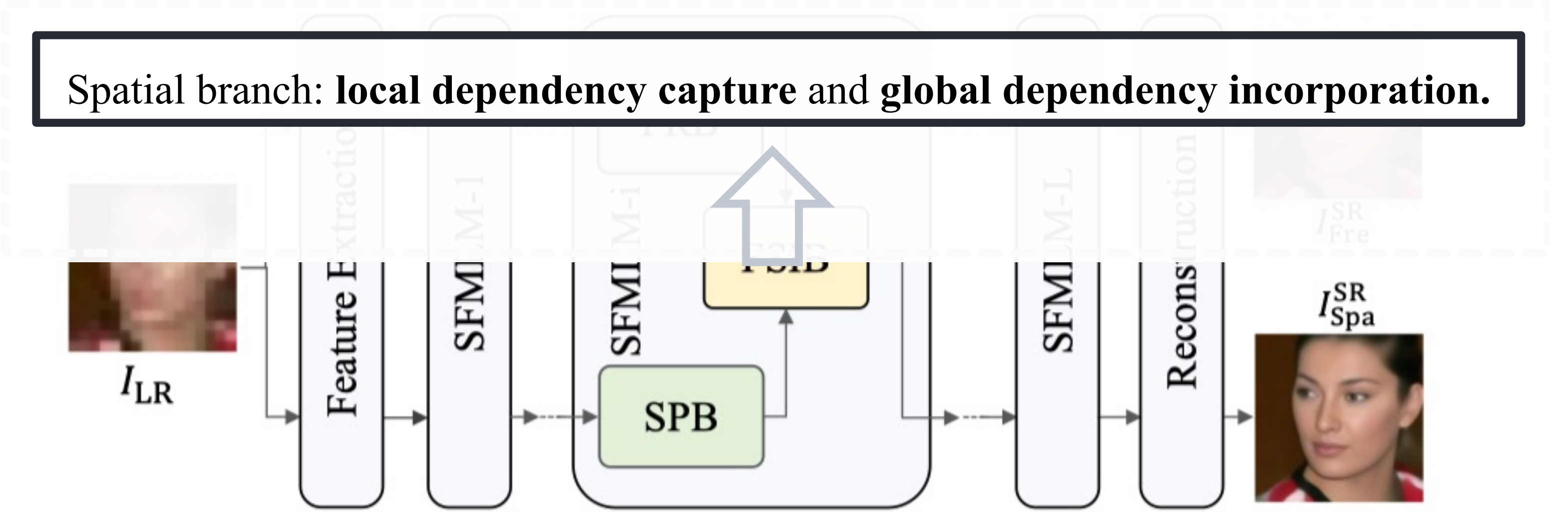


Overview of the proposed SFMNet.

- We develop a spatial-frequency mutual network (SFMNet) equipped with Fourier transform. To the best of our knowledge, this is the first method that explores the potential of both spatial and frequency information for face super-resolution.

SFMNet

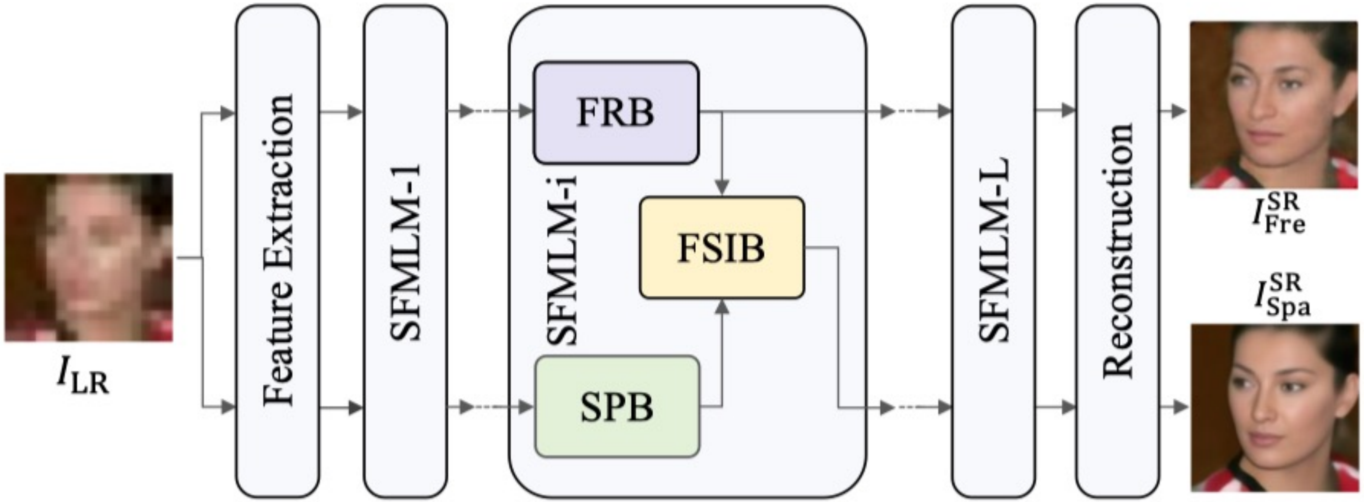
Spatial branch: **local dependency capture** and **global dependency incorporation.**



