Component Attention Guided Face Super-Resolution Network: CAGFace

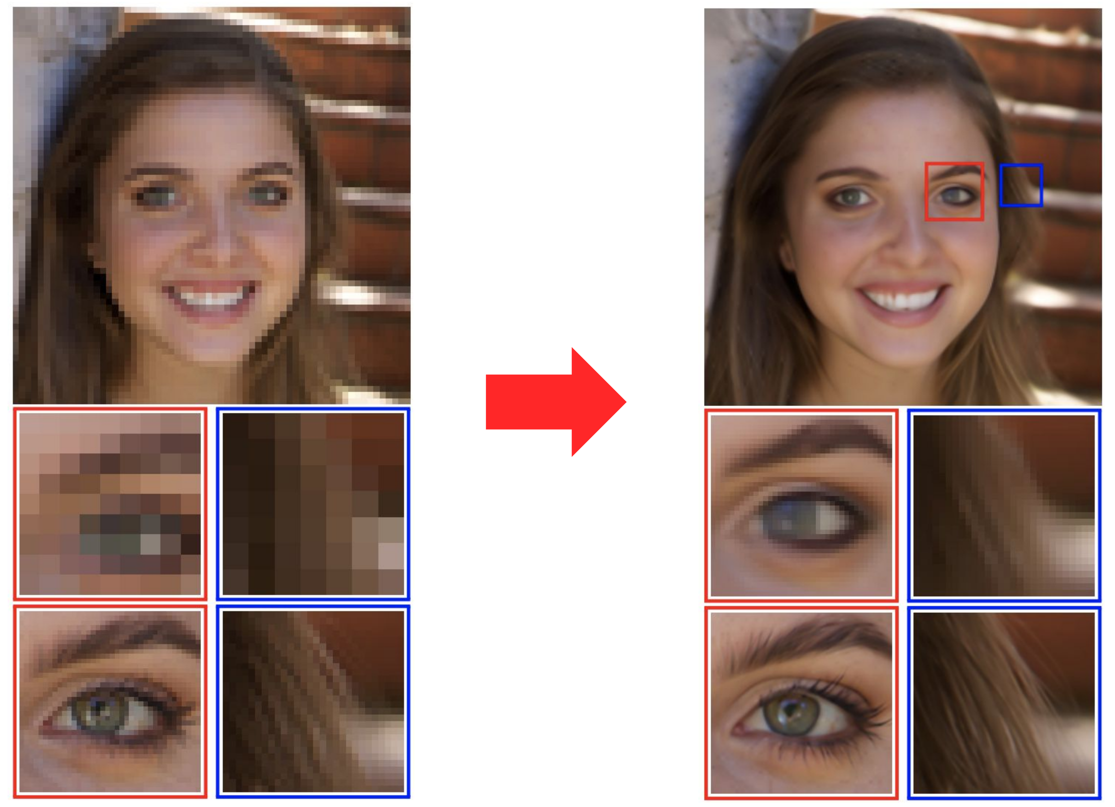
Ratheesh Kalarot kalarot@adobe.com

Tao Li taoli@purdue.edu

Fatih Porikli fatih.porikli@anu.edu.au

What is it about?

- Face Super Resolution (FSR) converts a low resolution (LR) face image to a corresponding high resolution(HR) image. We present the state-of-the-art FSR method results in a broad spectrum of real-life scenarios without inducing perceptual artifacts.



Our Contributions:

