VISION APPLICATIONS:

SUPER-RESOLUTION

Chih-Chung Hsu (許志仲)
Assistant Professor
ACVLab, Institute of Data Science
National Cheng Kung University
https://cchsu.info

Outline

- Overview of Deep Learning
  - Supervised – Unsupervised
- Deep super-resolution
  - Traditional super-resolution
  - Structured image super-resolution
    - Face hallucination
  - 2-D image super-resolution (generic images)
  - *N*-D image super-resolution (Hyperspectral images)
- Summary

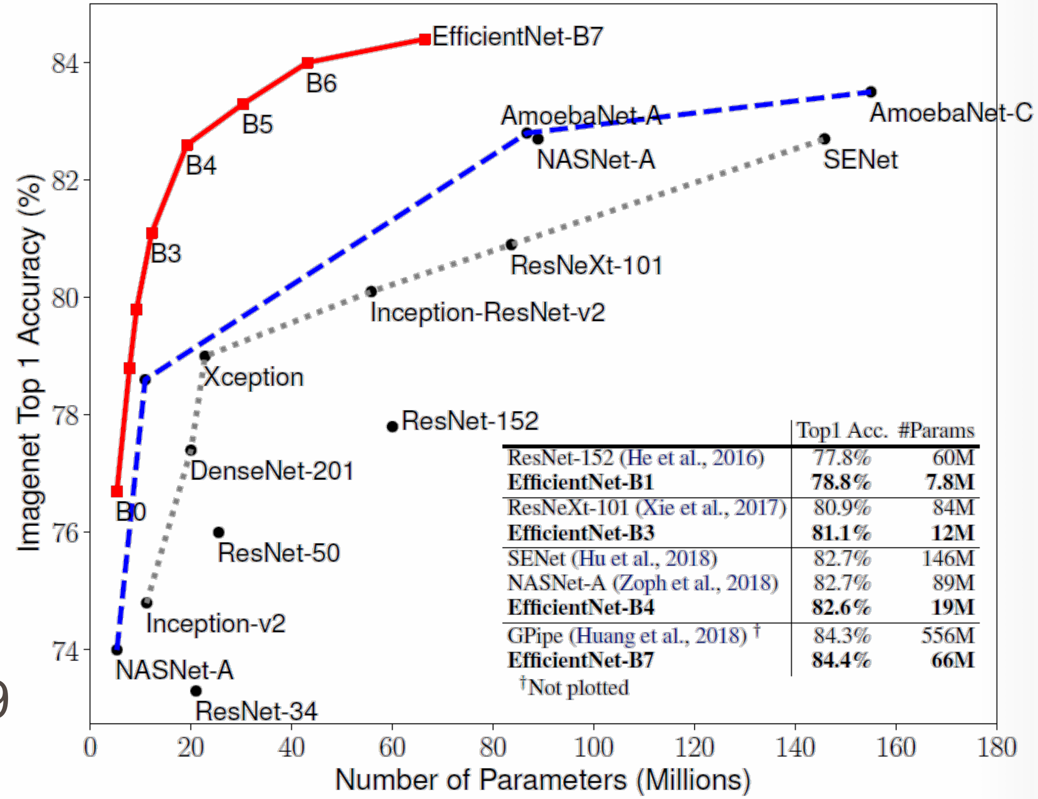
Outline

- Overview of Deep Learning
  - Supervised – Unsupervised
- Deep super-resolution
  - Traditional super-resolution
  - Structured image super-resolution
    - Face hallucination
  - 2-D image super-resolution (generic images)
  - N-D image super-resolution (Hyperspectral images)
- Summary

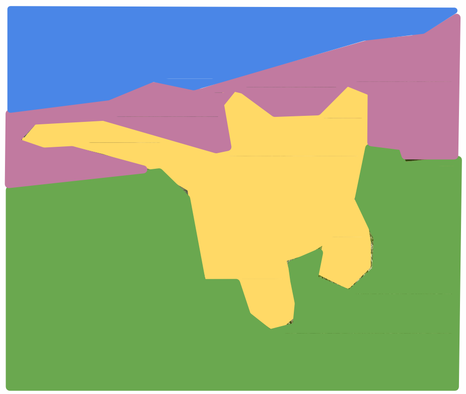
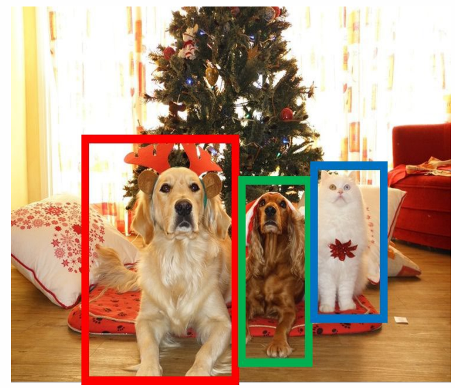
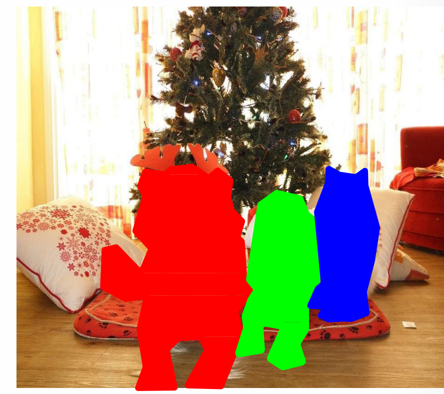
DEEP LEARNING APPLICATIONS

State-of-the-Art CNNs
(Supervised, Image Classification)

- We called those CNNs trained in supervision way are “backbone “ or “baseline” nets
- SOTA now
  - High-performance
    - ResNet
    - Wide-ResNet
    - ResNeSt
    - Swin-Transformer
    - CoAtNet [2021 late]
  - High-efficiency
    - MobileNet v3
    - EfficientNet v2
    - CSPNet (MIT!!)
- Anti-aliasing CNNs ICML19



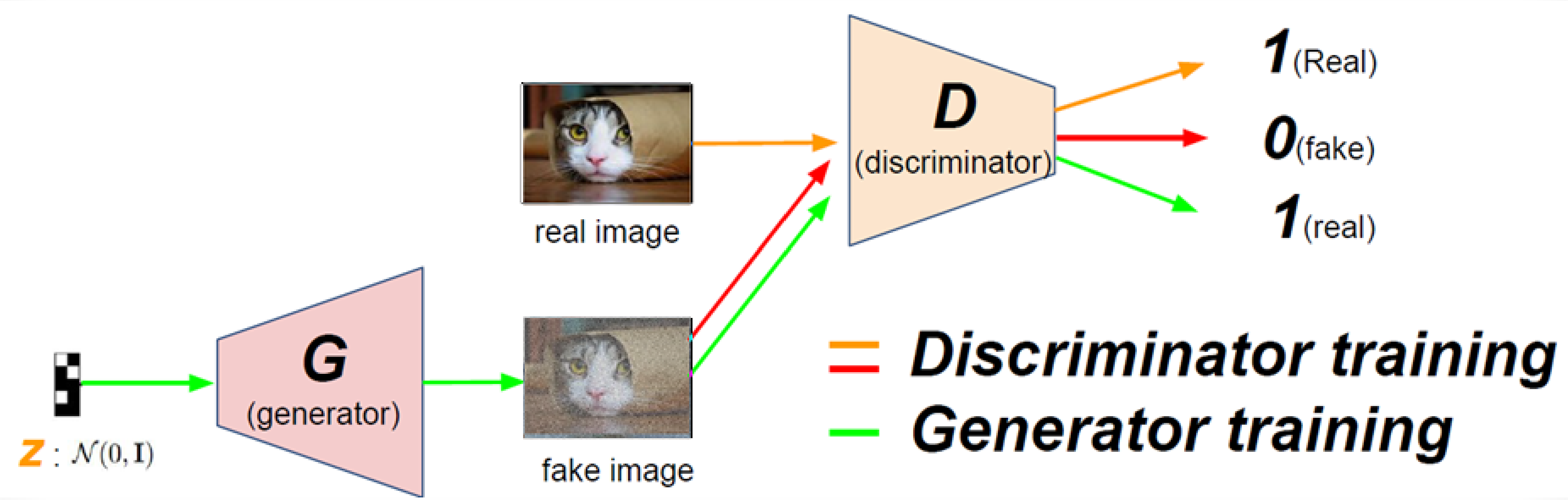
Computer Vision Applications (Supervised)

Classification	Semantic Segmentation	Object Detection	Instance Segmentation
			
CAT	GRASS, CAT, TREE, SKY	DOG, DOG, CAT	DOG, DOG, CAT
No spatial extent	No objects, just pixels	Multiple Object	

Slide credit: CS231n, Stanford

Unsupervised Deep Learning

- How to generate an image with good quality?
  - Generative adversarial network (GAN)



Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in neural information processing systems*. 2014.

