Face super-resolution with Attributes

Digital signal processing Lab
Presenter: KIM JONGHYUN

2018, CVPR

Super-Resolving Very Low-Resolution Face Images with Supplementary Attributes

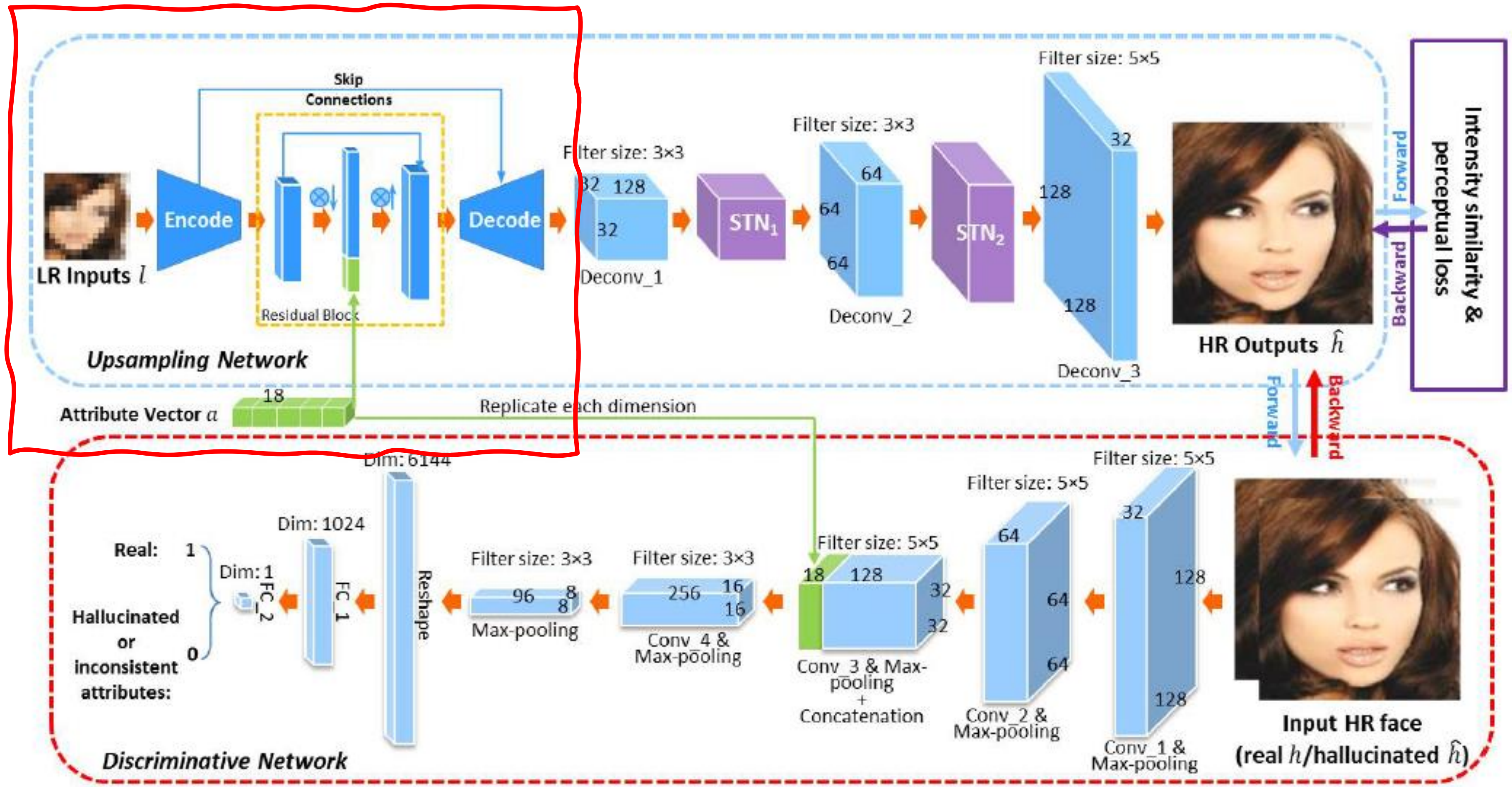
Xin Yu Basura Fernando Richard Hartley Fatih Porikli
Australian National University
xin.yu,basura.fernando,Richard.Hartley,fatih.porikli@anu.edu.au

