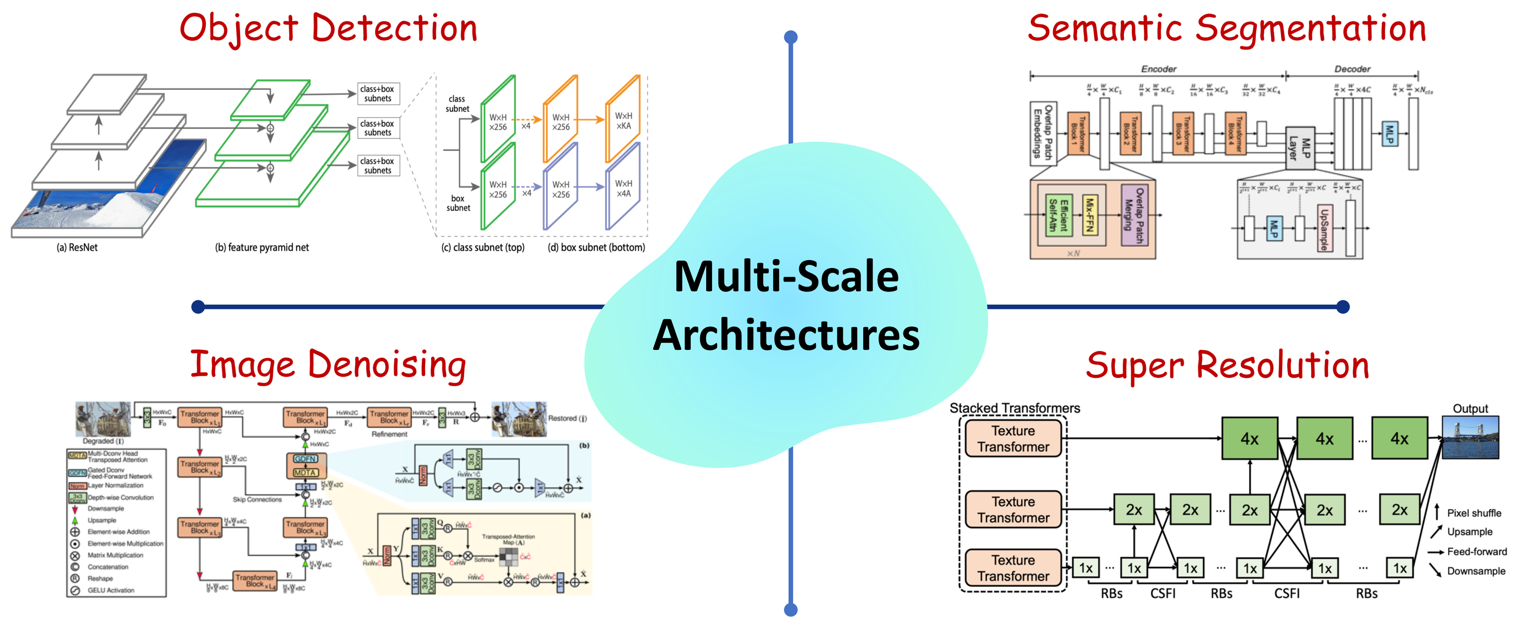
Multi-Scale Adaptive Network for Single Image Denoising

Yuanbiao Gou[1], Peng Hu[1], Jiancheng Lv[1], Joey Tianyi Zhou[2], Xi Peng[1]*
[1]Sichuan University, [2]A*STAR