Overview of the proposed SFMNet.

- We develop a spatial-frequency mutual network (SFMNet) equipped with Fourier transform. To the best of our knowledge, this is the first method that explores the potential of both spatial and frequency information for face super-resolution.

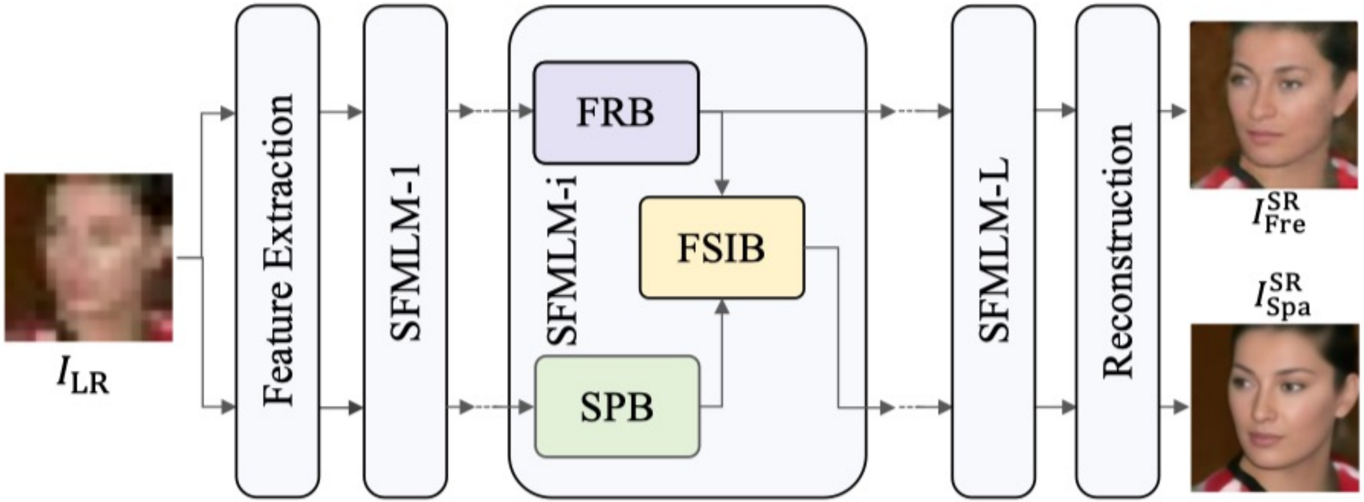
SFMNet



Overview of the proposed SFMNet.

- Component: feature extraction layer, L spatial-frequency mutual learning modules (SFMLM), reconstruction layer.

SFMNet

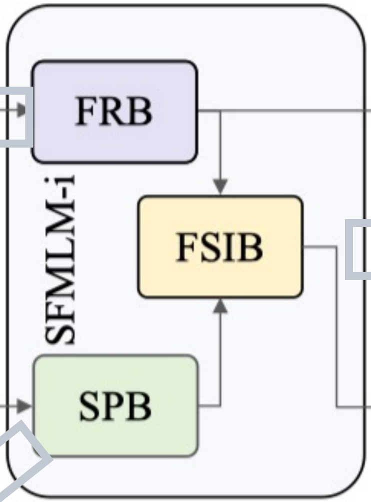


Overview of the proposed SFMNet.

