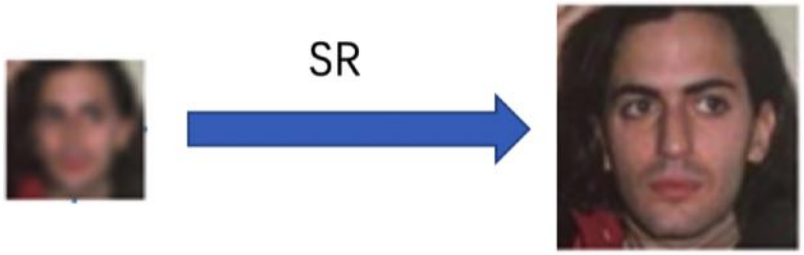
PARSING MAP GUIDED MULTI-SCALE ATTENTION NETWORK FOR FACE HALLUCINATION

Chenyang Wang, Zhiwei Zhong, Junjun Jiang, Deming Zhai, and Xianming Liu
School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China

Introduction–Super-resolution

Image super-resolution (SR):

transforms low-resolution (LR) images to high-resolution (HR) images



Introduction–Face hallucination

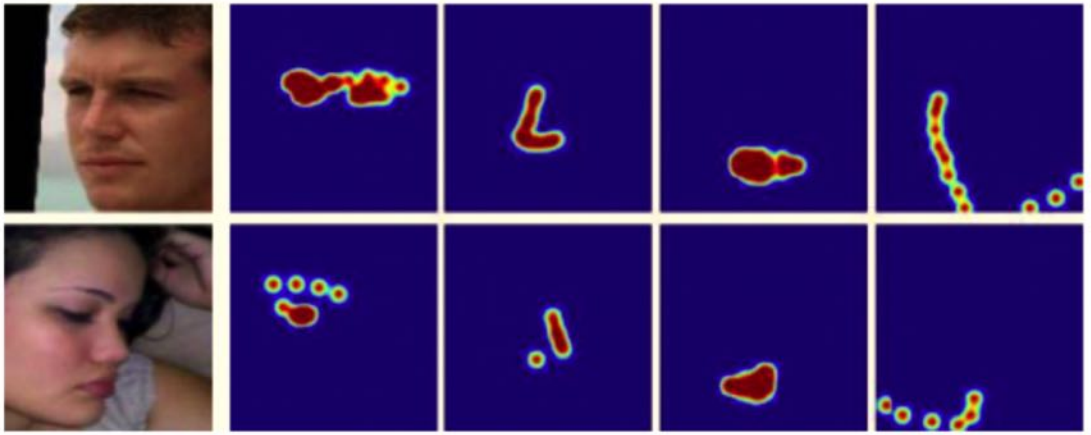
Face SR VS General image SR:

Face images have their special structural information and prior knowledge such as :

1. parsing map
2. landmark
3. heatmap

a) landmark

b) parsing map



c) heatmaps

Introduction

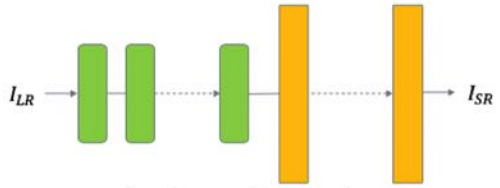
Our contributions:

1. We design a FishSRNet to generate features in a variety of resolution
2. We propose a multi-scale channel and spatial attention block (MSAB)
3. We get prior knowledge directly from input LR faces.

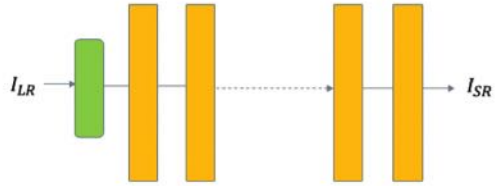
FishSRNet

Existing methods: pre or post-upsampling model, but features in low-resolution or high-resolution don't work well

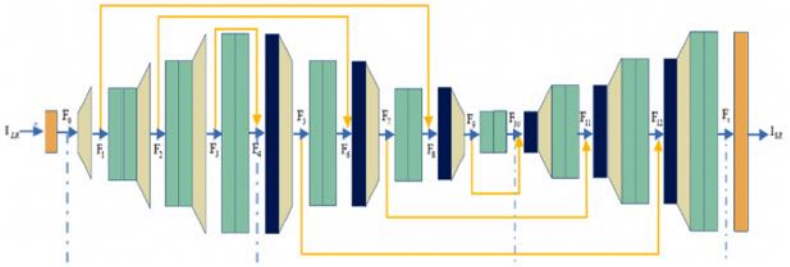
Our method: designs a FishSRNet to generate features in a variety of resolution



