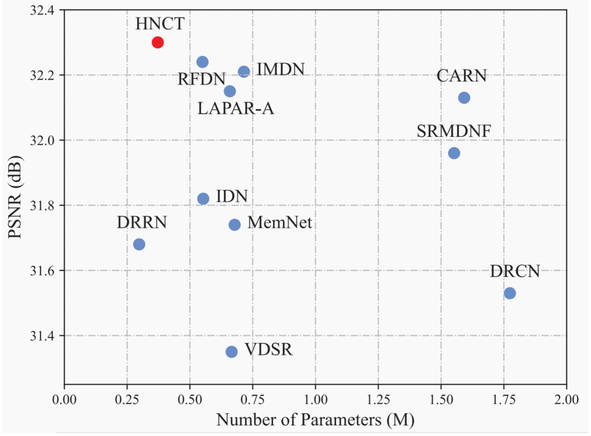
A Hybrid Network of CNN and Transformer for Lightweight Image Super-Resolution

Jinsheng Fang[1], Hanjiang Lin[1], Xinyu Chen[1], Kun Zeng[2]*
[1]Minnan Normal University, China [2]Minjiang University, China

Introduction

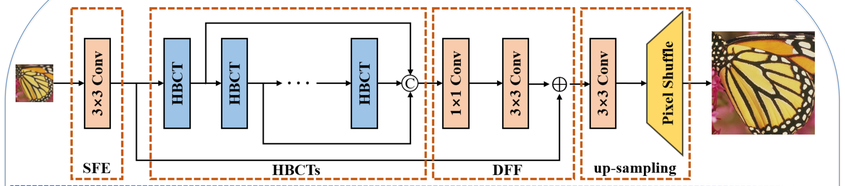
- A lightweight hybrid network of CNN and Transformer (HNCT) is proposed for image super-resolution, which achieves better SR performance with fewer parameters than other methods.
- A hybrid block of CNN and Transformer (HBCT) exploits local and non-local priors simultaneously to extract features beneficial for SR.

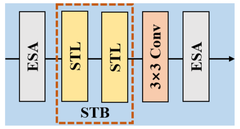


Methods

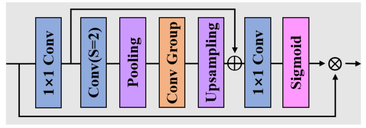
As shown in the following figure, the proposed HNCT consists of four parts:

- Shallow feature extraction (SFE)
- Hybrid blocks of CNN and Transformer (HBCTs)
  - A Swin Transformer block (STB) with two Swin Transformer layers inside
  - A convolutional layer
  - Two enhanced spatial attention (ESA) modules
- Dense feature fusion (DFF)
- Up-sampling module

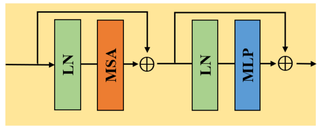




(a) Hybrid Block of CNN and Transformer(HBCT)



(c) Enhanced Spatial Attention (ESA)

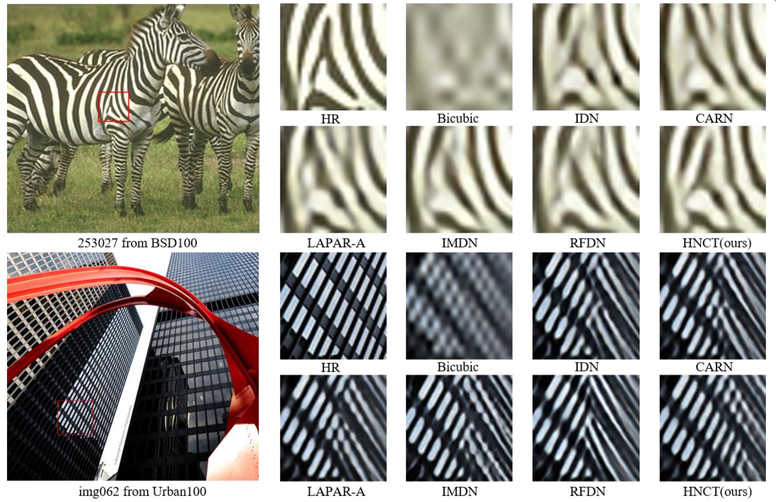


(b) Swin Transformer Layer (STL)