Outline

- Overview of Deep Learning
  - Supervised – Unsupervised
- Deep super-resolution
  - Traditional super-resolution
  - Structured image super-resolution
    - Face hallucination
  - 2-D image super-resolution (generic images)
  - *N*-D image super-resolution (Hyperspectral images)
- Summary

IMAGE SUPER-RESOLUTION

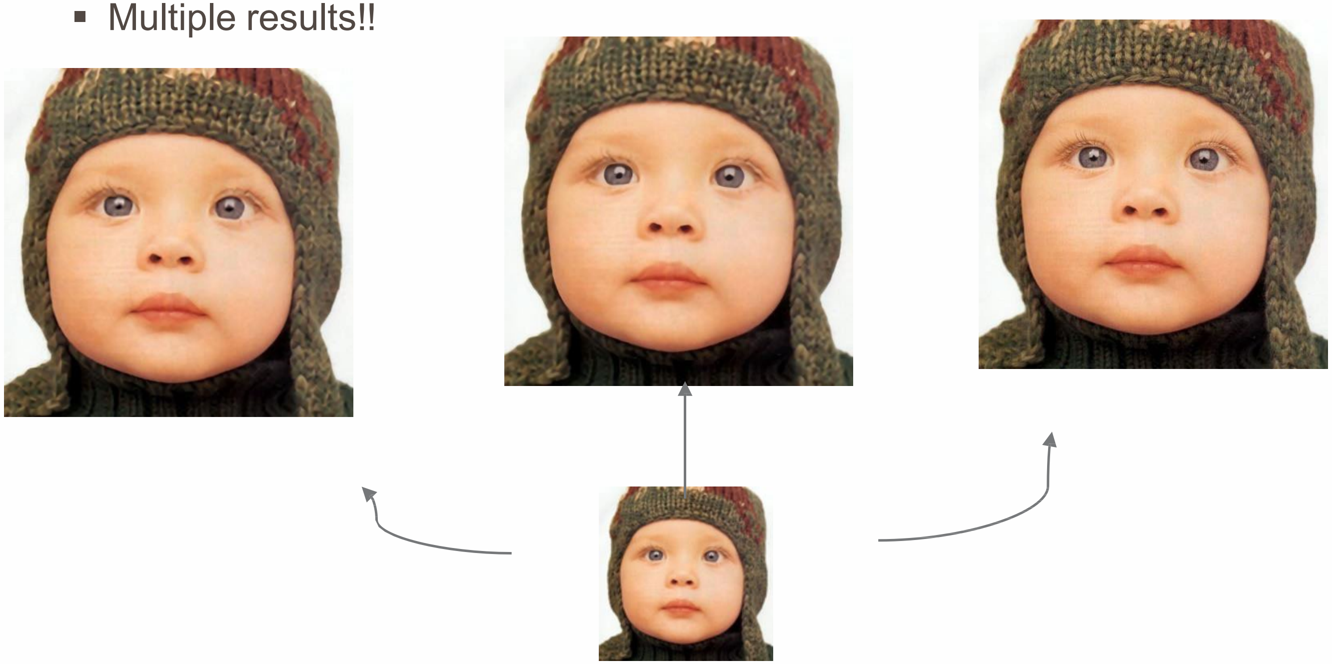
What is Super Resolution?

- Super Resolution
  - Restore High-Resolution(HR) image(or video) from Low-Resolution(LR) image(or video)
  - According to the number of input LR images, SR can be classified SISR or MISR
  - Efficient & Popular
  - Single Image Super Resolution

Super Resolution

What is Super Resolution?

- Single Image Super Resolution
  - Restore High-Resolution(HR) image(or video) from Low-Resolution(LR) image(or video)
  - Ill-Posed Problem.. (Regular Inverse Problem) → We can’t have ground truth from LR image
    - Multiple results!!



LR

What is Super Resolution?

- Interpolation-based Single Image Super Resolution
  - In image upscaling task, bicubic or bilinear or Lanczos interpolation is usually used.
  - Fast, easy.. but low quality..

Super Resolution

Deep SR

bilinear

What is Super Resolution?

- Single Image Super Resolution algorithms
  - Interpolation-based method
  - Reconstruction-based method
  - (Deep) Learning-based method

Applications of Super Resolution

- Satellite image processing
- Medical image processing
- Multimedia Industry and Video Enhancement

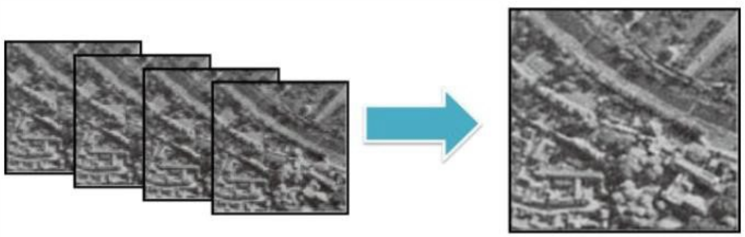


Fig 1: SR for satellite image [22]

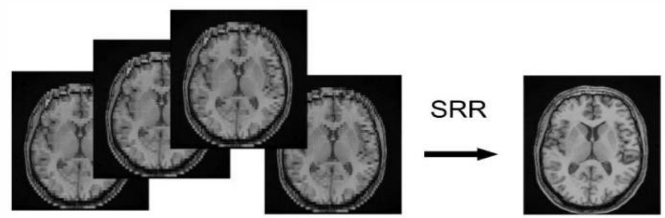
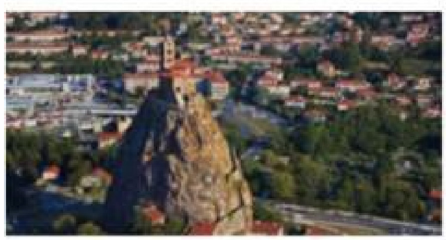


Fig 2: SR in Medical Imaging [23]

TV & Monitor



Before
HD(1280x720), FHD(1920x180)

AI Technology
UHD(3840x2160)

Reference: "Super Resolution Applications in Modern Digital Image Processing", 2016 IJCA

Deep Learning for Single Image Super Resolution

- First Deep Learning architecture for Single Image Super Resolution
- SRCNN(2014) – three-layer CNN, MSE Loss, Early upsampling
- Compared to traditional methods, it shows excellent performance.

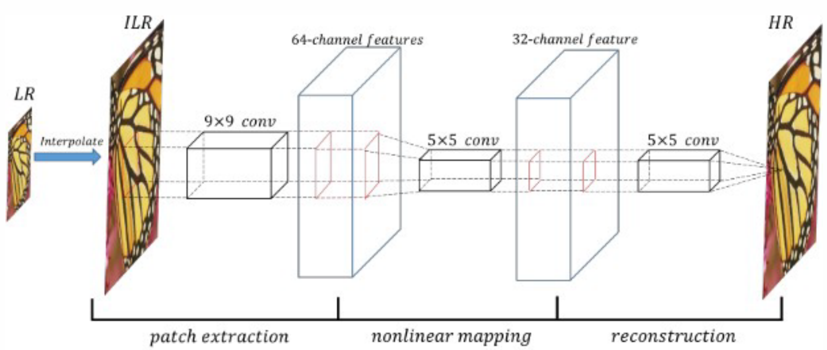
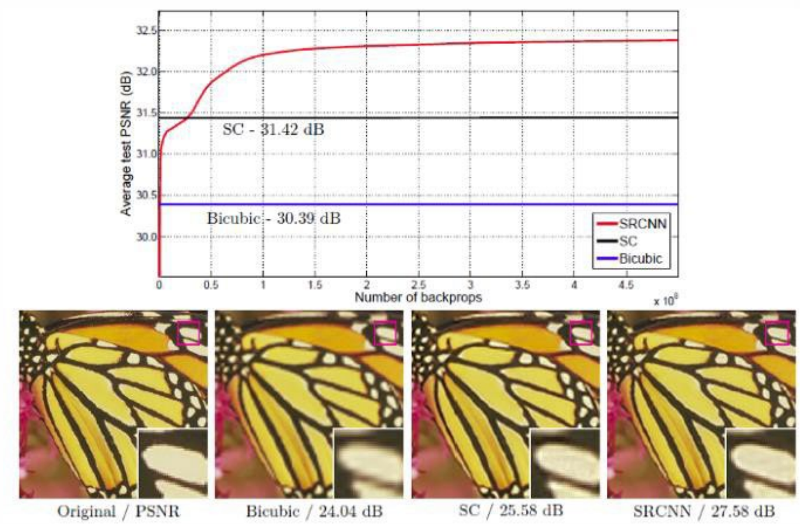


Figure 2: Sketch of the SRCNN architecture.