a) post-upsampling network



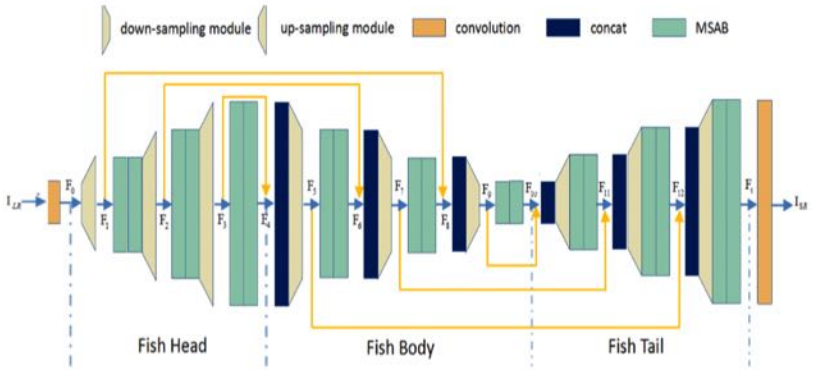
b) pre-upsampling network



c) our FishSRNet

FishSRNet

Our method: designs a FishSRNet to generate features in a variety of resolution

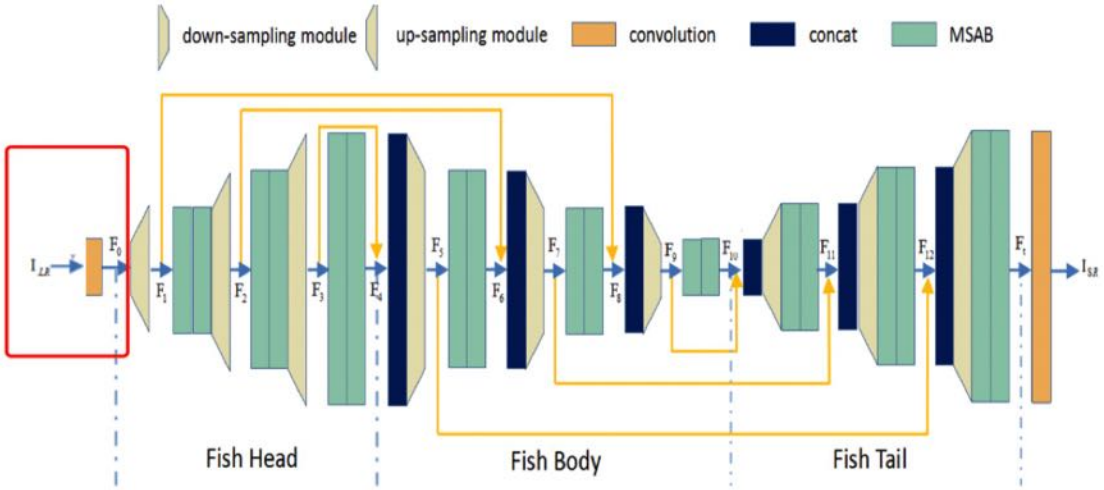


FishSRNet

FishSRNet first up-samples the input then down-samples and up-samples again.

- up-sampling module (UM)
- down-sampling module (DM)

FishSRNet



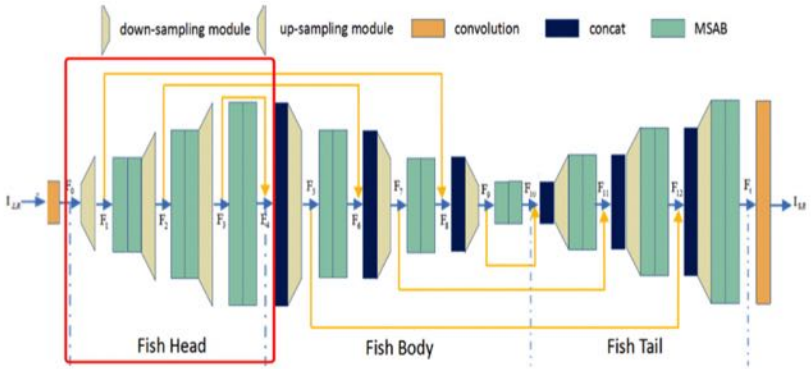
FishSRNet

Feature extraction layer: extracts features from the input

$$F_0 = Feature\ extraction(I_{LR}), \tag{1}$$

where F_0 is the output of the feature extraction layer.

FishSRNet



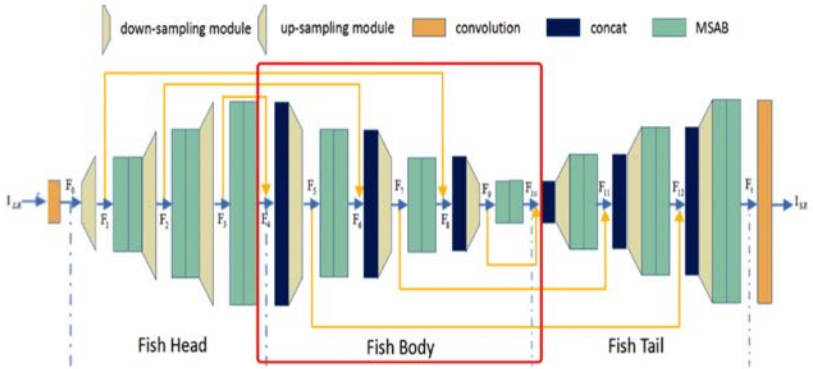
FishSRNet

Fish Head: up-samples features to increase the receptive field and the resolution of the features

$$F_1, F_2, F_3, F_4 = Fish\ Head(F_0), \tag{2}$$

where F_1, F_2, F_3 are the features from every UM for much richer variety of the features, F_4 is for deep layer.