001 Concept

1

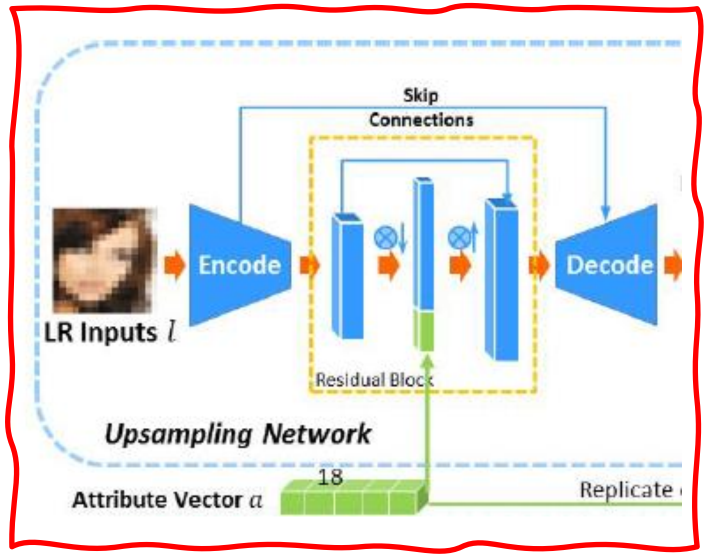
001 Concept

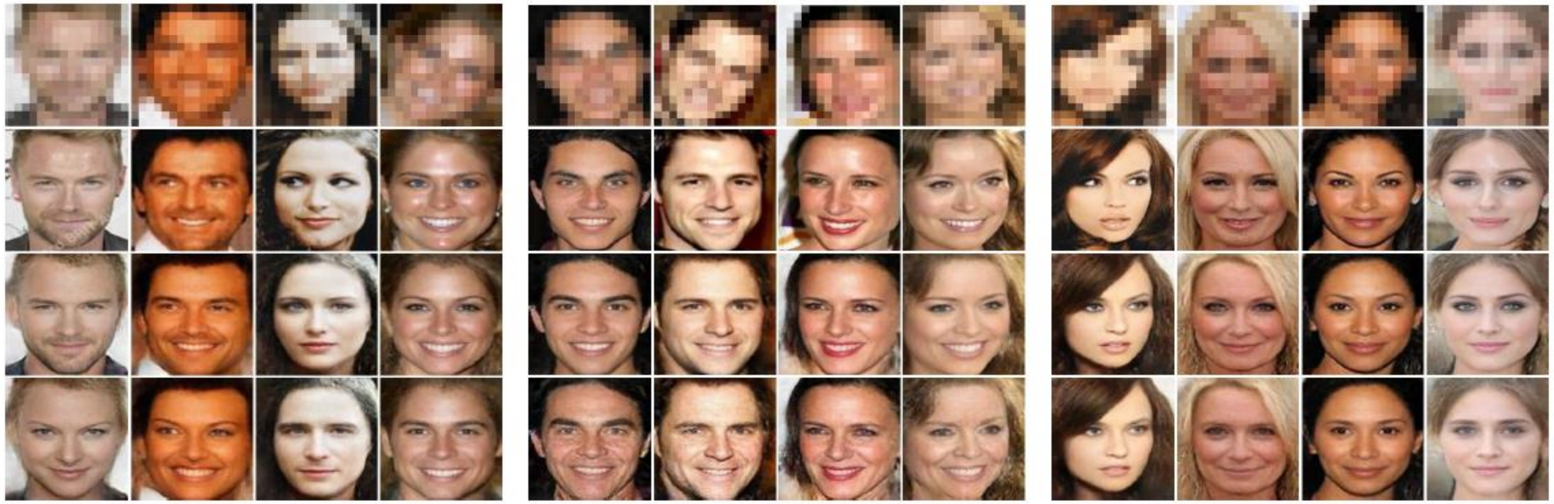
Overall



✓ Embedding semantic information into LR images

Attribute