- Objective functions:
  - Pixel-level loss: $\mathcal{L}_{\text{Pix}} = \left\| \boldsymbol{I}^{\text{SR}}_{\text{Spa}} - \boldsymbol{I}_{\text{HR}} \right\|_1 + \left\| \boldsymbol{I}^{\text{SR}}_{\text{Fre}} - \boldsymbol{I}_{\text{HR}} \right\|_1,$
  - Frequency-level loss: $\mathcal{L}_{\text{Fre}} = \left\| \mathcal{A}(\boldsymbol{I}^{\text{SR}}_{\text{Fre}}) - \mathcal{A}(\boldsymbol{I}_{\text{HR}}) \right\|_1 + \left\| \mathcal{P}(\boldsymbol{I}^{\text{SR}}_{\text{Fre}}) - \mathcal{P}(\boldsymbol{I}_{\text{HR}}) \right\|_1,$
  - Adversarial loss: $\mathcal{L}^{\text{Adv}}_{\text{Spa}} = -log(\mathcal{SD}(\boldsymbol{I}^{\text{SR}}_{\text{Spa}})), \quad \mathcal{L}^{\text{Adv}}_{\text{Fre}} = -log(\mathcal{FD}([\mathcal{A}(\boldsymbol{I}^{\text{SR}}_{\text{Spa}}), \mathcal{P}(\boldsymbol{I}^{\text{SR}}_{\text{Spa}})])),$
  - Perceptual loss: $\mathcal{L}_{\text{Per}} = \left\| \Phi(\boldsymbol{I}^{\text{SR}}_{\text{Spa}}) - \Phi(\boldsymbol{I}_{\text{HR}}) \right\|_1,$

SFMLM

Frequency block (FRB): captures global dependency and frequency information by Fourier Transform.

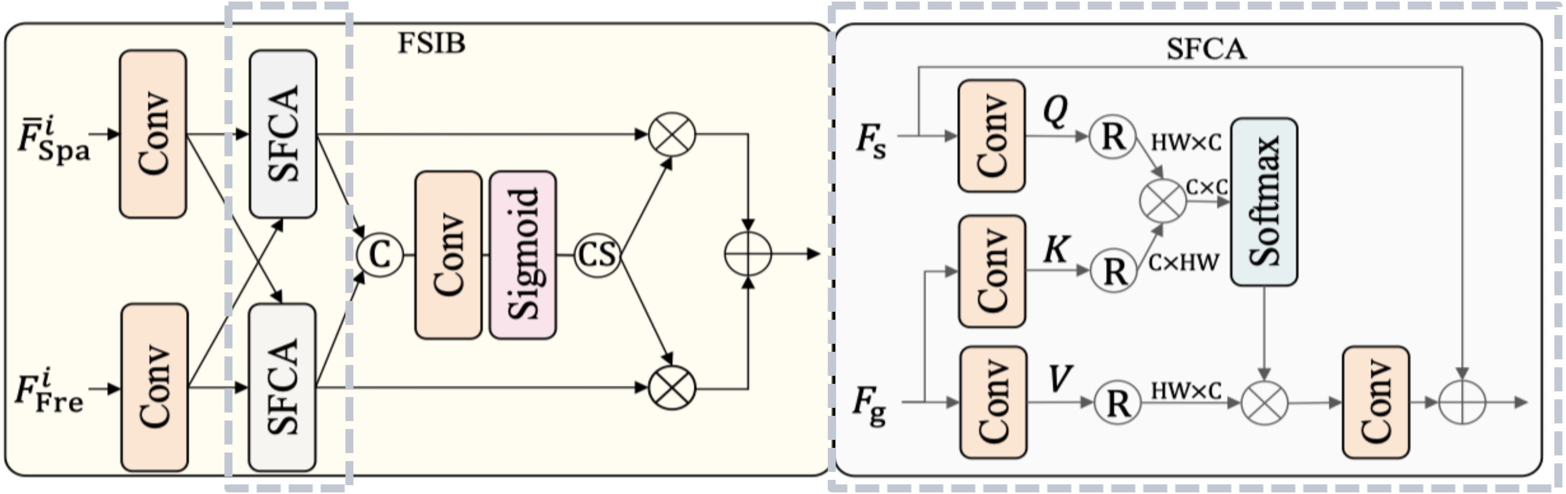


frequency-spatial interaction block (FSIB): mutually and adaptively fuse the global frequency information and local spatial information..