Reference: "Image Super-Resolution Using Deep Convolutional Networks", 2014 ECCV

Deep Learning for Single Image Super Resolution

- Efficient Single Image Super Resolution
- FSRCNN(2016), ESPCN(2016)
- Use Late Upsampling with deconvolution or sub-pixel convolutional layer

Inefficient in Memory, FLOPS

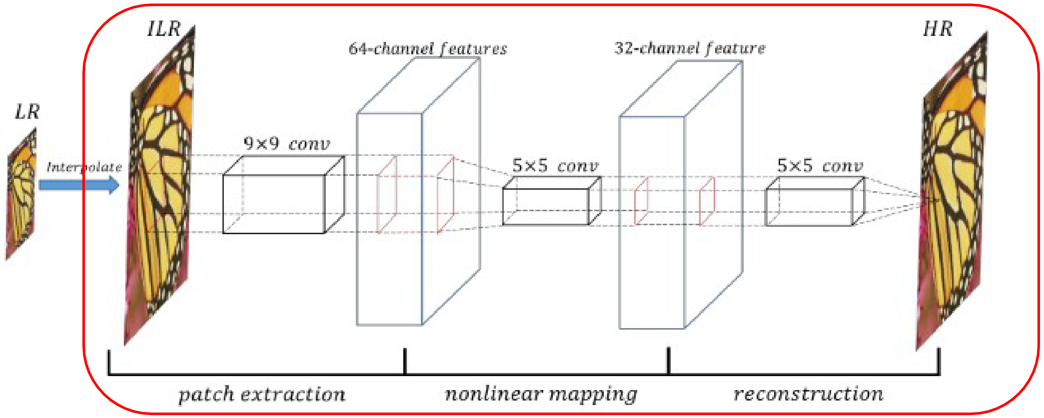
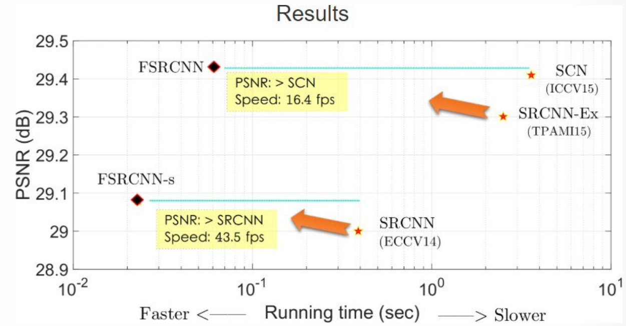


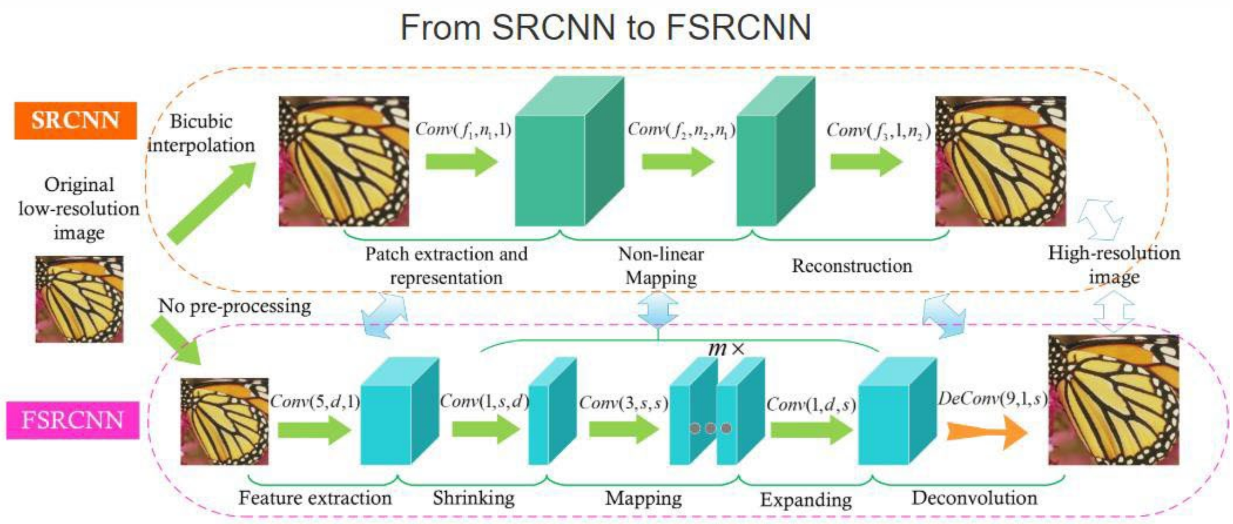
Figure 2: Sketch of the SRCNN architecture.

Reference: "Image Super-Resolution Using Deep Convolutional Networks", 2014 ECCV

Deep Learning for Single Image Super Resolution

- FSRCNN(Fast Super-Resolution Convolutional Neural Network)
  - Use Deconvolution layer instead of pre-processing(upsampling)
  - Faster and more accurate than SRCNN





Reference: "Accelerating the Super-Resolution Convolutional Neural Network", 2016 ECCV

Deep Learning for Single Image Super Resolution

- ESPCN(Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel
  - Convolutional Neural Network)
  - Use sub-pixel convolutional layer (pixel shuffler or depth_to_space)
  - This sub-pixel convolutional layer is used in recent SR models