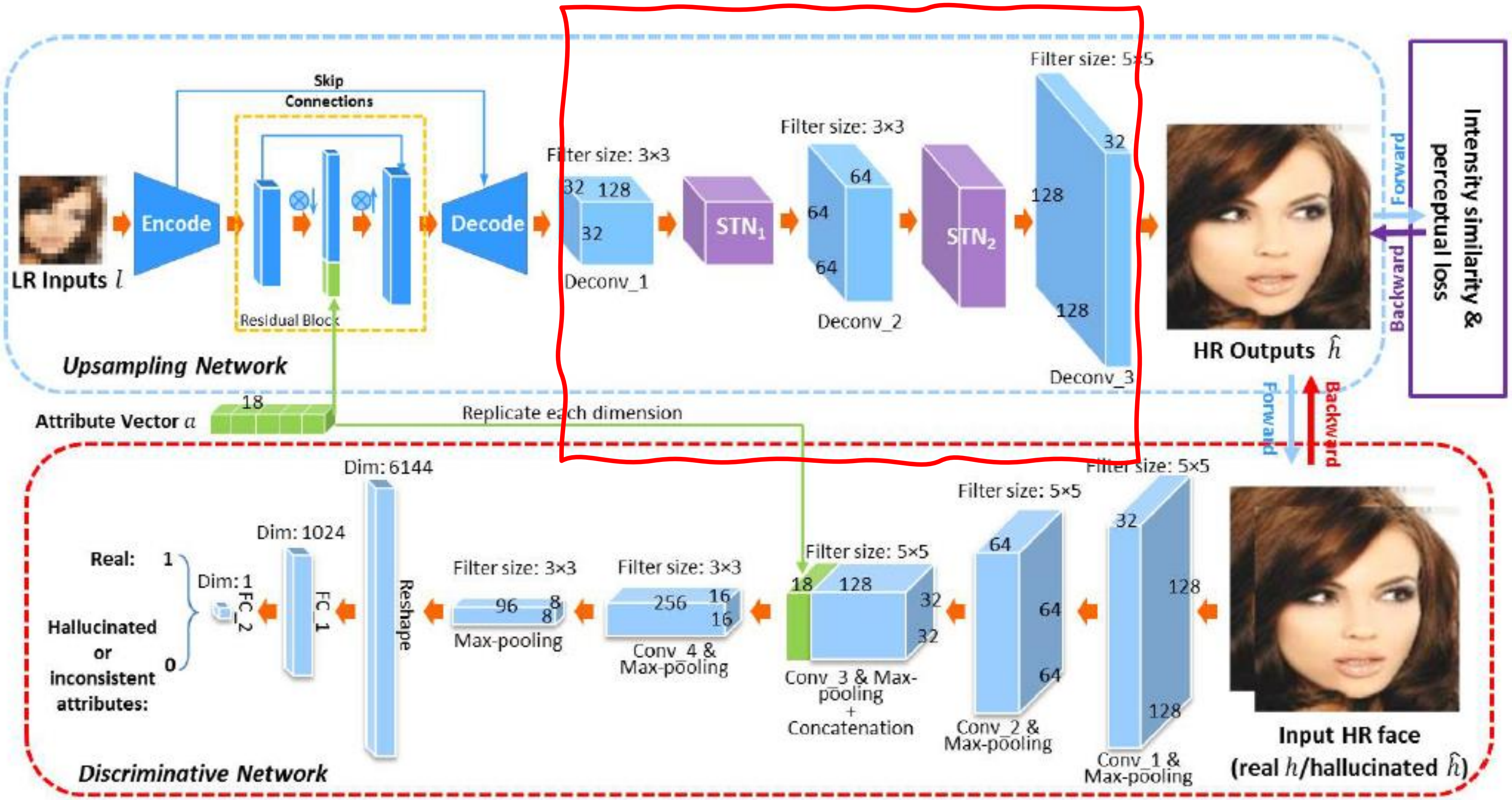




(a) Gender (b) Age (c) Makeup

- ✓ The attributes represent semantic information of input images
- ✓ 18 attributes, such as gender, age, and makeup, from 40 attributes in CelebA