Spatial block(SPB): extracts local dependency and spatial information.

Global and local dependencies are **complementary** and can **both facilitate FSR.**

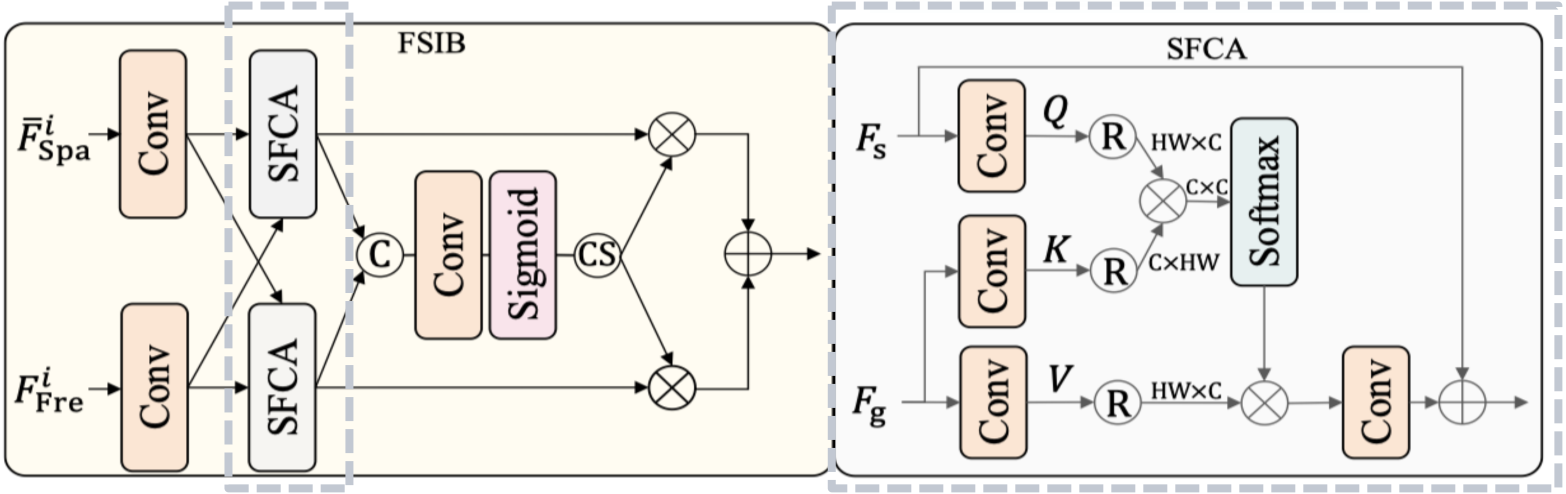
FSIB



Frequency-spatial interaction block (FSIB) (left) and spatial-frequency cross-attention (SFCA) (right).

- FSIB first applies two convolutional layers on spatial and frequency features.