Preliminary

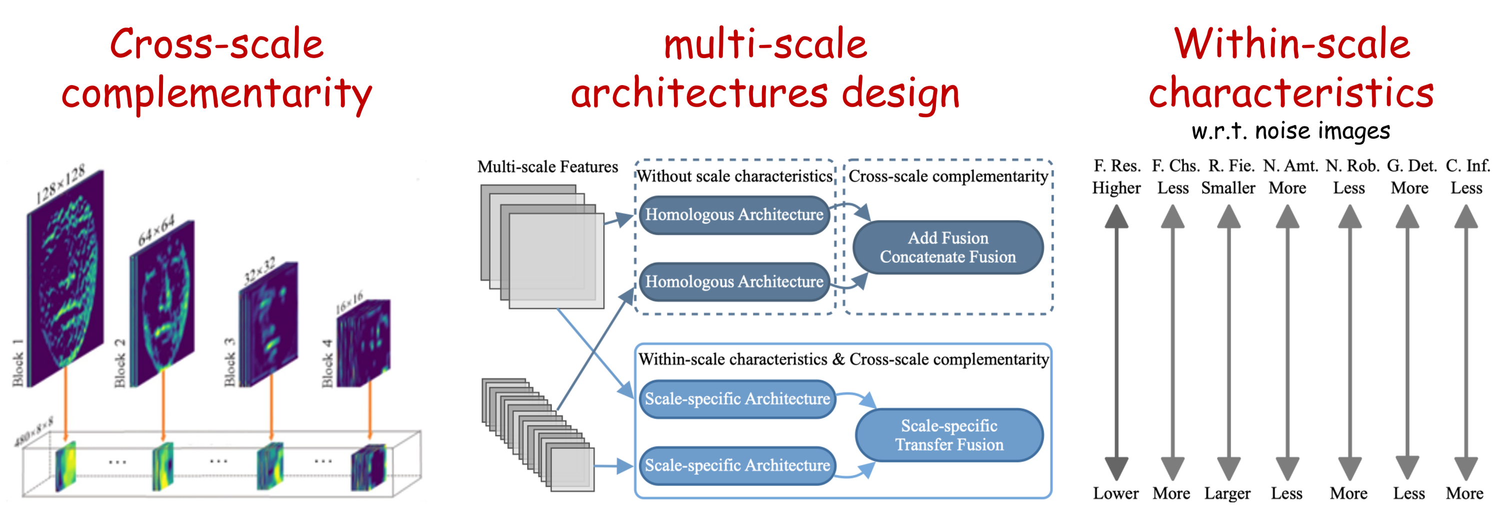
Multi-scale architectures have shown effectiveness in vision tasks, thanks to multi-scale features and cross-scale complementarity.



Feeding multi-/single-resolution images/features into single/multiple subnetworks, and fuse the outputs for exploiting cross-scale complementarity.

Observation & Motivation

A missing piece in modern multi-scale architecture design, i.e., the within-scale characteristics of multi-scale features are ignored.



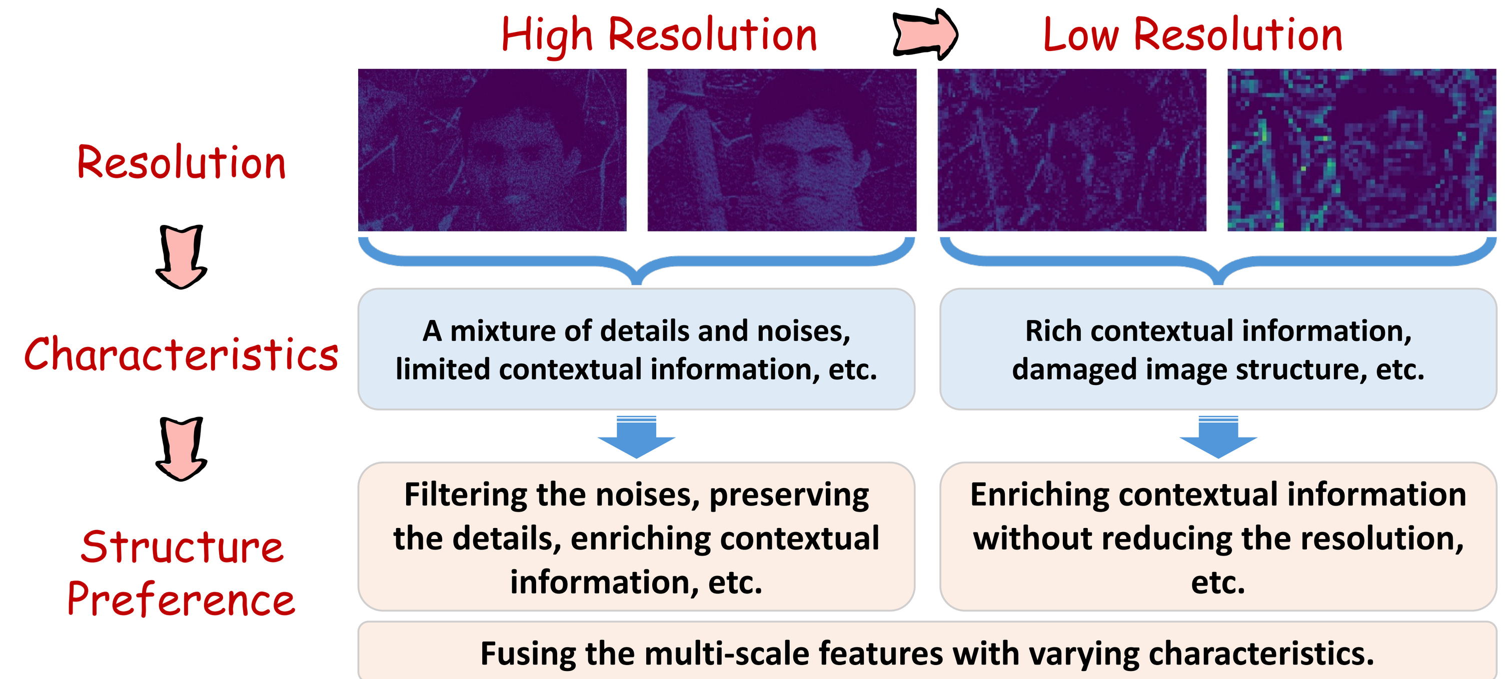
- Existing architectures process the different scale features by homologous architectures, such as skip connections and universal neural blocks, without considering the scale-specific characteristics.
- Within-scale characteristics involves feature resolutions (F. Res.), feature channels (F. Chs.), receptive field (R. Fie.), noise amount (N. Amt.), noise robustness (N. Rob.), geometric details (G. Det.), contextual information (C. Inf.), etc.
- Our Proposal: the different scale features show varying characteristics and should be processed by scale-specific structures rather than homologous architectures. In other words, the different network structures corresponding to the different scale features for adapting their varying characteristics.