Upscaling



✓ Upscaled by STN block

2015, NIPS

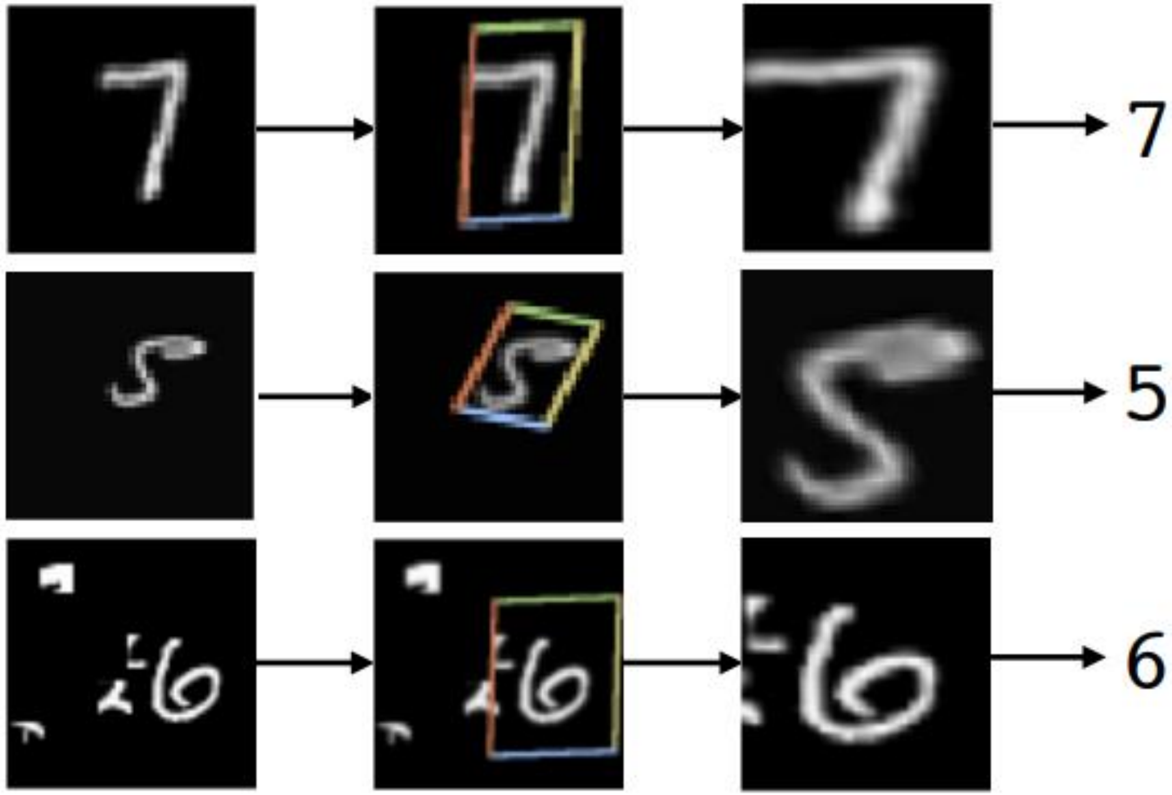
Spatial Transformer Networks

Max Jaderberg **Karen Simonyan** **Andrew Zisserman** **Koray Kavukcuoglu**

Google DeepMind, London, UK
{jaderberg,simonyan,zisserman,korayk}@google.com

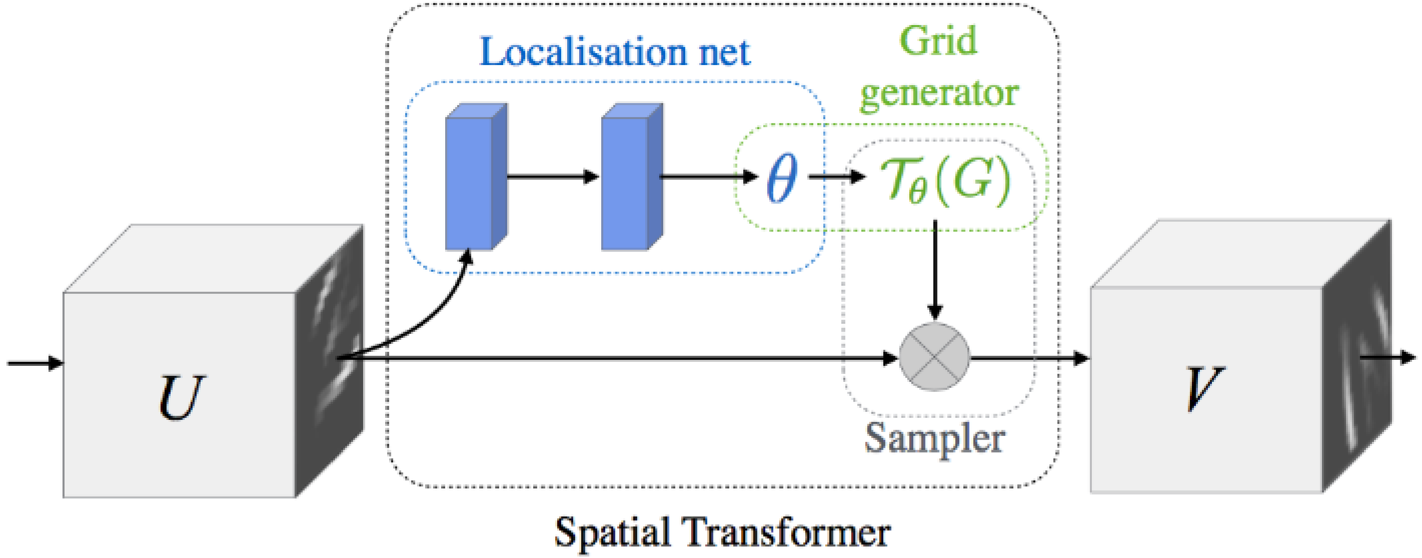
001 Concept

2015, NIPS

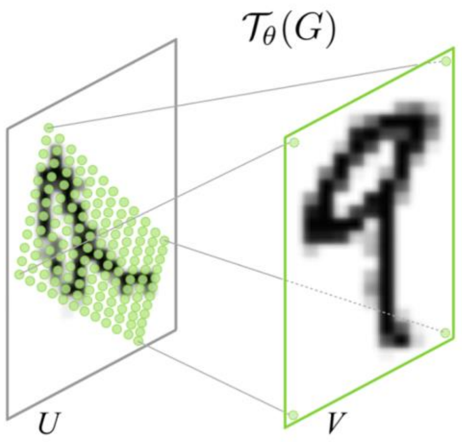


Region of interest & Spatial transformation

Upscaling

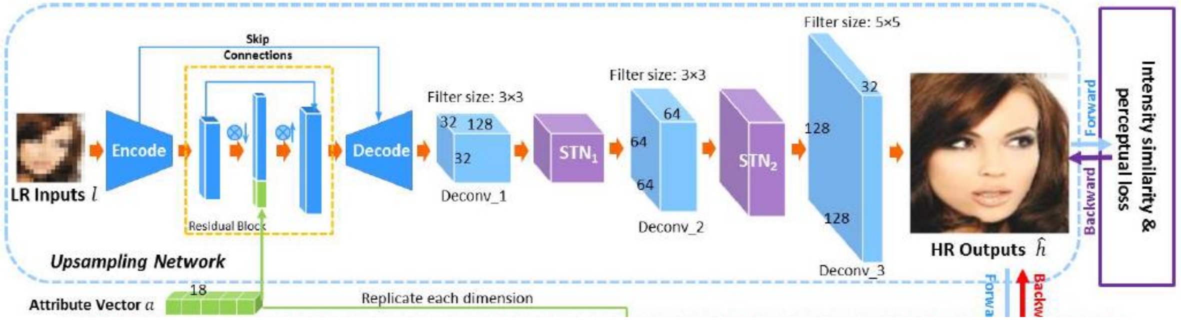