FSIB



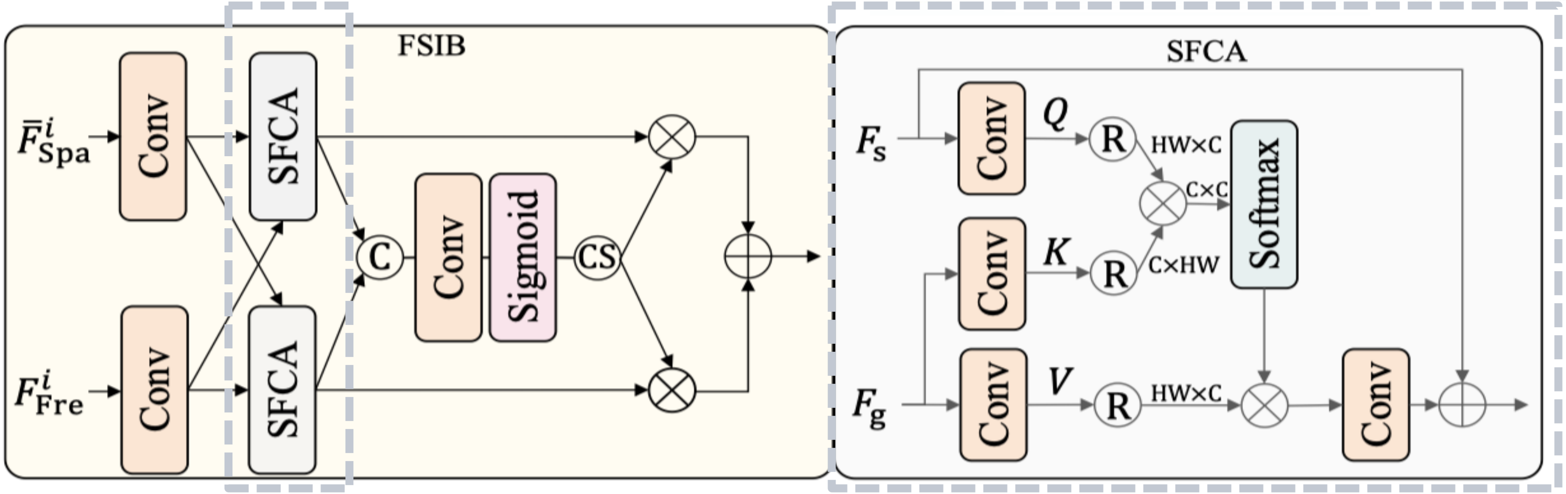
Frequency-spatial interaction block (FSIB) (left) and spatial-frequency cross-attention (SFCA) (right).

- Coarse fusion: spatial-frequency cross-attention (SFCA)
- SFCA has two inputs: source information F_s and guidance information F_g
- SFCA uses F_s to generate query Q and use F_g to generate key K and value V

$$\mathrm{Attention}(K, Q, V) = f_{\mathrm{Softmax}}(QK^T/\sqrt{d})V,$$

- Frequency feature and the spatial feature serve as source and guidance for each other.

FSIB



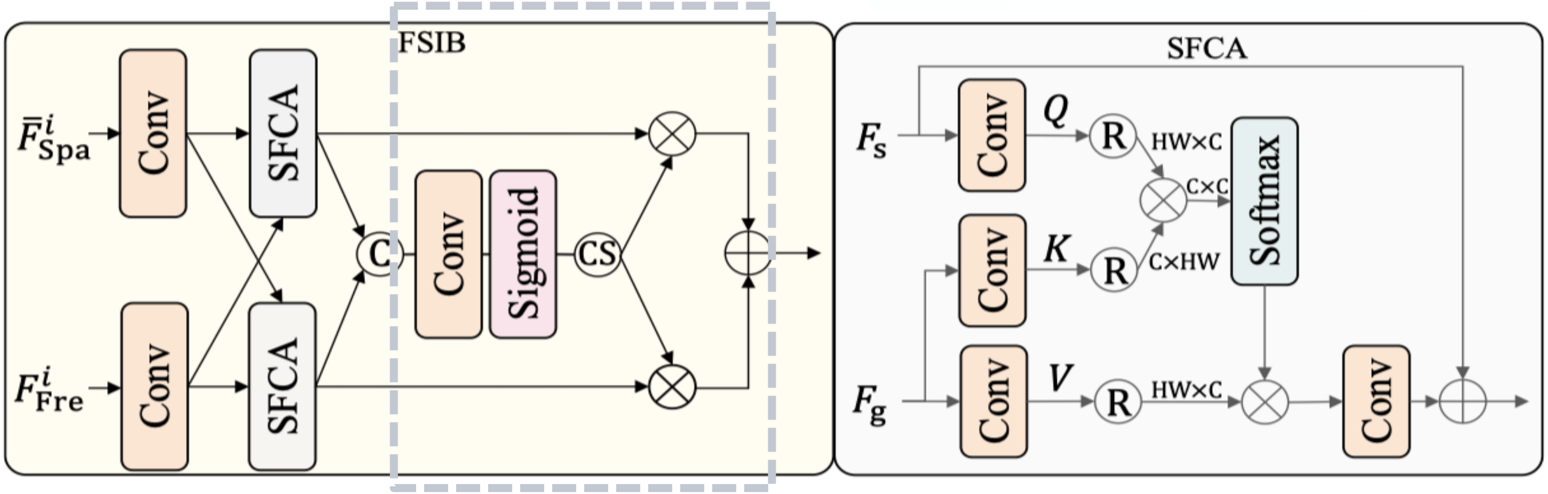
Frequency-spatial interaction block (FSIB) (left) and spatial-frequency cross-attention (SFCA) (right).