Key Contribution

We reveal the missing piece in multi-scale architecture design, and accordingly design a novel multi-scale adaptive network as well as two neural block for single image denoising, whose structures simultaneously consider within-scale characteristics and cross-scale complementarity.

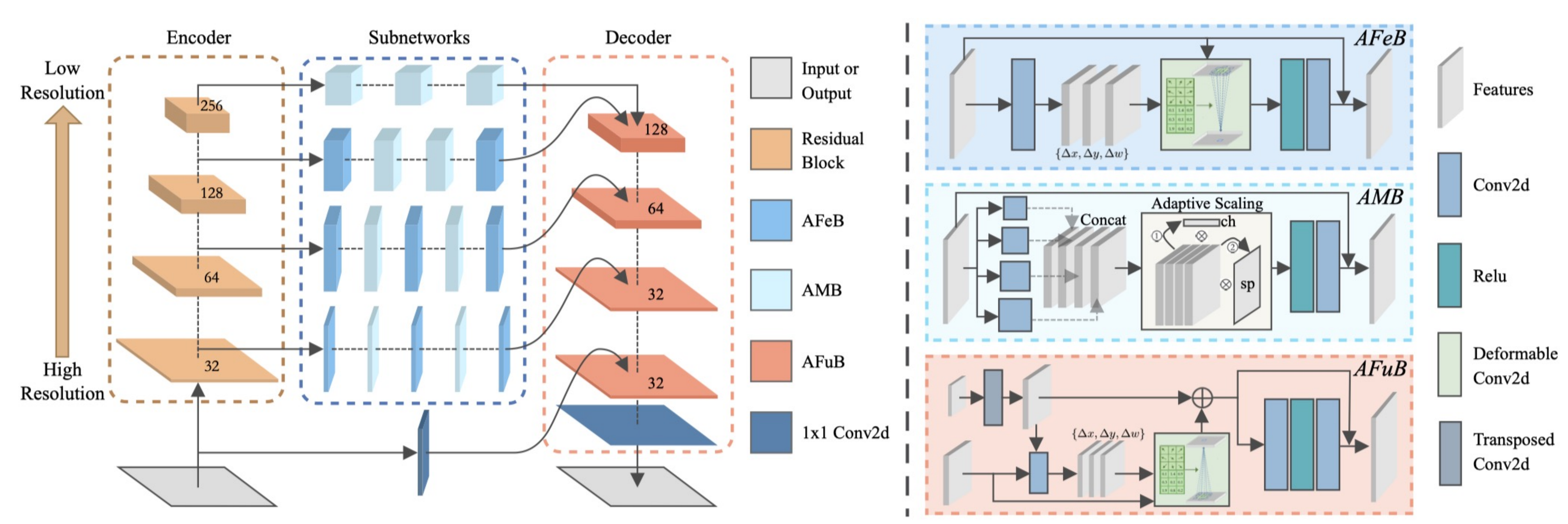
Method

Architecture Principles



The different scale features show varying characteristics, and thus enjoying the different network structure preferences.