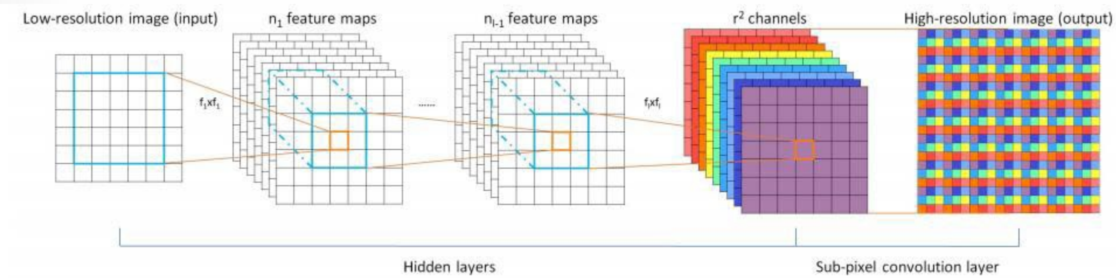


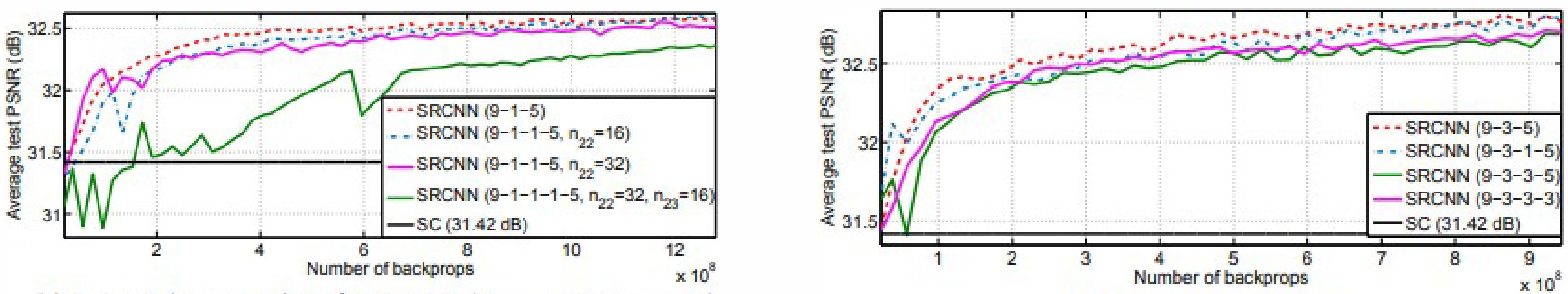
Figure 1. The proposed efficient sub-pixel convolutional neural network (ESPCN), with two convolution layers for feature maps extraction, and a sub-pixel convolution layer that aggregates the feature maps from LR space and builds the SR image in a single step.

```python
class Net(nn.Module):
    def __init__(self, upscale_factor):
        super(Net, self).__init__()

        self.conv1 = nn.Conv2d(1, 64, (5, 5), (1, 1), (2, 2))
        self.conv2 = nn.Conv2d(64, 64, (3, 3), (1, 1), (1, 1))
        self.conv3 = nn.Conv2d(64, 32, (3, 3), (1, 1), (1, 1))
        self.conv4 = nn.Conv2d(32, 1 * (upscale_factor ** 2), (3, 3), (1, 1), (1, 1))
        self.pixel_shuffle = nn.PixelShuffle(upscale_factor)

    def forward(self, x):
        x = F.tanh(self.conv1(x))
        x = F.tanh(self.conv2(x))
        x = F.tanh(self.conv3(x))
        x = F.sigmoid(self.pixel_shuffle(self.conv4(x)))
        return x
```

Reference: "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network", 2016 CVPR
Code: https://github.com/leftthomas/ESPCN

Deep Learning for Single Image Super Resolution

- Deeper Networks for Super-Resolution
  - SRCNN, FSRCNN, ESPCN are shallow network → Why not deep network?
  - Failed to train deeper models.. → Use shallow network → how to use deeper network?



(a) 9-1-1-5 ($n_{22} = 32$) and 9-1-1-1-5 ($n_{22} = 32, n_{23} = 16$)

(b) 9-3-3-5 and 9-3-3-3

Fig. 9. Deeper structure does not always lead to better results.

Reference: "Image Super-Resolution Using Deep Convolutional Networks", 2014 ECCV

Deep Learning for Single Image Super Resolution

- VDSR(Accurate Image Super-Resolution Using Very Deep Convolutional Networks)
  - VGG based deeper model(20-layer) for Super-Resolution → large receptive field
  - Residual learning & High learning rate with gradient clipping
- MSE Loss, Early upsampling

Epoch	10	20	40	80
Residual	36.90	36.64	37.12	37.05
Non-Residual	27.42	19.59	31.38	35.66
Difference	9.48	17.05	5.74	1.39

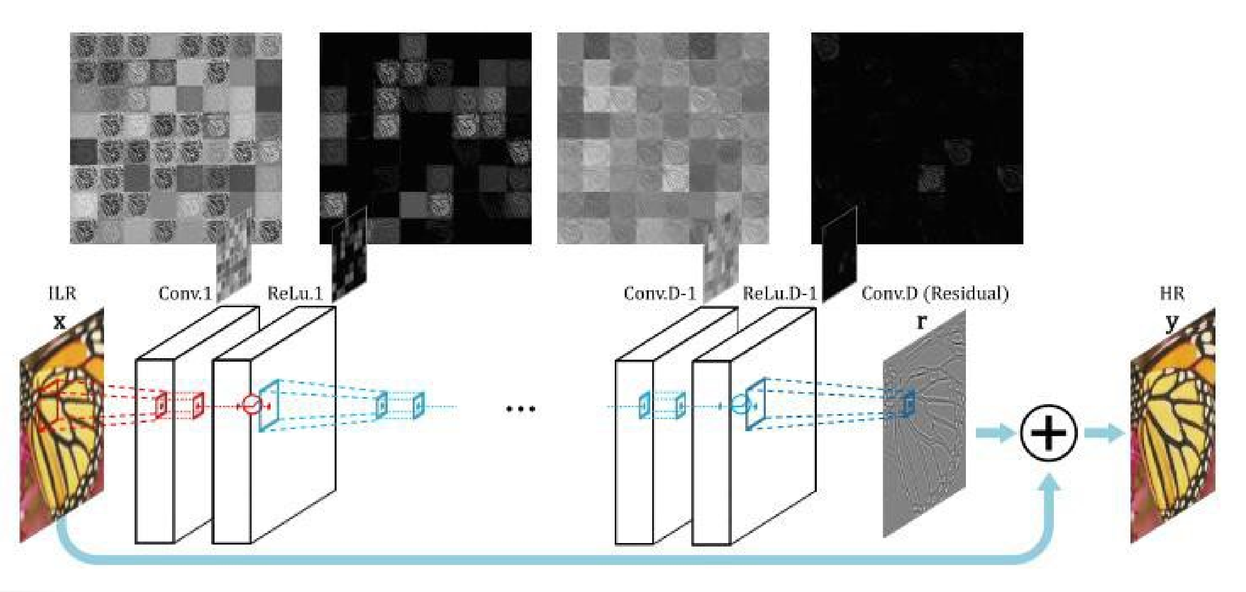
(a) Initial learning rate 0.1

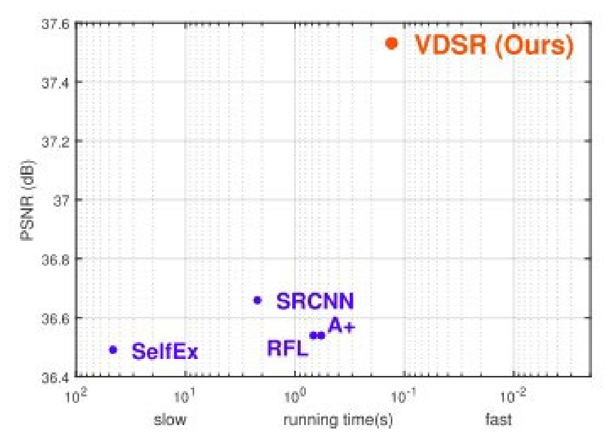
Epoch	10	20	40	80
Residual	36.74	36.87	36.91	36.93
Non-Residual	30.33	33.59	36.26	36.42
Difference	6.41	3.28	0.65	0.52

(b) Initial learning rate 0.01

Epoch	10	20	40	80
Residual	36.31	36.46	36.52	36.52
Non-Residual	33.97	35.08	36.11	36.11
Difference	2.35	1.38	0.42	0.40

(c) Initial learning rate 0.001

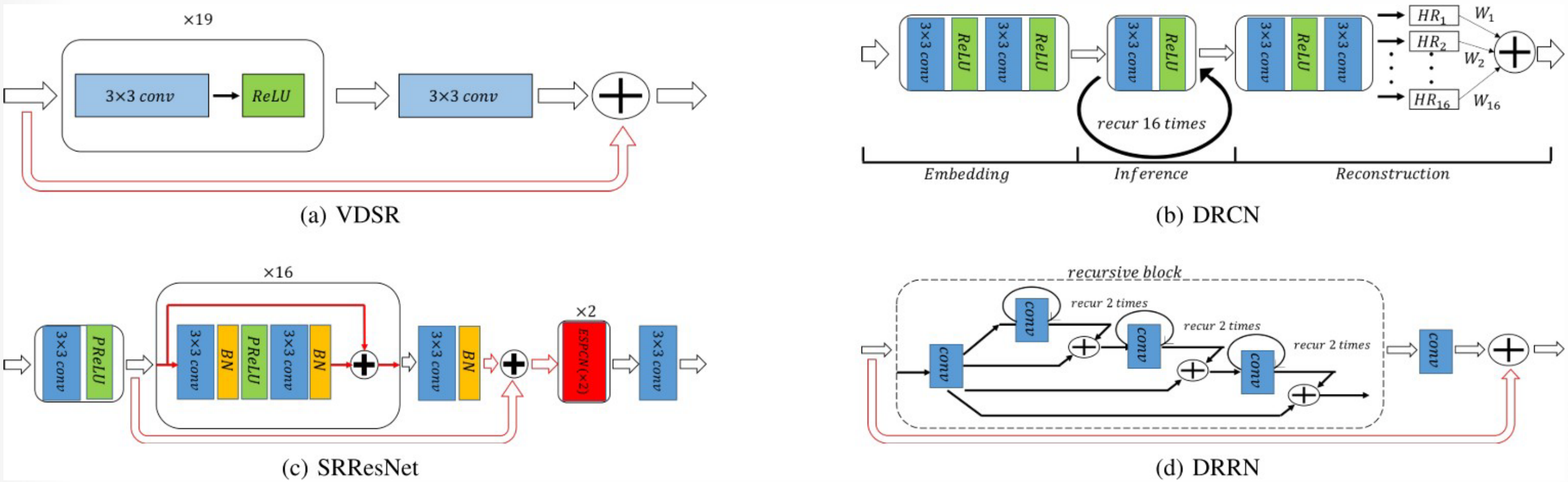




Reference: "Accurate Image Super-Resolution Using Very Deep Convolutional Networks", 2016 CVPR