- Effectiveness of SFCA: replace SFCA with Concatenation-Convolution (CC)

	CelebA		Helen	
	PSNR	SSIM	PSNR	SSIM
CC	27.40	0.8022	27.10	0.8072
SFCA	**27.56**	**0.8082**	**27.22**	**0.8141**

SFCA can improve face super-resolution performance.

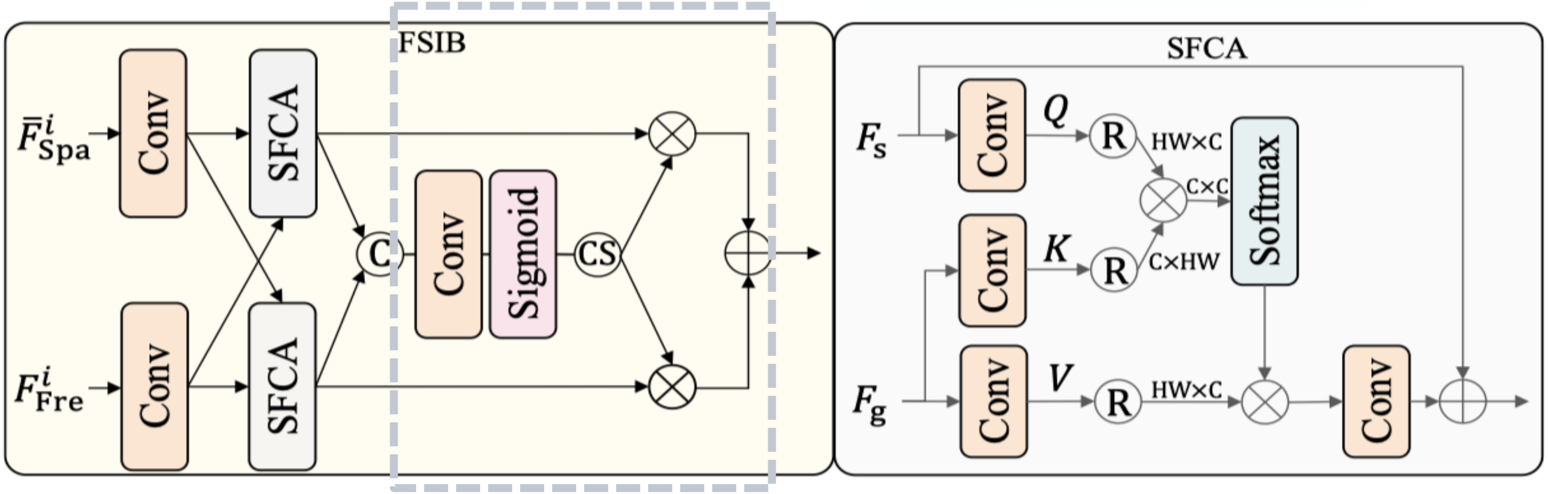
FSIB



Frequency-spatial interaction block (FSIB) (left) and spatial-frequency cross-attention (SFCA) (right).

- Fine fusion: use coarsely fused feature to generate attention map for refinement.

FSIB



Frequency-spatial interaction block (FSIB) (left) and spatial-frequency cross-attention (SFCA) (right).

- Effectiveness of FSIB: replace FSIB with concatenation-convolution (CC)

	CelebA		Helen	
	PSNR	SSIM	PSNR	SSIM
CC	27.39	0.8033	27.01	0.8079
FSIB	**27.56**	**0.8082**	**27.22**	**0.8141**

FSIB can improve face super-resolution performance.