Ⓒ Concatenation ⊕ Element-wise addition
⊗ Element-wise multiplication
SFE: Shallow Feature Extraction
DFF: Dense Feature Fusion
HBCTs: Hybrid Blocks of CNN and Transformer
Conv(S=2): Convolutional layer with stride of 2
Conv Group: Three convolutional layers

Results

Method	Scale	Params	Set5	Set14	BSD100	Urban100	Manga109
			PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic	x3	-	30.39/0.8682	27.55/0.7742	27.21/0.7385	24.46/0.7349	26.95/0.8556
VDSR		666K	33.66/0.9213	29.77/0.8314	28.82/0.7976	27.14/0.8279	32.01/0.9340
DRCN		1774K	33.82/0.9226	29.76/0.8311	28.80/0.7963	27.15/0.8276	32.24/0.9343
DRRN		298K	34.03/0.9244	29.96/0.8349	28.95/0.8004	27.53/0.8378	32.71/0.9379
MemNet		678K	34.09/0.9248	30.00/0.8350	28.96/0.8001	27.56/0.8376	32.51/0.9369
IDN		553K	34.11/0.9253	29.99/0.8354	28.95/0.8013	27.42/0.8359	32.71/0.9381
SRMDNF		1528K	34.12/0.9254	30.04/0.8382	28.97/0.8025	27.57/0.8398	33.00/0.9403
CARN		1592K	34.29/0.9255	30.29/0.8407	29.06/0.8034	28.06/0.8493	33.50/0.9440
LAPAR-A		544k	34.36/0.9267	30.34/0.8421	29.11/0.8054	28.15/0.8523	33.51/.09441
IMDN		703K	34.36/0.9270	30.32/0.8417	29.09/0.8046	28.17/0.8519	33.61/0.9445
RFDN		541K	34.41/0.9273	30.34/0.8420	29.09/0.8050	28.21/0.8525	33.67/0.9449
HNCT(Ours)		363K	34.47/0.9275	30.44/0.8439	29.15/0.8067	28.28/0.8557	33.81/0.9459
Bicubic	x4	-	28.42/0.8104	26.00/0.7027	25.96/0.6675	23.14/0.6577	24.89/0.7866
VDSR		666K	31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524	28.83/0.8870
DRCN		1774K	31.53/0.8854	28.02/0.7670	27.23/0.7233	25.14/0.7510	28.93/0.8854
DRRN		298K	31.68/0.8888	28.21/0.7720	27.38/0.7284	25.44/0.7638	29.45/0.8946
MemNet		678K	31.74/0.8893	28.26/0.7723	27.40/0.7281	25.50/0.7630	29.42/0.8942
IDN		553K	31.82/0.8903	28.25/0.7730	27.41/0.7297	25.41/0.7632	29.41/0.8942
SRMDNF		1552K	31.96/0.8925	28.35/0.7787	27.49/0.7337	25.68/0.7731	30.09/0.9024
CARN		1592K	32.13/0.8937	28.60/0.7806	27.58/0.7349	26.07/0.7837	30.47/0.9084
LAPAR-A		659K	32.15/0.8944	28.61/0.7818	27.61/0.7366	26.14/0.7871	30.42/0.9074
IMDN		715K	32.21/0.8948	28.58/0.7811	27.56/0.7353	26.04/0.7838	30.45/0.9075
RFDN		550K	32.24/0.8952	28.61/0.7819	27.57/0.7360	26.11/0.7858	30.58/0.9089
HNCT(Ours)		372K	32.31/0.8957	28.71/0.7834	27.63/0.7381	26.20/0.7896	30.70/0.9112



- The table shows quantitative results of five benchmark datasets.
- The figure shows three visual comparisons between HNCT and the other lightweight competitors on ×4.

Conclusion

- We propose a hybrid network of CNN and Transformer (HNCT) for lightweight image SR.
- HNCT can exploit both local and non-local priors and extract deep features more beneficial for image SR.
- Enhanced spatial attention (ESA) is employed to further improve SR results.

References

[1] Jingyun Liang, et al. SwinIR: Image restoration using swin transformer. CVPR 2021.
[2] Jie Liu, et al. Residual feature distillation network for lightweight image super-resolution. ECCV 2020.
[3] Wenbo Li, et al. Lapar: Linearly-assembled pixel-adaptive regression network for single image super-resolution and beyond. NeurIPS 2020.