Deep Learning for Single Image Super Resolution

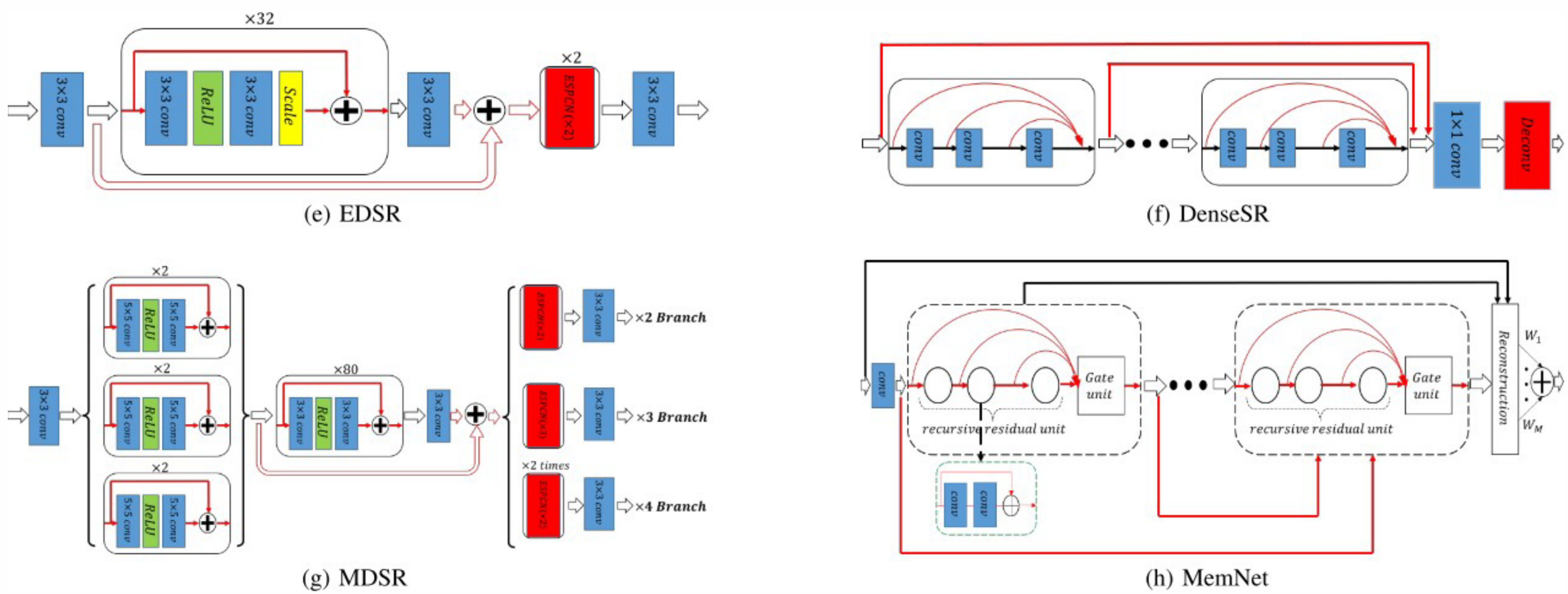
- Deeper Networks for Super-Resolution after VDSR
  - DRCN(Deeply-recursive Convolutional network), 2016 CVPR
  - SRResNet, 2017 CVPR
  - DRRN(Deep Recursive Residual Network), 2017 CVPR



Reference: “Deep Learning for Single Image Super-Resolution: A Brief Review”, 2018 IEEE Transactions on Multimedia (TMM)

Deep Learning for Single Image Super Resolution

- Deeper Networks for Super-Resolution after VDSR
  - EDSR, MDSR (Enhanced Deep Residual Network, Multi Scale EDSR), 2017 CVPRW
  - DenseSR, 2017 CVPR
  - MemNet, 2017 CVPR

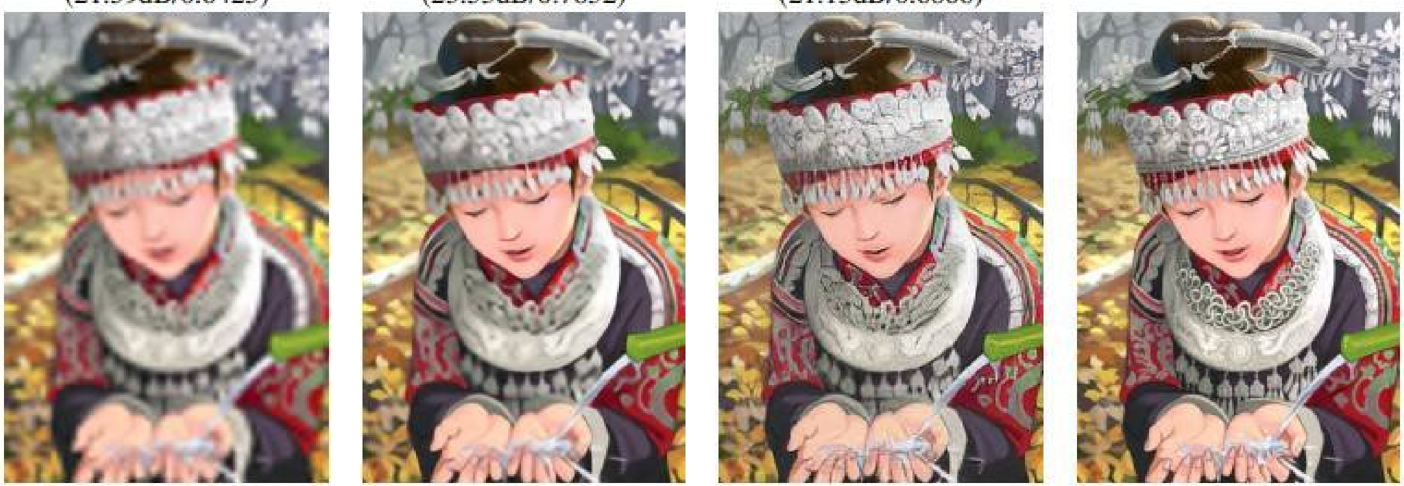


Reference: "Deep Learning for Single Image Super-Resolution: A Brief Review", 2018 IEEE Transactions on Multimedia (TMM)

Deep Learning for Single Image Super Resolution

- Generative Adversarial Network(GAN) for Super-Resolution
  - SRGAN(Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network)
  - First GAN-based SR Model, MSE Loss → Blurry Output → GAN loss + Content loss = Perceptual loss