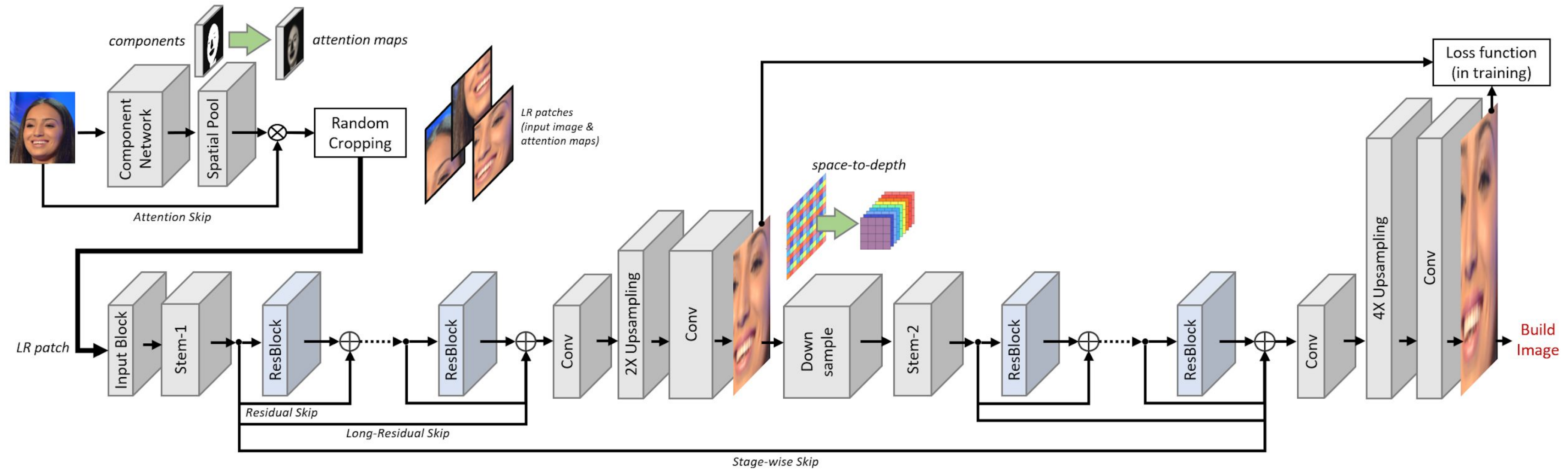
- A novel a patch-based, fully convolutional network for face image face super-resolution, where we
  - processes patches in original low-resolution throughout its backbone.
  - drives **networks attention by face component masks**.
- **Multi stage architecture** that recurrently applys the super-resolution stages to leverage on the reconstructed high-resolution outputs from the previous stage to enhance estimated high resolution details progressively.
- The experiments demonstrate **SOTA with best SSIM/PSNR/FID results** compared to existing methods and without much perceptual artifacts!

(a) Input (b) Skin (c) Hair (d) Other parts

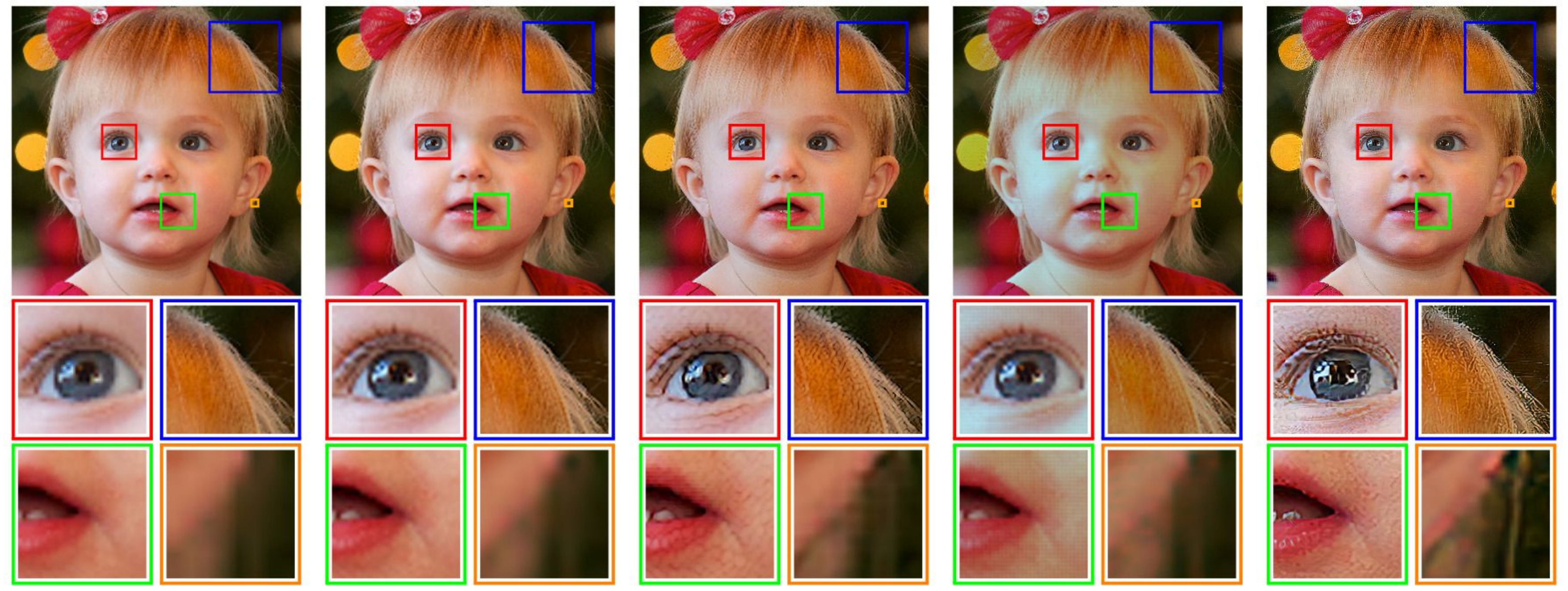
Figure 3: Sample attention maps from component network.