Architecture Overview



Multi-scale Adaptive Network

- Encoder
  - Residual block only
  - Extracting features of different scales
- Decoder
  - AFuB only
  - Transferring details into contexts
- High-resolution branches
  - Alternately stack AFeB and AMB
  - Deeper structure for effectiveness
- Low-resolution branches
  - AMB only
  - Shallower depth for efficiency

Adaptive Neural Blocks

- AFeB preserves the indispensable details while filtering unpleasant noises via

$$\{\Delta x, \Delta y, \Delta w\}_{(x,y)} = F(f_i) \qquad f_{i+1}(x,y) = \sum_{j=1}^{k} w_j * f_i(x+\Delta x_j, y+\Delta y_j) * \Delta w_j$$

$$f_{out} = f_i + F_{conv}(F_{relu}(f_{i+1}))$$

- AMB enriches context information without losing details and damaging structure via

$$f_{i+1} = Concat(\{F_k^d(f_i) | d, k \in \mathbb{N}^+\})$$

$$f_{out} = f_i + F_{conv}(F_{relu}(f_{i+3}))$$

$$ch = 2 * F_{sig}(F_{fc}(avg\_pool(f_{i+1}))),$$
$$f_{i+2} = ch * f_{i+1},$$
$$sp = 2 * F_{sig}(F_{conv}(mean(f_{i+2}))),$$
$$f_{i+3} = sp * f_{i+2},$$

- AFuB fuses multi-scale features with varying characteristic via

$$f_{coarse} = F_{TConv}(f_{coarse}^{low}) \qquad \{\Delta x, \Delta y, \Delta w\}_{(x,y)} = F(f_{coarse}, f_{fine})$$

$$f_{coarse}^{fine} = f_{coarse} + \sum_{j=1}^{k} w_j * f_{fine}(x+\Delta x_j, y+\Delta y_j) * \Delta w_j \qquad f_{out} = f_{coarse}^{fine} + F_{conv}(F_{relu}(F_{conv}(f_{coarse}^{fine})))$$

Experiments

Table 1: Quantitative results on SIDD sRGB validation dataset.

Method	CDnCNN-B	CBM3D	CBDNet	PD	RIDNet	SADNet	DeamNet	MSANet
PSNR	26.21	30.88	33.07	33.96	38.71	39.46	39.47	**39.56**
SSIM	-	-	0.8324	0.8195	0.9052	0.9103	0.9105	**0.9118**

Table 2: Quantitative results on Nam dataset with JPEG compression.

Method	CDnCNN-B	CBM3D	CBDNet	PD	RIDNet	SADNet	DeamNet	MSANet
PSNR	37.49	39.84	41.31	41.09	41.04	42.92	42.03	**43.52**
SSIM	0.9272	0.9657	0.9784	0.9780	0.9814	0.9839	0.9790	**0.9863**

Table 3: Quantitative results on DnD sRGB dataset.

Method	CDnCNN-B	CBM3D	FFDNet+	CBDNet	N3Net	PR	RIDNet	SADNet	DeamNet	MSANet
PSNR	32.43	34.51	37.61	38.06	38.32	39.00	39.26	39.59	39.63	**39.65**
SSIM	0.7900	0.8507	0.9415	0.9421	0.9384	0.9542	0.9528	0.9523	0.9531	**0.9553**

Figure 5: Qualitative results on real noise image from DnD dataset. From left to right, we show the real noise image, the results of CBDNet, RIDNet, PD, SADNet, DeamNet, and MSANet.

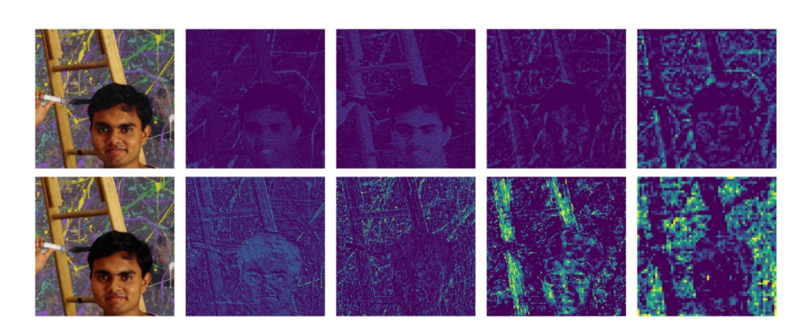
Table 4: Quantitative results on synthetic color noise image datasets.