FishSRNet



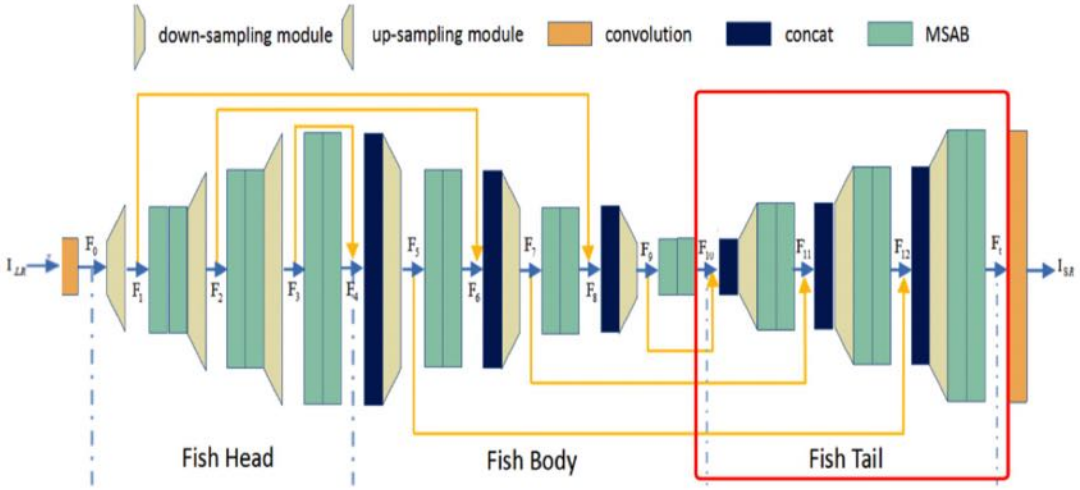
FishSRNet

Fish Body: down-samples features to improve the diversity of resolution

$$F_5, F_7, F_9, F_{10} = Fish\ Body(F_4, F_1, F_2, F_3), \tag{3}$$

where F_5, F_7, F_9 are the features from every DM for much richer variety of the features, F_{10} is for deep layer.

FishSRNet



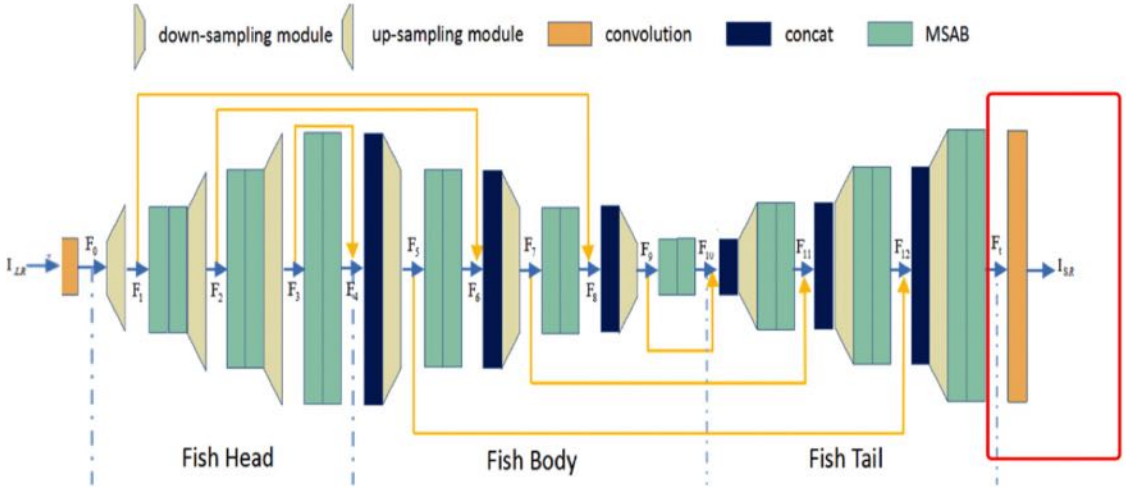
FishSRNet

Fish Tail: up-samples the feature maps to the same resolution as HR

$$F_t = Fish\ Tail(F_{10}, F_9, F_7, F_5), \tag{4}$$

where F_t is for deep layer.

FishSRNet



FishSRNet

Reconstruction layer: generates the final output

$$I_{SR} = Reconstruction(F_t), \tag{5}$$

where I_{SR} is the result of our network.