Method:

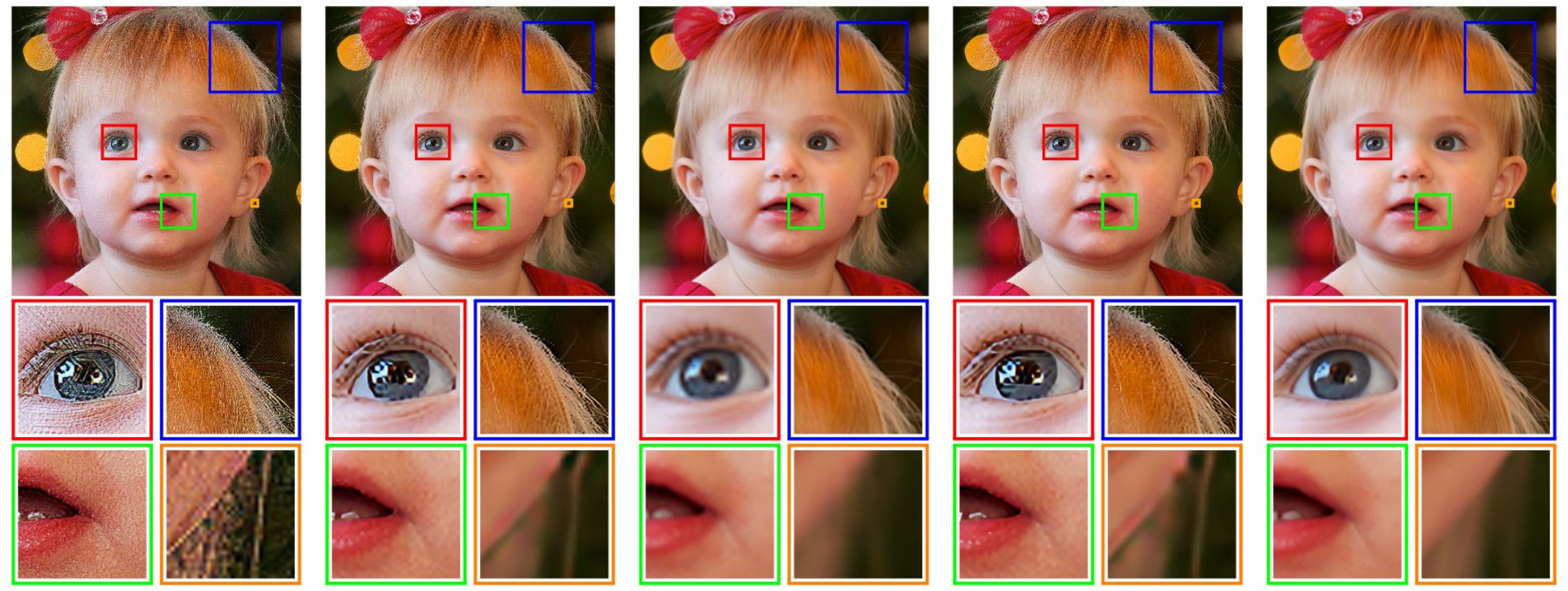
- First, facial components are segmented, and component-wise attention maps are generated. For training, random patches are sampled.
- The super-resolution network has two stages:
  - The first stage estimates a 2× intermediate HR image;
  - The second stage builds on the space-to-depth converted intermediate HR image and uses the original features of the first stem layer through a stage-wise skip-connection while implicitly imposing the component-wise attention.