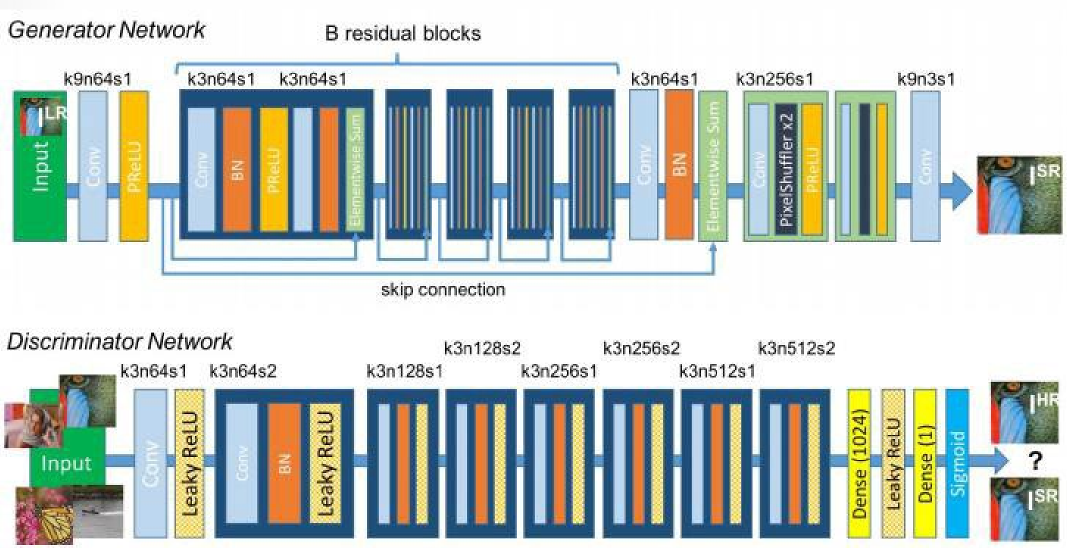
bicubic (21.59dB/0.6423)	SRResNet (23.53dB/0.7832)	SRGAN (21.15dB/0.6868)	original



Reference: "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network", 2017 CVPR

Deep Learning for Single Image Super Resolution

- Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network
  - MSE Loss → Blurry Output → GAN loss + Content loss = Perceptual loss
  - Replace MSE loss to VGG loss (used in style transfer) and add adversarial loss



$$\cancel{l^{SR}_{MSE} = \frac{1}{r^2WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}(I^{HR}_{x,y} - G_{\theta_G}(I^{LR})_{x,y})^2} \quad (4)$$

$$l^{SR}_{VGG/i.j} = \frac{1}{W_{i,j}H_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}(\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2 \quad (5)$$

$$l^{SR}_{Gen} = \sum_{n=1}^{N} -\log D_{\theta_D}(G_{\theta_G}(I^{LR})) \quad (6)$$

$$\boxed{l^{SR} = \underbrace{\underbrace{l^{SR}_{X}}_{\text{content loss}} + \underbrace{10^{-3} l^{SR}_{Gen}}_{\text{adversarial loss}}}_{\text{perceptual loss (for VGG based content losses)}} \quad (3)}$$

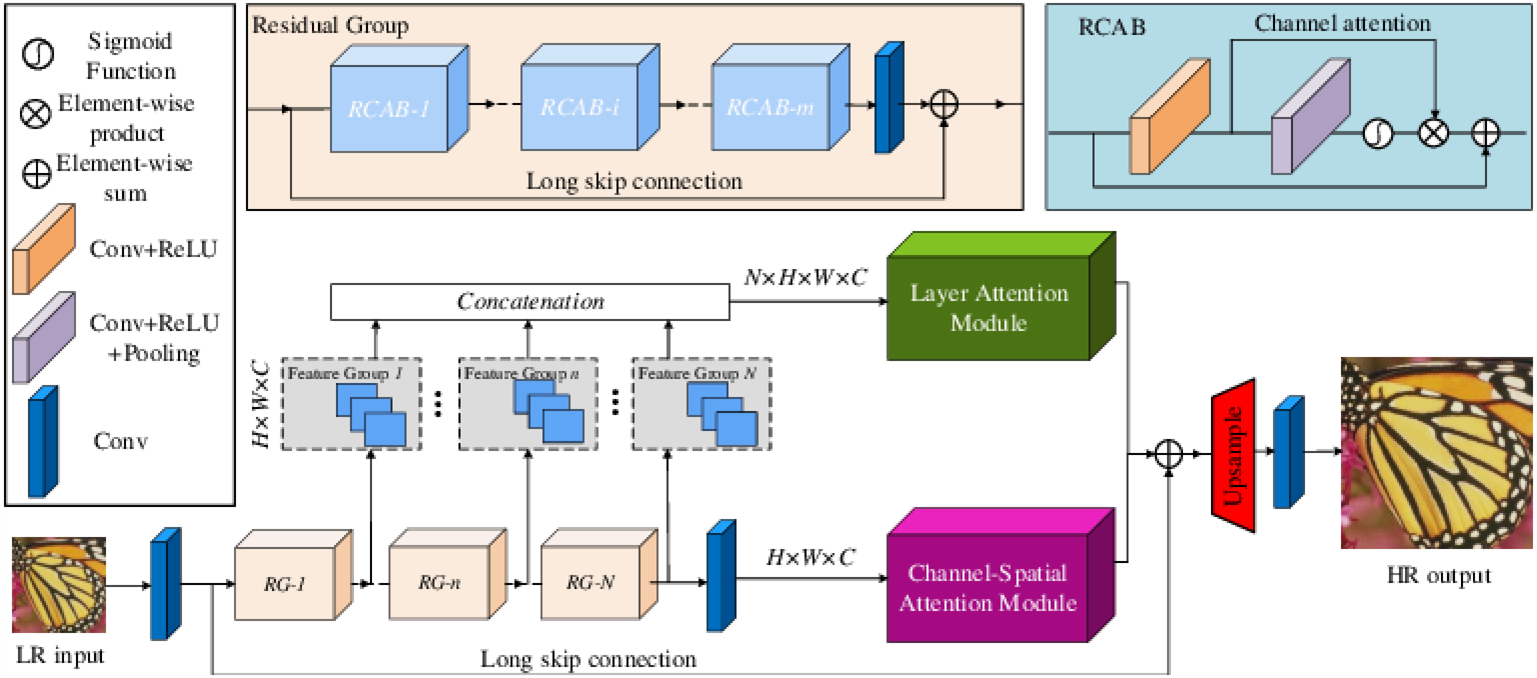
Reference: "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network", 2017 CVPR

Deep Learning for Single Image Super Resolution

- Generative Adversarial Network(GAN) for Super-Resolution
  - SRGAN, EnhanceNet, SRFeat, ESRGAN

Reference: "A Deep Journey into Super-resolution: A survey", 2019 arXiv

The SOTA so far (HANet, ECCV 2020)



- Bring the “attention” module to the generator

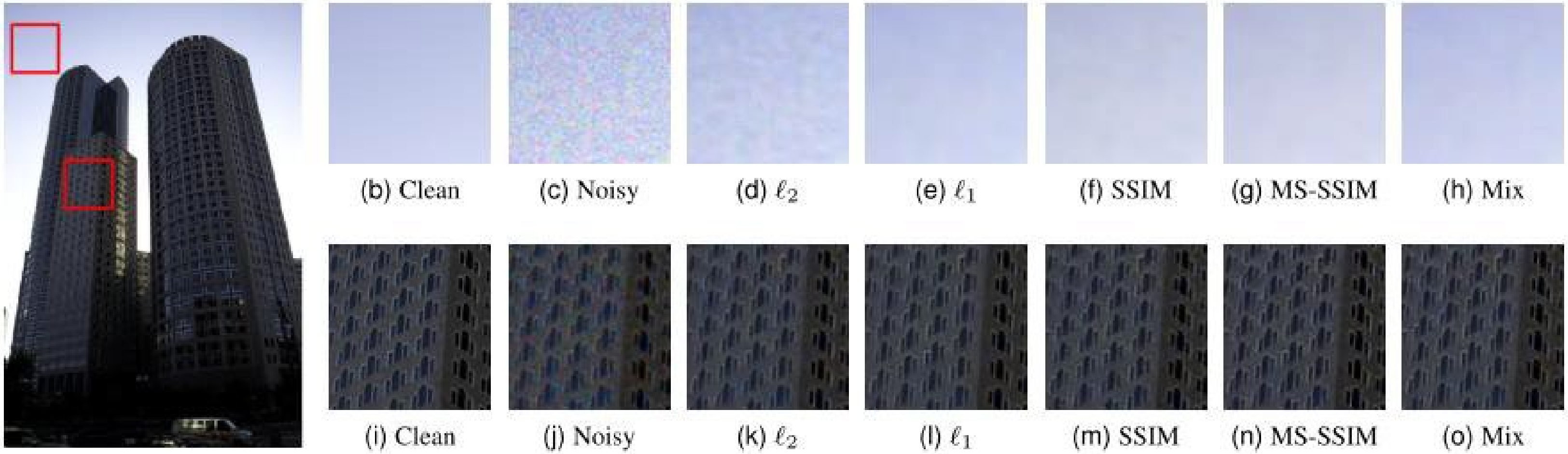
WHAT'S NEXT?