Dataset	σ	CBM3D	DnCNN	FFDNet	CLEARER	SADNet	RNAN	DeamNet	MSANet (Ours)
CMcMaster	30	29.58/0.8107	29.64/0.8098	30.05/0.8221	30.83/0.8522	31.96/0.8857	32.01/0.8848	32.00/0.8862	**32.07/0.8876**
	50	25.92/0.7153	25.99/0.7147	26.23/0.7244	28.92/0.8143	29.72/0.8374	29.69/0.8333	29.78/0.8393	**29.82/0.8403**
	70	23.12/0.6398	23.03/0.6297	23.19/0.6406	26.96/0.7504	28.25/0.7988	28.14/0.7918	28.27/0.8000	**28.35/0.8028**
CKodak24	30	30.33/0.8417	30.77/0.8548	30.62/0.8542	31.17/0.8590	31.72/0.8730	31.72/0.8716	31.76/0.8736	**31.78/0.8744**
	50	27.28/0.7572	27.63/0.7718	27.54/0.7687	28.94/0.7977	29.49/0.8149	29.43/0.8102	29.53/0.8155	**29.57/0.8169**
	70	24.84/0.6890	24.90/0.6912	24.88/0.6890	27.59/0.7503	28.10/0.7715	27.99/0.7635	28.14/0.7721	**28.17/0.7731**
CBSD68	30	29.22/0.8378	29.72/0.8556	29.51/0.8526	30.35/0.8665	30.63/0.8749	30.61/0.8733	30.65/0.8749	**30.67/0.8758**
	50	26.06/0.7378	26.48/0.7600	26.38/0.7550	28.01/0.7996	28.31/0.8089	28.25/0.8050	28.34/0.8093	**28.36/0.8107**
	70	23.70/0.6548	23.86/0.6626	23.80/0.6584	26.58/0.7433	26.91/0.7577	26.81/0.7511	26.92/0.7574	**26.96/0.7591**

Table 5: Quantitative results on synthetic grayscale noise image datasets.

Dataset	σ	BM3D	DnCNN	FFDNet	CLEARER	SADNet	RNAN	DeamNet	MSANet (Ours)
GMcMaster	30	29.45/0.8151	29.80/0.8119	29.89/0.8292	30.29/0.8491	30.92/0.8649	30.92/0.8629	30.94/0.8656	**30.96/0.8661**
	50	26.23/0.7218	26.24/0.7281	26.46/0.7319	28.31/0.7945	28.61/0.8052	28.57/0.8014	28.65/0.8070	**28.68/0.8072**
	70	23.78/0.6517	23.63/0.6682	23.64/0.6466	26.83/0.7427	27.18/0.7606	27.06/0.7526	27.20/0.7616	**27.22/0.7620**
GKodak24	30	28.71/0.7854	29.21/0.7946	29.15/0.8077	29.49/0.8132	29.87/0.8238	29.89/0.8208	29.90/0.8241	**29.91/0.8248**
	50	26.22/0.6996	26.52/0.7190	26.52/0.7177	27.27/0.7412	27.77/0.7559	27.73/0.7494	27.79/**0.7567**	**27.81**/0.7564
	70	24.37/0.6393	24.31/0.6647	24.28/0.6419	26.12/0.6931	26.51/0.7090	26.42/0.6989	26.53/**0.7107**	**26.54**/0.7091
GBSD68	30	27.43/0.7721	27.96/0.7762	27.89/0.7982	28.27/0.8112	28.58/0.8165	28.59/0.8140	28.59/0.8165	**28.61/0.8174**
	50	24.90/0.6715	25.19/0.6826	25.19/0.6909	26.09/0.7295	26.50/0.7382	26.46/0.7333	26.50/0.7392	**26.51/0.7393**
	70	23.07/0.5985	23.04/0.6107	22.98/0.5942	25.03/0.6734	25.23/0.6828	25.15/0.6736	25.23/**0.6831**	**25.25**/0.6826

Ablations	ED	ResB	AFeB	AMB	AFuB	AFeB+AMB	AFeB+AFuB	AMB+AFuB	MSANet
PSNR	31.70	31.93	31.94	31.94	32.01	31.98	32.04	32.03	32.07
SSIM	0.8801	0.8851	0.8854	0.8851	0.8864	0.8860	0.8869	0.8866	0.8876



Multi-scale features with our networks (the bottom row) show more significant within-scale characteristics and cross-scale complementarity.

Feel free to send me an email!

gouyuanbiao@gmail.com

The code could be accessed from https://pengxi.me

[1] Lin T Y, et al. Focal Loss for Dense Object Detection. CVPR, 2017.
[2] Xie E, et al. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. NeurIPS, 2021.
[3] Zamir S W, et al. Restormer: Efficient Transformer for High-resolution Image Restoration. CVPR, 2022.
[4]Yang F, et al. Learning Texture Transformer Network for Image Super-Resolution. CVPR, 2020.