FishSRNet–Experiment

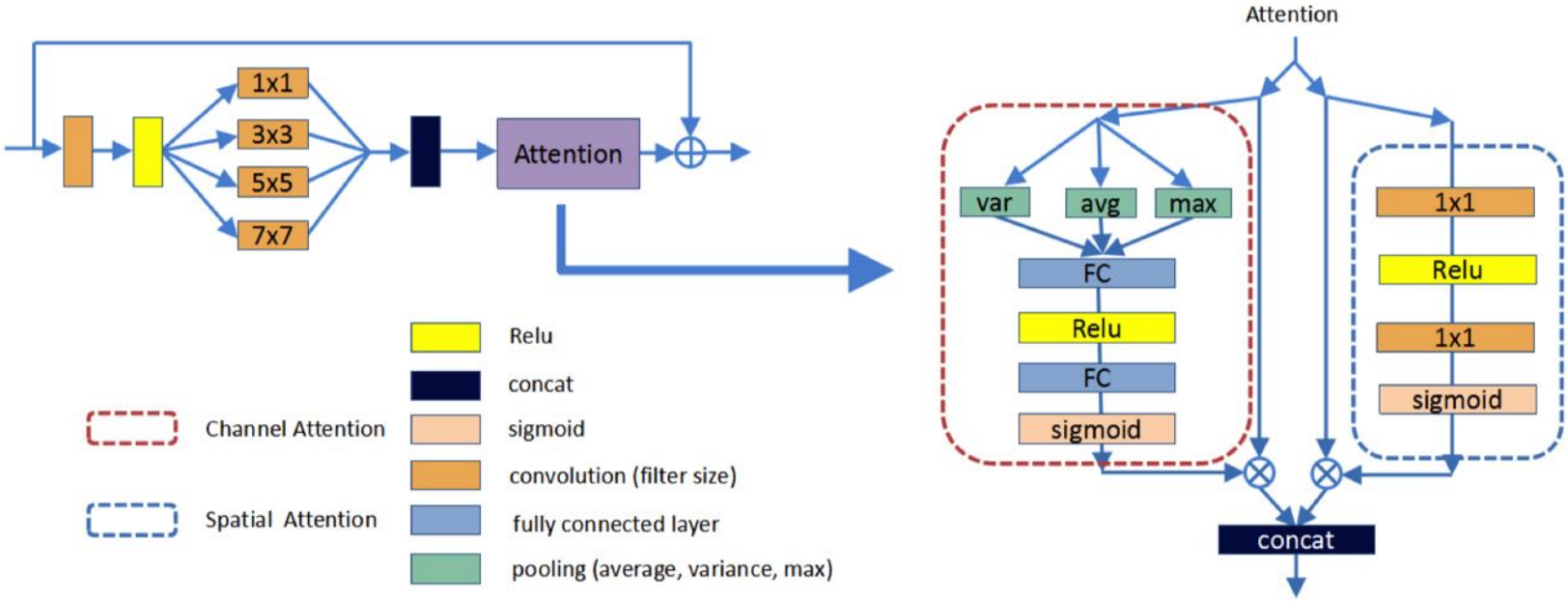
The effectiveness of FishSRNet

Model	PSNR	SSIM
post-upsampling	25.12	0.8705
FishSRNet	25.26	0.8745

MSAB

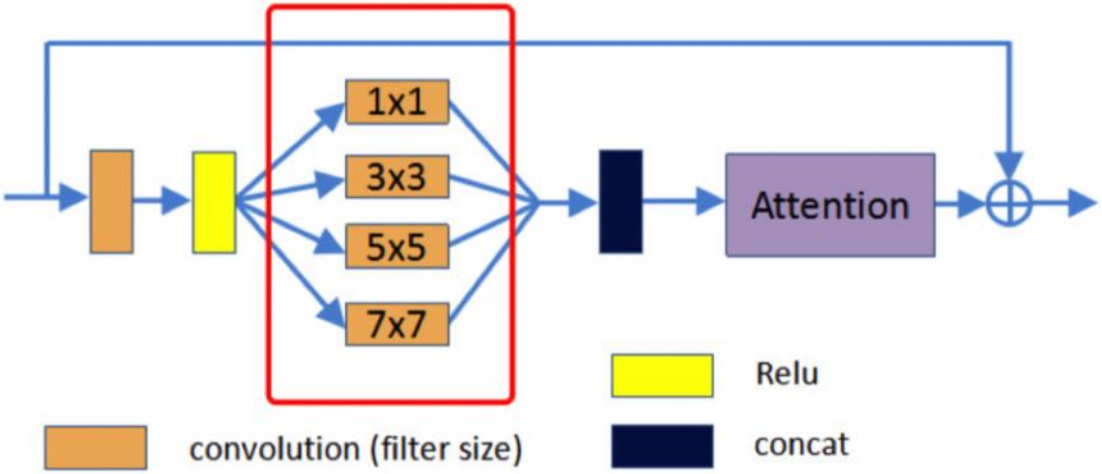
Existing methods: ignore the attention mechanism which is proved useful in general image SR.

Our method: introduces attention mechanism to face SR and constructs a multi-scale channel and spatial attention block (MSAB).