Finding the issues in current SRs

Some Issues for Super Resolution

- Loss function
  - Propose a various loss function methods in Image Restoration task
  - Report the best result when using mixed loss with MS-SSIM loss + $l1$ loss

$$\mathcal{L}^{\text{Mix}} = \alpha \cdot \mathcal{L}^{\text{MS-SSIM}} + (1-\alpha) \cdot G_{\sigma_G^M} \cdot \mathcal{L}^{\ell_1}, \qquad (14)$$



(a) Clean image (b) Clean (c) Noisy (d) ℓ_2 (e) ℓ_1 (f) SSIM (g) MS-SSIM (h) Mix (i) Clean (j) Noisy (k) ℓ_2 (l) ℓ_1 (m) SSIM (n) MS-SSIM (o) Mix

Reference: "Loss Functions for Image Restoration with Neural Networks", 2016 IEEE TCI

Some Issues for Super Resolution

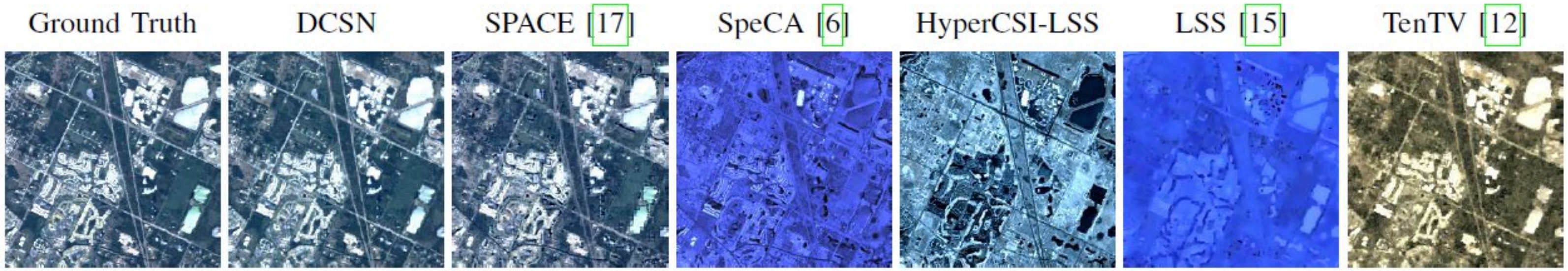
- GAN Loss achieves a high visual quality
- L1/SSIM losses achieves a high fidelity
- However, we don’t have a metric that can consider both of them

- We show that one of the critical problem in loss functions is “resolution-aware” information
  - Feature distance does not fit “resolution”
    - Good quality != High resolution
      - E.g., defocused sample/background?

Some Issues for Super Resolution

- How about the structured image super-resolution?
  - Face hallucination

- How about multilinear super-resolution
  - E.g. Hyperspectral data



Hsu, Chih-Chung, et al. "Sigan: Siamese generative adversarial network for identity-preserving face hallucination." *IEEE Transactions on Image Processing* 28.12 (2019): 6225-6236.

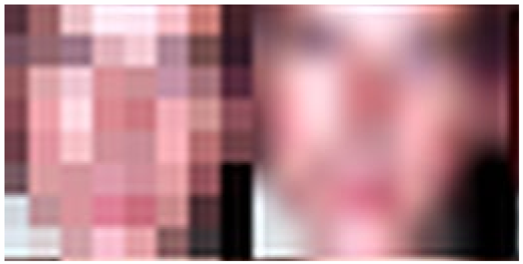
Outline

- Overview of Deep Learning
  - Supervised – Unsupervised
- Deep super-resolution
  - Traditional super-resolution
  - Structured image super-resolution
    - Face hallucination
  - 2-D image super-resolution (generic images)
  - *N*-D image super-resolution (Hyperspectral images)
- Summary

IDENTITY-PRESERVING FACE HALLUCINATION

ICIP 18, IEEE Transactions on Image Processing (TIP), Dec. 2019.

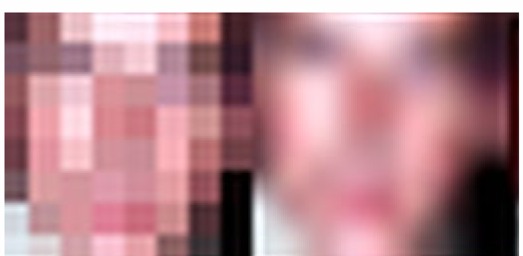
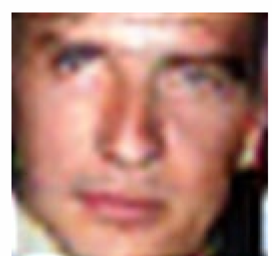
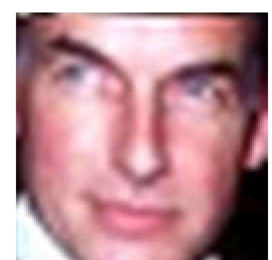
Traditional Face Hallucination



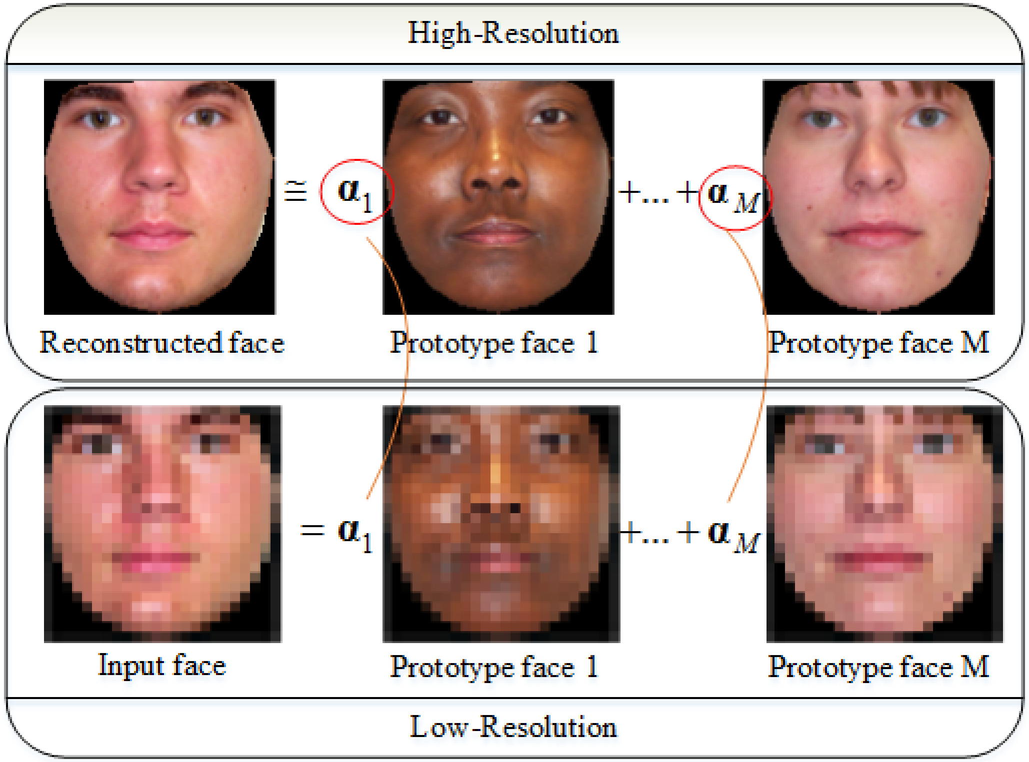
LR Bicubic SR HR

Amazing but identity unrecognizable!