Dataset	CelebA [30]						Helen [25]						Par	Time
	×4			×8			×4			×8				
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓		
Bicubic	27.48	0.8166	0.1841	23.58	0.6285	0.2692	28.22	0.6628	0.1771	23.88	0.6628	0.2560	-	-
SRCNN [10]	28.04	0.8369	0.1599	23.93	0.6348	0.2559	28.77	0.8730	0.0556	24.27	0.6770	0.2430	19.6k	9.1ms
EDSR [29]	31.45	0.9095	0.0518	26.84	0.7787	0.1159	31.87	0.9286	0.0574	26.60	0.7851	0.1400	3.4M	10.0ms
FSRNet [9]	31.46	0.9084	0.0519	26.66	0.7714	0.1098	31.93	0.9283	0.0543	26.43	0.7799	0.1356	3.2M	53.0ms
DIC [32]	31.53	0.9107	0.0532	27.37	0.8022	0.0920	31.98	0.9303	0.0576	26.94	0.8026	0.1144	20.8M	84.6ms
SPARNet [8]	31.71	0.9129	0.0476	27.42	0.8036	0.0891	31.98	0.9300	0.0592	26.95	0.8029	0.1169	10.0M	45.0ms
SISN [31]	31.88	0.9157	0.0476	27.31	0.7978	0.0998	32.41	0.9351	0.0535	27.08	0.8083	0.1225	8.4M	63.8ms
SFMNet(Ours)	**32.01**	**0.9175**	0.0441	**27.56**	**0.8074**	0.0869	**32.51**	**0.9362**	0.0498	**27.22**	**0.8141**	0.1061	8.1M	51.8ms
SFMNet+GAN	30.99	0.8051	**0.0291**	26.48	0.7662	**0.0594**	31.54	0.9187	**0.0323**	26.39	0.7792	**0.0760**	8.1M	51.8ms

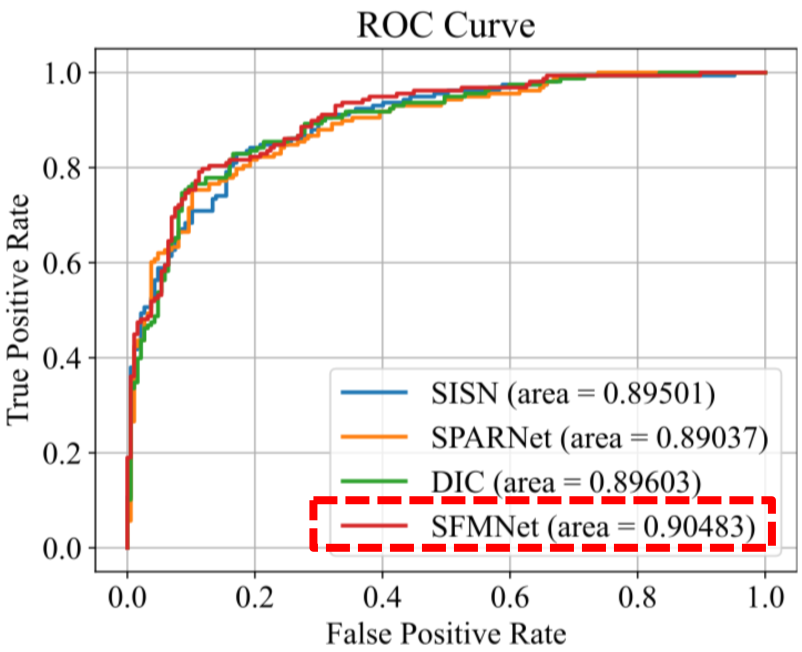
SFMNet achieves a good balance between performance and model complexity.

Experiments

LR SRCNN EDSR FSRNet DIC SPARNet SISN SFMNet SFMGAN GT

SFMNet can recover more accurate and realistic details than other methods.

Experiments



SFMNet outperforms other FSR methods in face recognition task.

Although the results of SFMNet are not as high quality as those of VQFR, they are realistic and natural, and contain key facial details.

Conclusion

- We develop a spatial-frequency mutual network (SFMNet) for face super-resolution, which is the first work to explore the interaction between spatial domain and frequency domain in this field.

- We carefully design a frequency-spatial interaction block that can fuse these dependencies mutually and boost face super-resolution performance.

- Experimental results demonstrate that our proposed method can achieve state-of-the-art performance.

Thank you for your attention！

https://aiialabhit.github.io/

Artificial Intelligence & Image Analysis (AIIA) Lab