Ex) Affine transformation



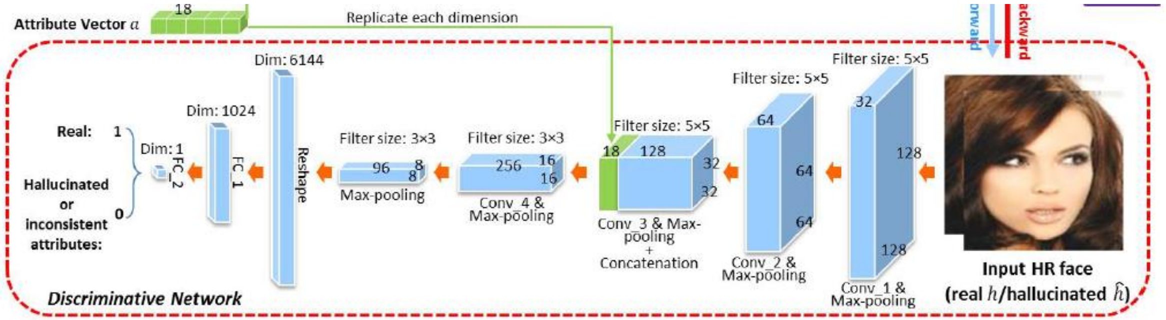
- ✓ θ is transformer methods, i.e., scale, rotation, translation, skew, cropping.
- ✓ Upscaling is used as the spatial transformer.

Training method



$$L_U = \mathbb{E}\left[\|\hat{h}-h\|_F^2 + \alpha\|\phi(\hat{h})-\phi(h)\|_F^2 - \beta \log D_d(\hat{h},a)\right]$$

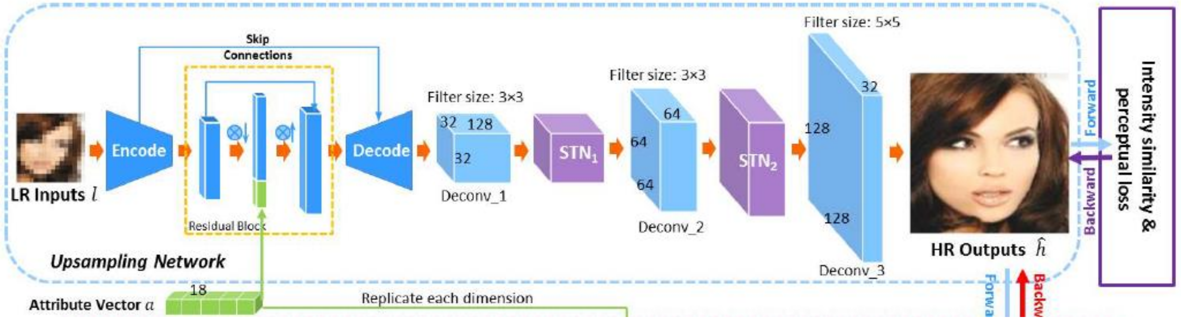
$h, \hat{h}, \emptyset, D, a$
$= [HR, SR, VGG19, Discriminator, Attributes]$