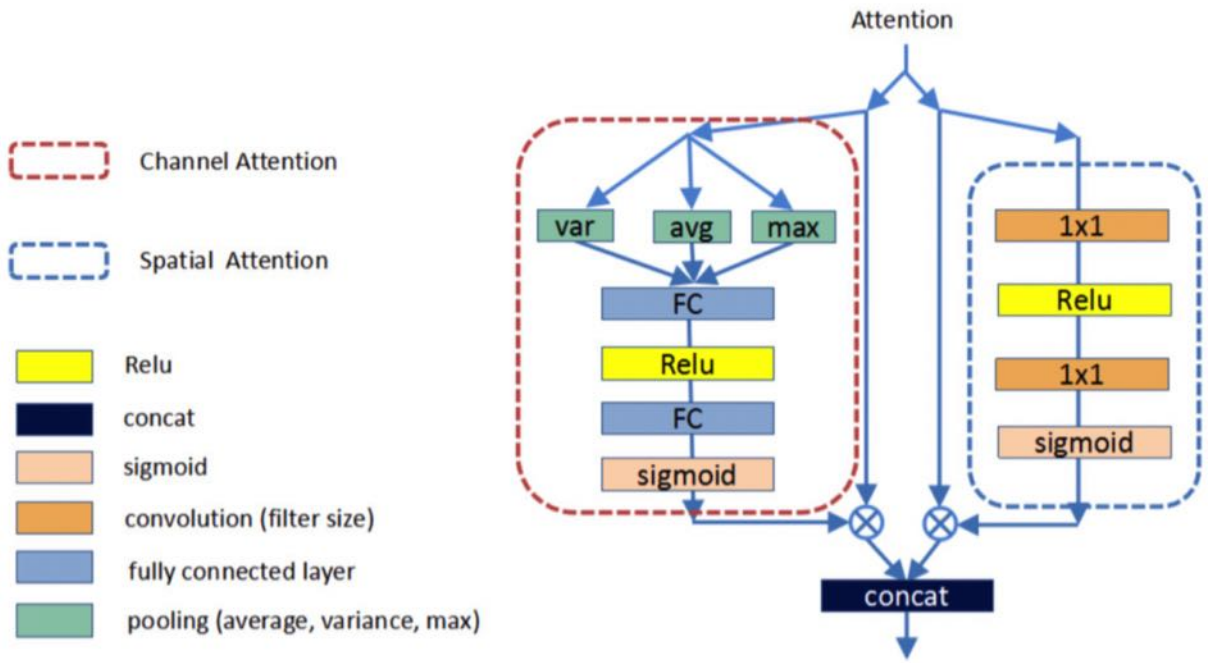
MSAB

MSAB



Multi-scale convolution: extracts multi-scale information

MSAB



- Channel attention: generates channel mask
- Spatial attention: generates spatial mask