(a) Input (PSNR / SSIM) (b) SRCNN [10] (22.82 / 0.668) (c) EDSR [33] (21.78 / 0.689) (d) SRGAN [30] (17.48 / 0.420) (e) E-Net [41] (23.08 / 0.679) (f) SRFBN [32] (21.12 / 0.673) (g) Ours (26.79 / 0.800)



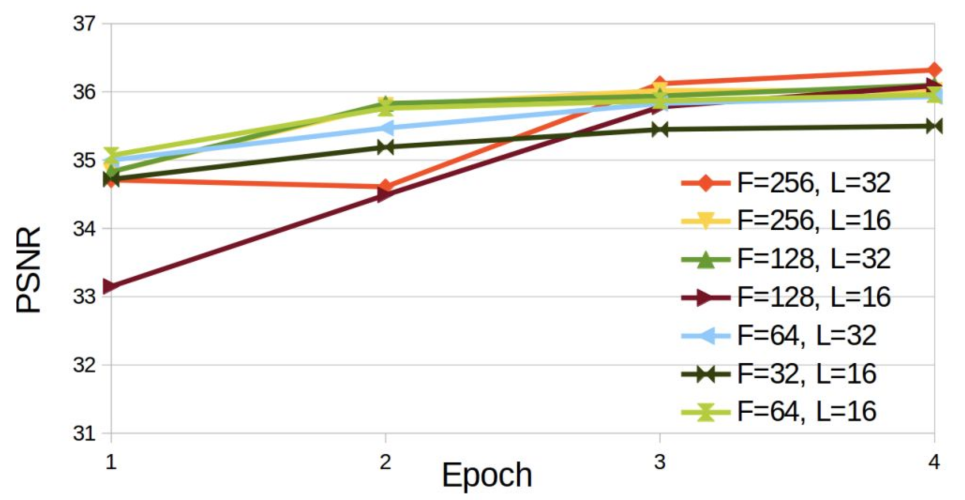
(a) Input (PSNR / SSIM) (b) Bicubic (30.96 / 0.830) (c) SRCNN [10] (27.54 / 0.750) (d) FSRCNN [10] (23.56 / 0.749) (e) SRGAN [30] (22.32 / 0.482)



(f) ESRGAN [55] (17.41 / 0.183) (g) EDSR [33] (27.14 / 0.773) (h) EnhanceNet [41] (29.24 / 0.799) (i) SRFBN [32] (26.65 / 0.765) (j) Ours (33.92 / 0.893)

	PSNR	SSIM	MS-SSIM	FID
Bicubic	31.87	0.872	0.956	**10.65**
SRCNN [10]	27.40	0.801	0.924	31.84
FSRCNN [11]	24.71	0.804	0.951	23.97
EDSR [33]	28.34	0.827	0.933	15.54
SRGAN [30]	21.49	0.515	0.807	60.67
ESRGAN [55]	19.84	0.353	0.782	72.73
EnhanceNet [41]	29.42	0.832	0.934	19.07
SRFBN [32]	27.90	0.822	0.931	17.14
Ours	**34.10**	**0.906**	**0.971**	12.40

Table 2: Comparison results for 1024×1024 outputs. Our method is trained with **patches**.



	PSNR	SSIM	MS-SSIM	FID
Bicubic	25.57	0.766	0.935	135.51
SRCNN [10]	23.12	0.688	0.900	147.21
FSRCNN [11]	22.45	0.709	0.930	139.78
EDSR [33]	22.47	0.706	0.901	129.14
SRGAN [30]	17.57	0.415	0.757	156.07
ESRGAN [55]	15.43	0.267	0.747	166.36
EnhanceNet [41]	23.64	0.701	0.897	116.38
SRFBN [32]	21.96	0.693	0.895	132.59
Ours	**27.42**	**0.816**	**0.958**	**74.43**

Table 3: Comparison results for 256×256 outputs. Our method is trained with **whole-faces**.

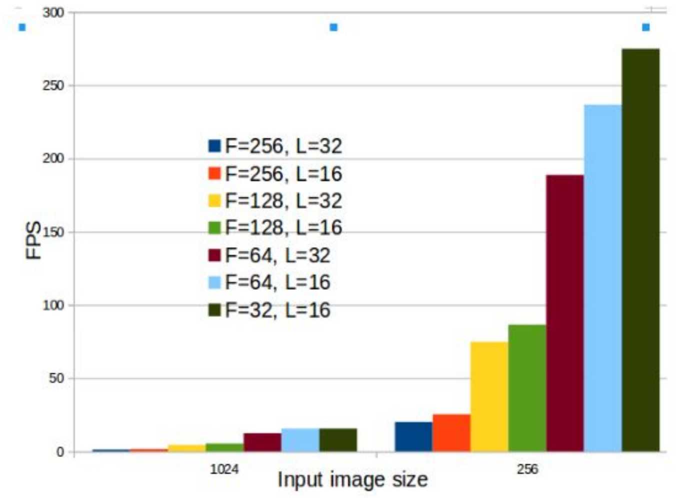


Figure 7: Effect of different network parameters on speed.