$$L_D = -\mathbb{E}\left[\log D_d(h,a)\right] - \mathbb{E}\left[\log(1-D_d(\hat{h},a)) + \log(1-D_d(h,\tilde{a}))\right]$$

$h, \hat{h}, \tilde{a} = [HR, SR, mismatched\ Attributes]$

Training method



$$L_U = \mathbb{E}\left[\|\hat{h} - h\|_F^2 + \alpha \|\phi(\hat{h}) - \phi(h)\|_F^2 - \beta \log D_d(\hat{h}, a) \right]$$

$$h, \hat{h}, \emptyset, D, a = [HR, SR, \boxed{VGG19,} Discriminator, Attributes]$$

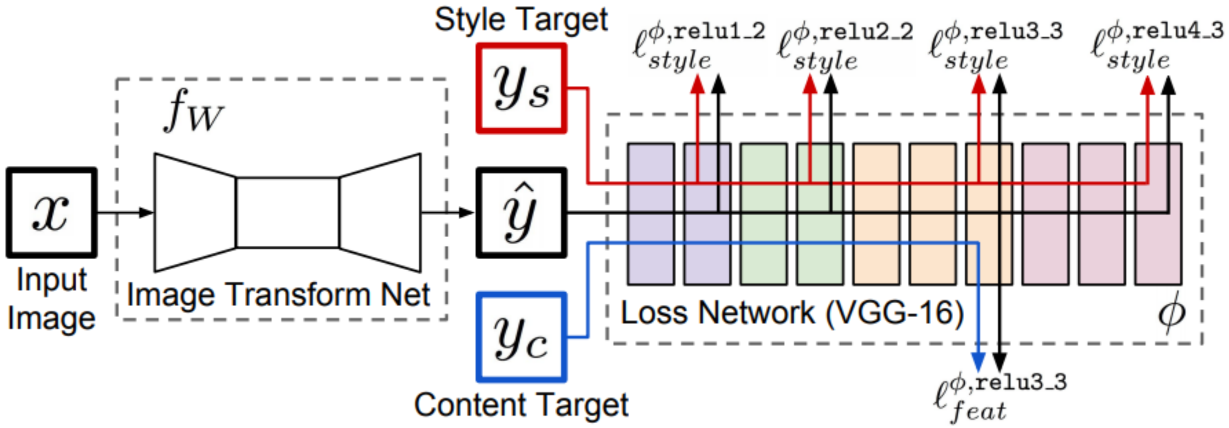
Perceptual Losses for Real-Time Style Transfer and Super-Resolution

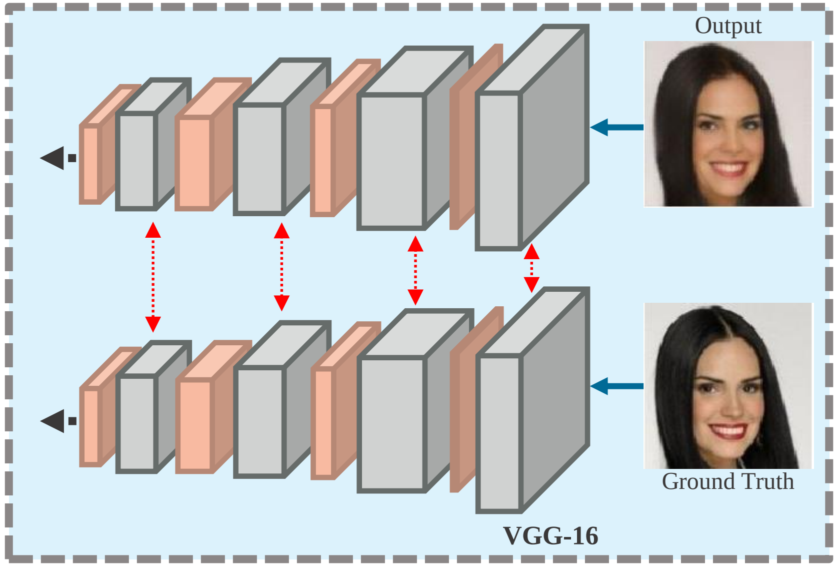
Justin Johnson, Alexandre Alahi, Li Fei-Fei
{jcjohns, alahi, feifeili}@cs.stanford.edu