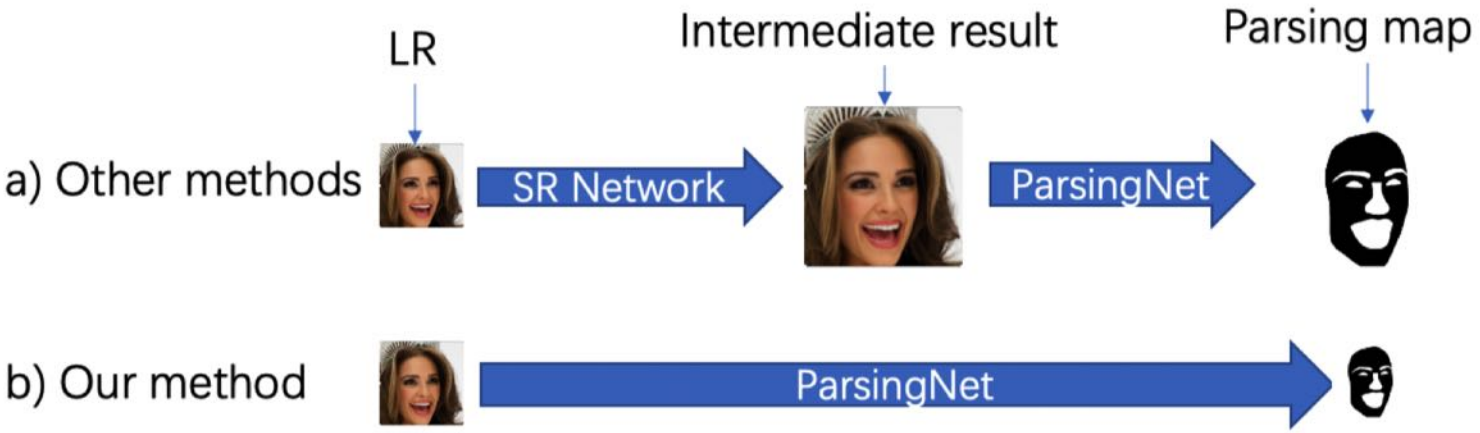
MSAB–Experiment

The effectiveness of MSAB

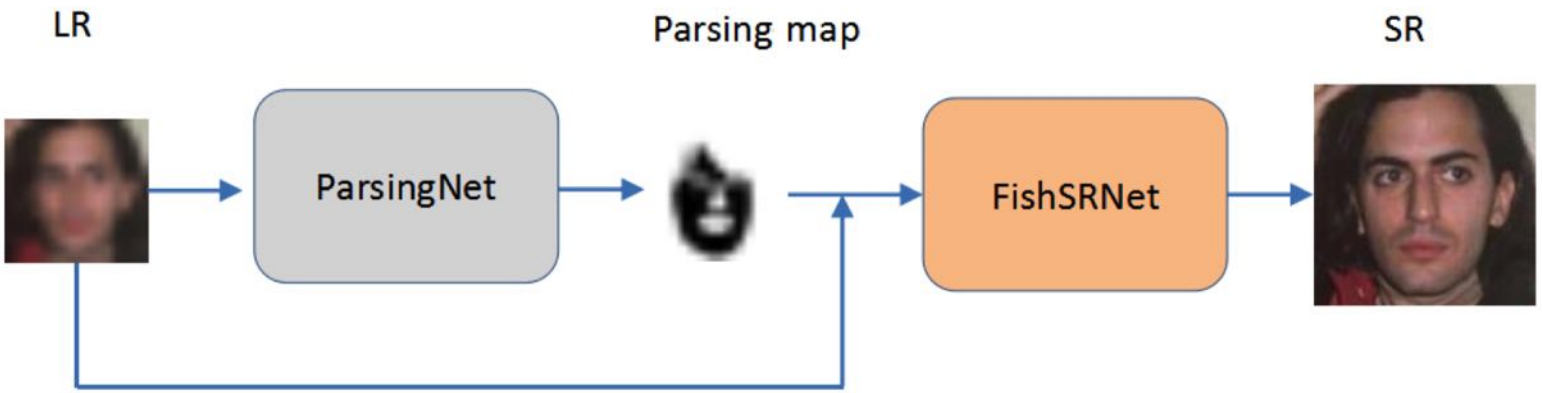
Model	PSNR	SSIM
FishSRNet + Resblock	25.26	0.8745
FishSRNet + MSAB	25.39	0.8773

ParsingNet

Existing methods: the prior knowledge derived from the intermediate results is directly affected by the quality of intermediate results.
Our method: gets prior knowledge directly from input LR faces.



