SGSR: A SALIENCY-GUIDED IMAGE SUPER-RESOLUTION NETWORK

Dayeon Kim, Munchurl Kim†

Korea Advanced Institute of Science and Technology (KAIST)

INTRODUCTION

Motivation of Research

- ✓ Humans pay attention to some essential parts or objects when viewing an image.
  - The human visual system (HVS) identifies these salient regions automatically.
  - Such salient regions play substantial roles in determining the quality of images.
- ✓ Problem?
  - Existing SR methods restore all regions of LR images in the same manner.

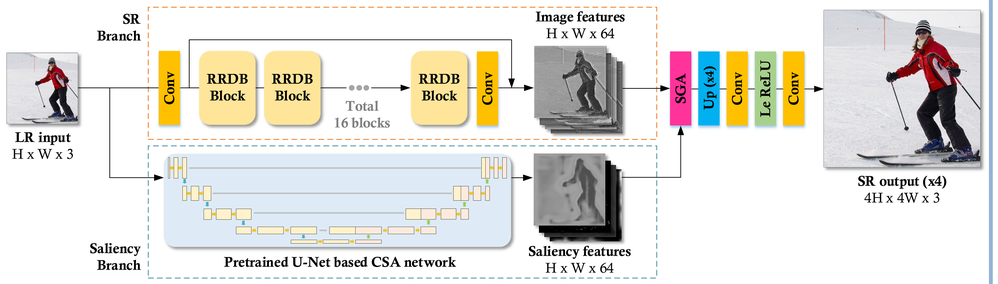
Saliency-Guided Image Super-Resolution (SGSR)

- ✓ Concentrate on salient objects while SR
  - Maximizing the image quality of salient object region
  - Suppressing excessive generation of unnecessary structures causing artifacts in background
- ✓ To determine salient object regions in an image
  - Saliency Detection: determines important parts in an given image
  - Instance Segmentation: detects and extracts object instance from an given image

To the best of our knowledge, this is the first attempt to induce discriminatory results guided by object saliency in the field of natural image SR

PROPOSED METHOD

Saliency-Guided Image Super-Resolution (SGSR) Network



- ✓ SR network to have relatively better quality in the area of important objects.
- ✓ Two branch saliency-guided SR network architecture: SR branch, Saliency branch
- ✓ Features are fused within a Saliency-Guided Attention (SGA)

Saliency Learning Scheme

- ✓ Saliency score
  - Propose a new metric which represents the saliency degree of each object area.

$$s_i = \frac{1}{2}\left(\frac{A(o_i)}{\sum_{k=1}^{N'} A(o_k)} + \frac{A(o_i) \cap \mathrm{S(I)}}{A(o_i)}\right)$$

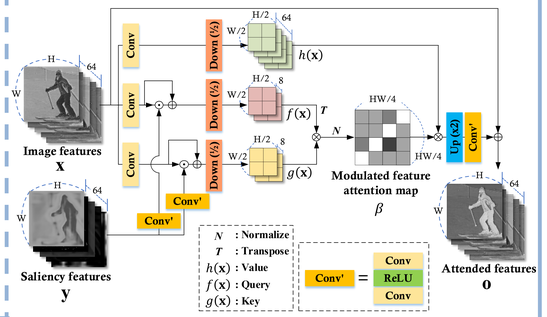
- ✓ U-Net based CSA network

Saliency-Weighted Loss Function

- ✓ The saliency of each object can be reflected with saliency score maps M.

$$L_{pix} = \frac{l_1(\mathrm{I_{HR}}, \mathrm{I_{SR}}) + l_1(\mathrm{I_{HR}} \odot M, \mathrm{I_{SR}} \odot M)}{2}$$

Saliency-Guided Attention (SGA)



- ✓ Two features are downscaled and combined in the SGA.
- ✓ SGA utilizes salient features as a guide and manipulates the image features to concentrate on salient object areas.

EXPERIMENTS

Quantitative Results

	Overall images						Most salient object regions					
	Bicubic	ESRGAN[11]	SFTGAN[13]	SPSAGAN[14]	SPSR[15]	**Ours**	Bicubic	ESRGAN[11]	SFTGAN[13]	SPSAGAN[14]	SPSR[15]	**Ours**
PSNR	25.552	25.920	25.943	24.921	24.097	**26.384**	25.053	25.725	25.604	24.708	23.930	**25.794**
SSIM	0.677	0.693	0.710	0.684	0.643	**0.726**	0.896	0.905	**0.910**	0.901	0.903	**0.910**
LPIPS	0.514	0.192	0.151	0.182	0.158	**0.137**	0.578	0.212	0.178	0.200	0.254	**0.140**

Qualitative Results

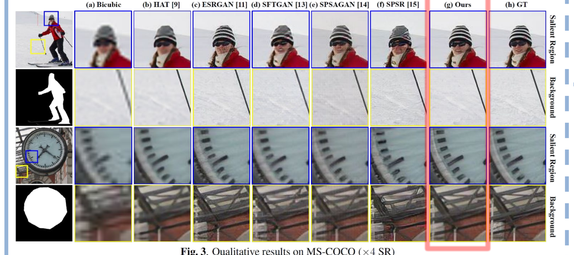


Fig. 3. Qualitative results on MS-COCO (×4 SR)

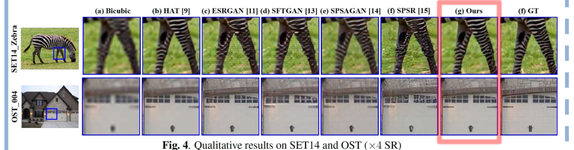


Fig. 4. Qualitative results on SET14 and OST (×4 SR)

User Study

Methods	Bicubic	ESRGAN	SFTGAN	SPSAGAN	**Ours**	HR
MOS	1.30	3.78	3.08	3.10	**4.04**	4.49

Table 2. The experimental results of the MOS test (×4 SR). The best and second best results except ground-truth HR are highlighted in **bold** and underlined.

Effect of SGA

Fig. 5. Examples of feature maps that are inputs to the proposed SGA module.

Fig. 6. The visualized results of modulated feature attention maps in the proposed SGA module. (a) Saliency score map (b) Salient object region (c) Background region

- The brighter values indicate the higher attention levels

- ✓ The proposed SGSR significantly outperforms other GAN-based methods.
- ✓ The proposed SGSR well coincides with the HVS's perceptual quality.
- ✓ The SGA analysis highlights that the SGA can lead to the saliency-guided SR.

CONCLUSION

- ✓ We propose the SGSR which performs SR to have relatively better quality in salient object regions compared to the non-salient backgrounds.
- ✓ To achieve our objective, we introduce (1) the saliency learning scheme, (2) the saliency-guided attention (SGA), and (3) the saliency-weighted loss function.
- ✓ While existing GAN-based models suffer from undesirable artifacts, our SGSR achieves stable SR results by concentrating on salient object regions and alleviating the unnecessary generative learning for the non-salient regions.