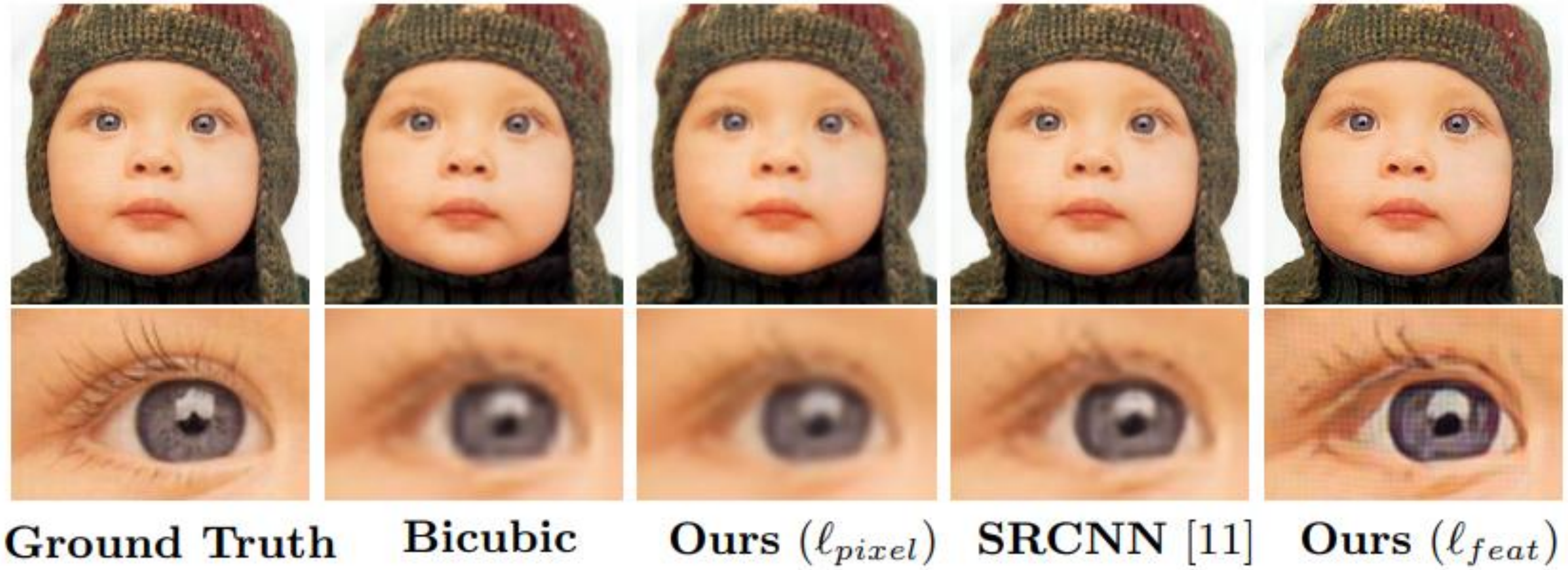
Department of Computer Science, Stanford University