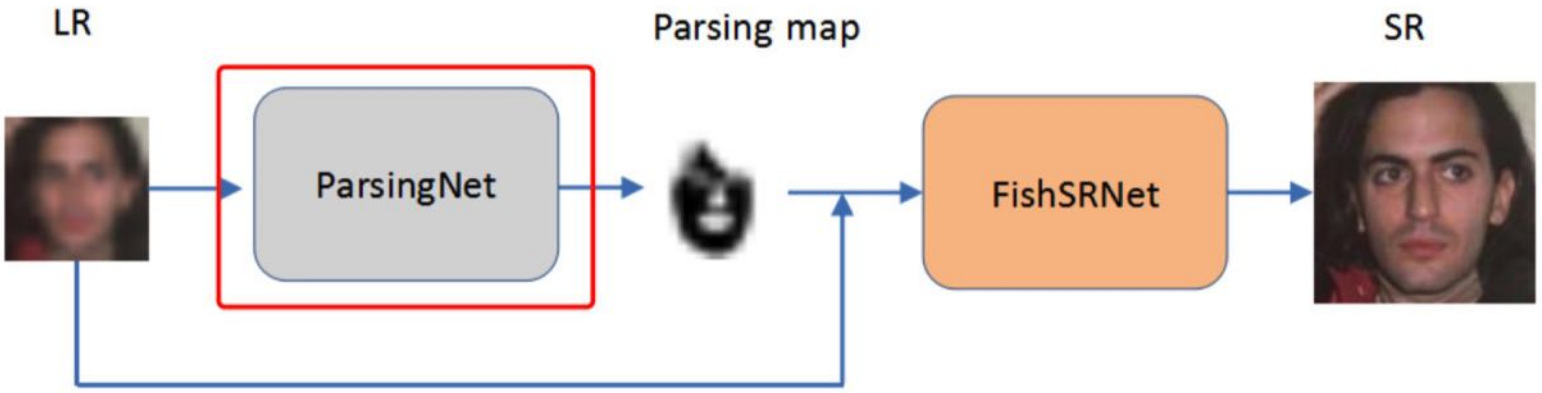
Overall Framework



Overall framework

- LR denotes the input LR face
- SR denotes the output of our network

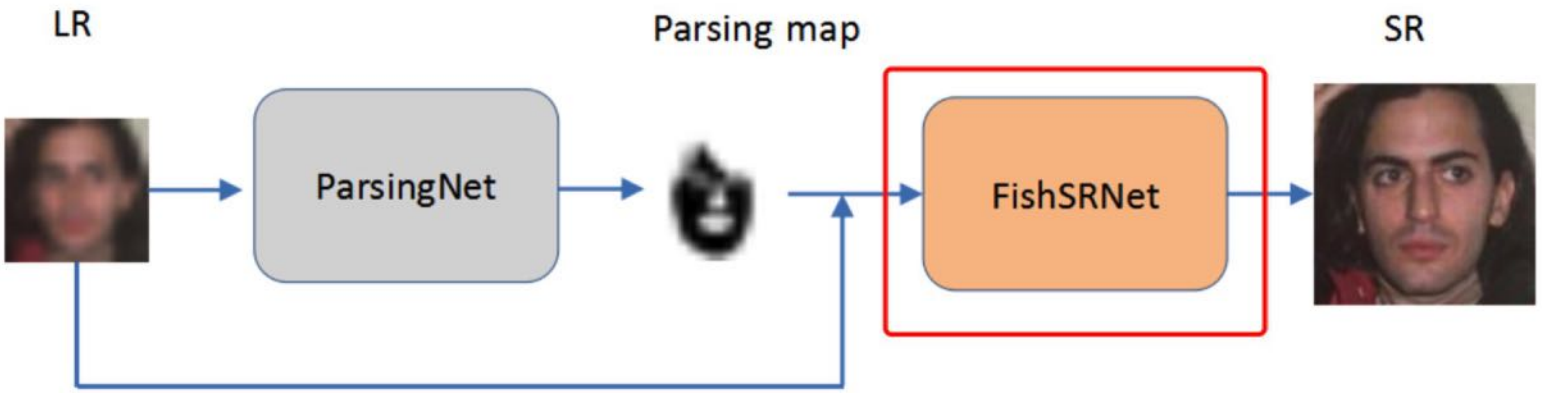
ParsingNet



Overall framework

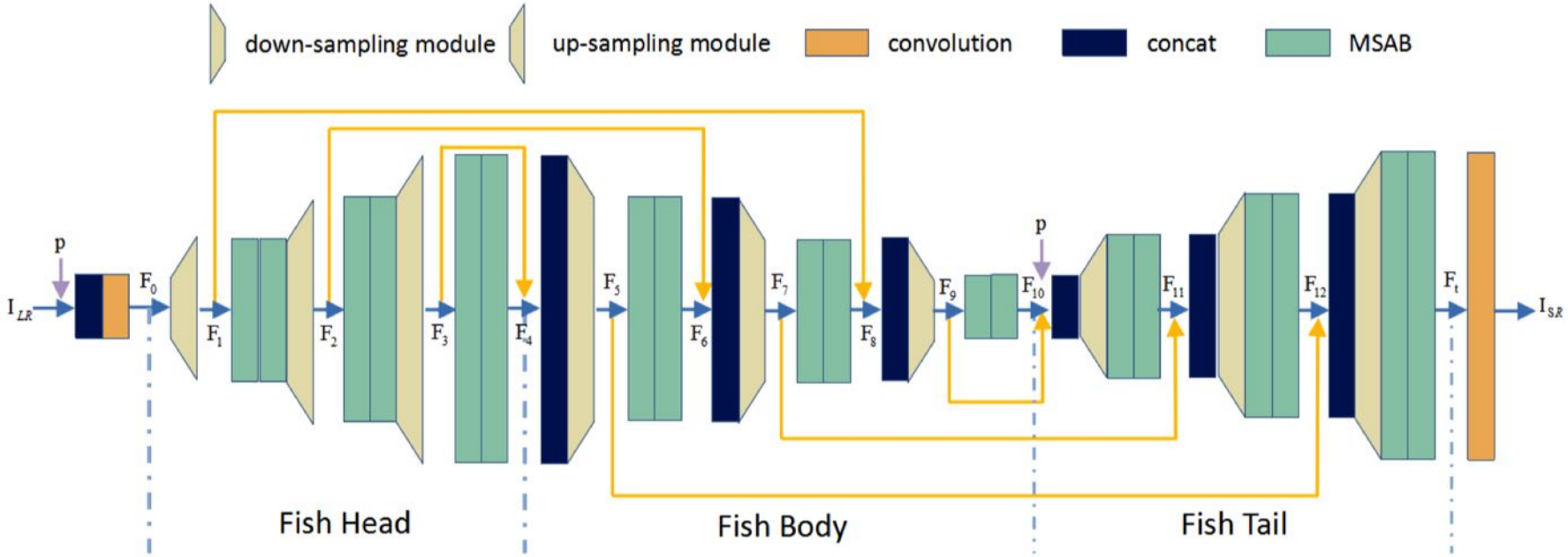
- common residual network
- parsing map: mask matrix with 0 in skin region and 255 in other components
- other components have much richer textual and structural information

Overall Framework–ParsingNet



Overall framework

Overall Framework–FishSRNet



FishSRNet with parsing map

P denotes the parsing map. We concat the paring map at the front of the FishSRNet and before the Fish Tail.