001 Concept

Training method







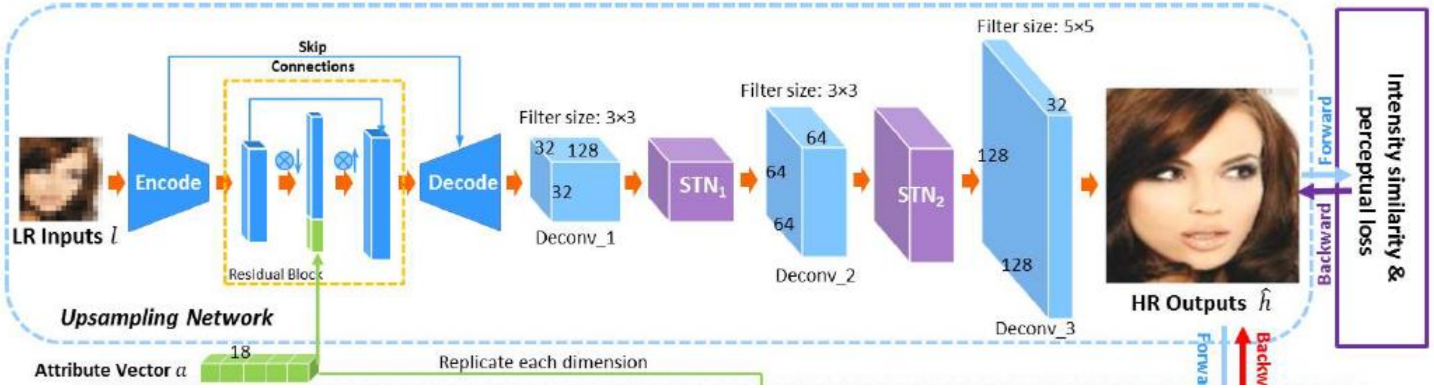
002 Result

002 Result

Ablation

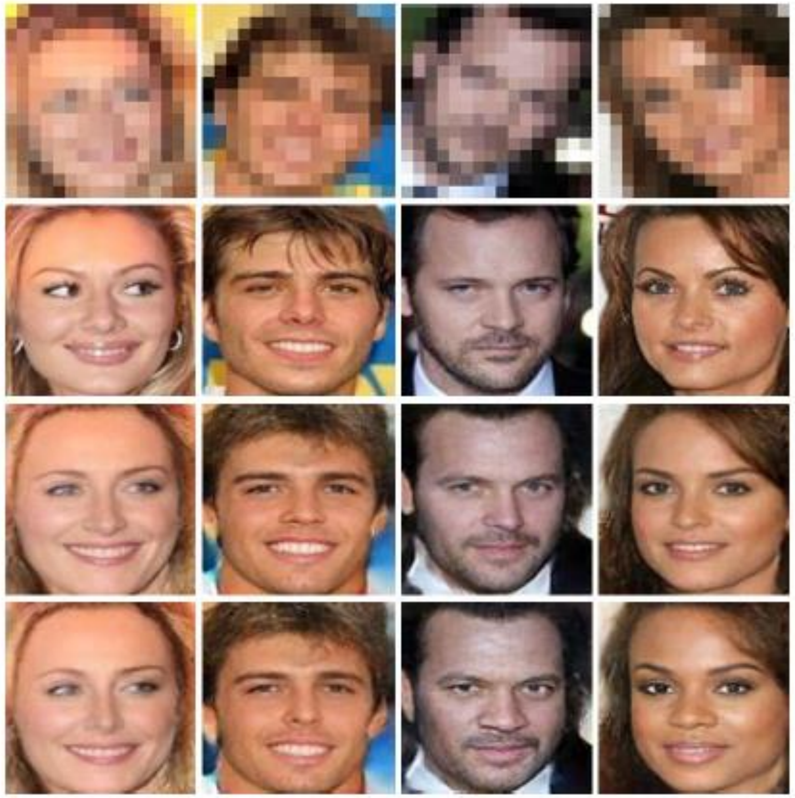
(a) (b) (c) (d) (e) (f) (g) (h)

Figure 3. Ablation study of our network. (a) 16×16 LR input image. (b) 128×128 HR ground-truth image, its ground-truth attributes are male and old. (c) Result without using an autoencoder. Here, the attribute vectors are replicated and then concatenated with the LR input directly. (d) Result without using skip connections in the autoencoder. (e) Result by only using ℓ_2 loss. (f) Result without using the attribute embedding but with a standard discriminative network. In this case, the network is similar to the decoder in [29]. (g) Result without using the perceptual loss. (h) Our final result.



- ✓ Autoencoder
- ✓ Skip connection
- ✓ Attribute

Attribute



(d) Nose

(e) Beard

(f) Eyes

- ✓ The final super-resolved results are manipulated according to user's descriptions in testing phase.

002 Result

Comparison

(a) (b) (c) (d) (e) (f) (g) (h)

Figure 5. Comparison with the state-of-the-arts methods. (a) Unaligned LR inputs. (b) Original HR images. (c) Bicubic interpolation. (d) Results of Kim *et al.*'s method (VDSR) [11]. (e) Results of Ma *et al.*'s method [23]. (f) Results of Zhu *et al.*'s method (CBN) [35]. (g) Results of Yu and Porikli's method (TDAE) [29]. (h) Our results.

Table 2. Quantitative evaluations on the test dataset.

Method	Bicubic	VDSR [11]	VDSR† [11]	Ma [23]	CBN [35]	TDAE [29]	Ours
PSNR	19.23	19.58	20.12	19.11	18.77	20.40	**21.82**
SSIM	0.56	0.57	0.57	0.54	0.54	0.57	**0.62**

THANK

YOU