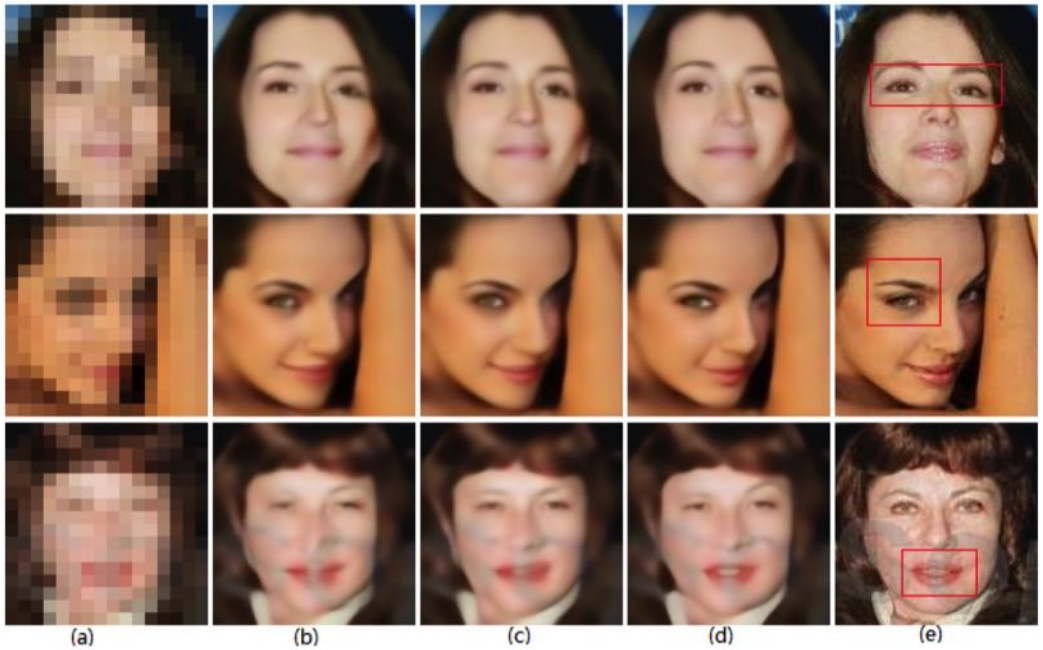
ParsingNet–Experiment

The effectiveness of ParsingNet

Model	PSNR	SSIM
FishSRNet + MSAB	25.39	0.8773
FishSRNet + MSAB + ParsingNet	25.34	0.8758

ParsingNet can't improve PSNR and SSIM.

ParsingNet–Experiment



Illustrations of influences of our different components: (a) LR. (b) The results of FishSRNet. (c) The results of FishSRNet + MSAB . (d) The results of FishSRNet + MSAB + ParsingNet. (e) Ground truth.

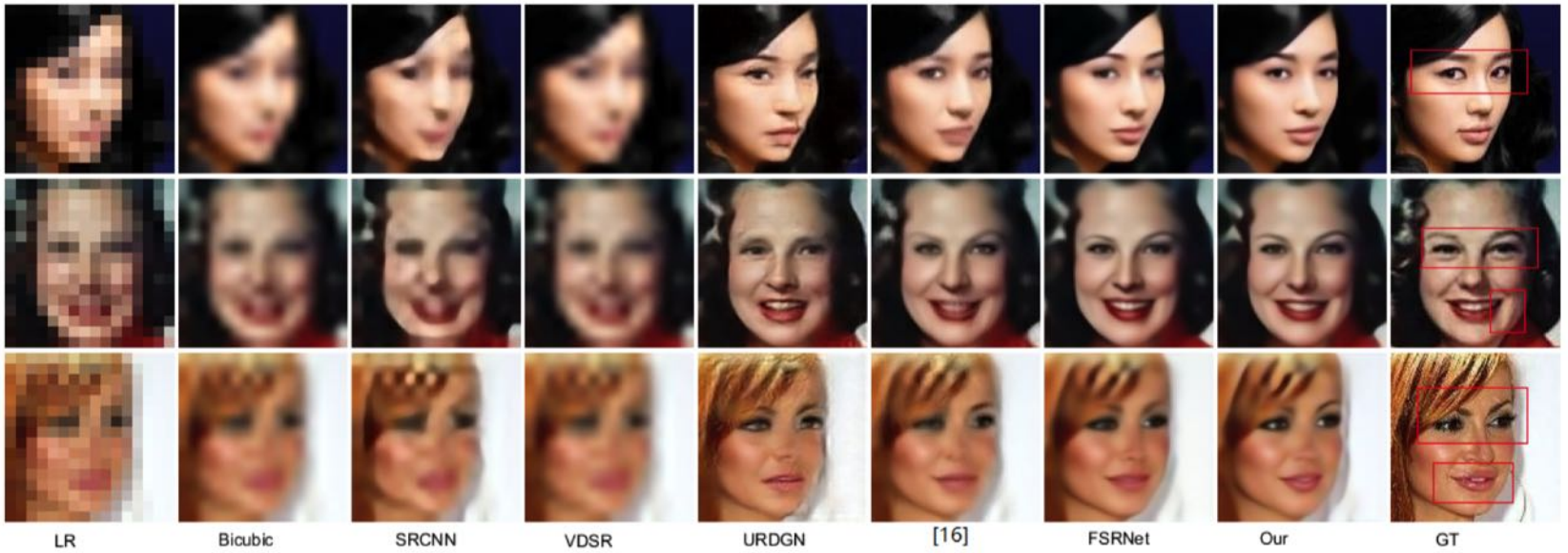
ParsingNet contributes to visual quality.

Experiment

Quantitative evaluation of various face hallucination methods

	Bicubic	SRCNN	VDSR	URDGN	[26]	FSRNet	Ours
PSNR	22.60	23.18	22.60	23.42	24.71	25.08	**25.34**
SSIM	0.8104	0.8301	0.8164	0.8375	0.8587	0.8670	**0.8758**

Experiment



Qualitative comparison of various face hallucination methods

Experiment

Failure cases

Our method exhibits poor performances when encountering special faces.

Thanks for your attention!

Any questions?