We achieve

Face Hallucination



$$\mathbf{I} \cong \boxed{\mathbf{P}} \cdot \boldsymbol{\alpha} = \mathbf{R}$$

$$\boldsymbol{\alpha}^* = ((\mathbf{P}_L)^T \cdot \mathbf{P}_L)^{-1} \cdot (\mathbf{P}_L)^T \cdot \mathbf{I}_L$$

Dictionary

Learning to Hallucinating Face

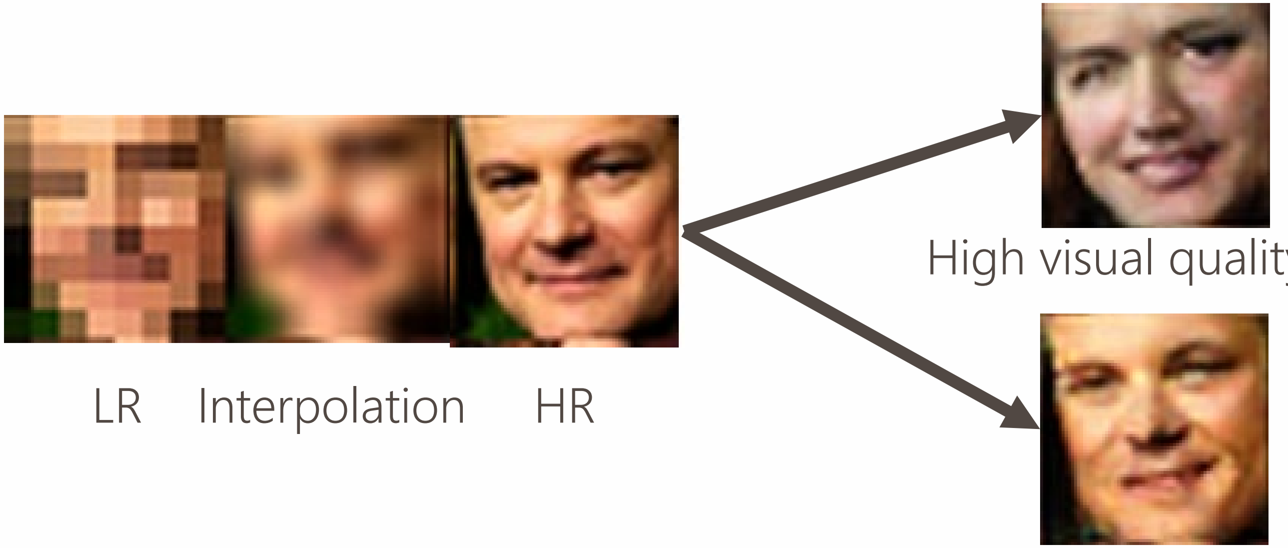
- Traditional approach
  - Dictionary learning by PCA, NMF, ONMF,…etc
- Deep learning-based approach
  - End-to-end architecture
    - Input low-resolution face image, out high-resolution face image directly.
- Deep neural network has different structures
  - CNN-based (Convolutional neural network)
    - Upsampling layer upscales input signal
  - GAN-based (Generative adversarial network)
    - High quality result
    - May result in identity-unrecognizable

GAN-based Face Hallucination

- Pros:
  - High visual quality of the reconstructed image
- Cons:
  - May be identity-unrecognizable

Our Goal

- High visual quality reconstruction
  - Even in extreme low-resolution inputs
- Identity-recognizable reconstruction
  - As similar to the ground truth as possible



LR Interpolation HR

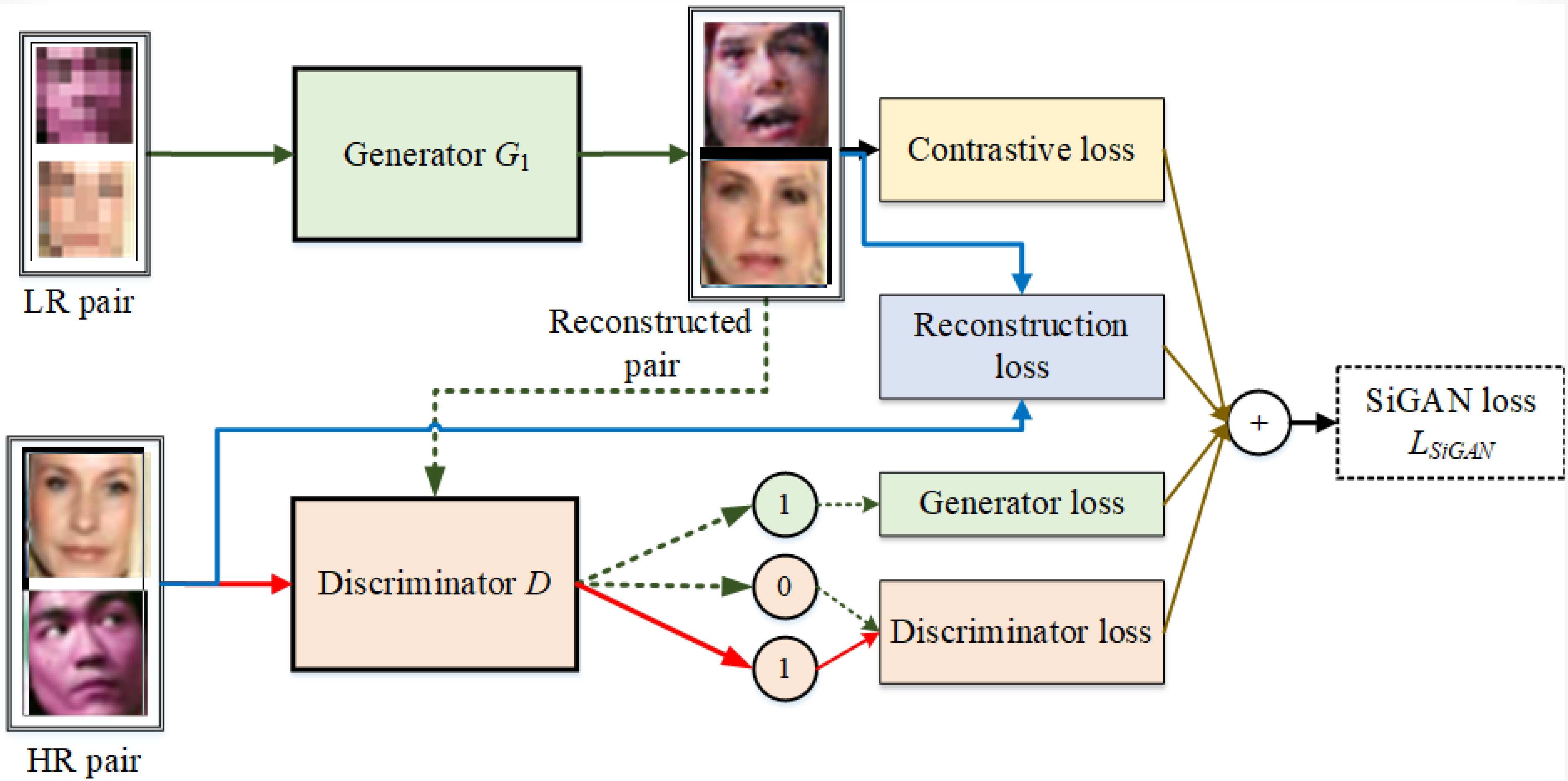
High visual quality only

Identity-recognizable & high visual quality

Our Solution

- Key idea
  - Label embedding
    - Use the label information to fine-tune the generator
    - Identity-recognizable reconstruction
  - We propose “Siamese GAN” (SiGAN)
    - Label information will guide the “generator” how to obtain both high-visual quality and identity-recognizable result
    - Partial label information needs only

The Proposed SiGAN

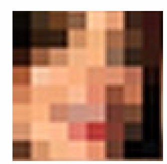
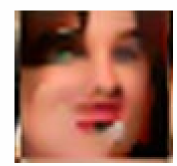


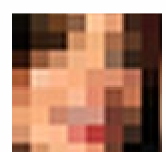
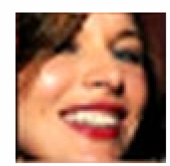
The Loss Function of The Proposed SiGAN

- Loss function for our generator

$$\min_{G}\max_{D} V(D,G) = E_D\left[\log D(\mathbf{x}_1^{HR})\right]$$
$$+E_G\left[\log\left(1 - D(G(\mathbf{x}_1^{LR}))\right)\right] + E_C\left[G(\mathbf{x}_1^{LR}), G(\mathbf{x}_2^{LR})\right],$$

  - subject to $\|y^{HR} - y^{SR}\|_1 < \epsilon$
  - SR result: G($\mathbf{x}^{LR}$)
  - E_C represents contrastive loss

D [G() = ] = 0

D [G() = ] = 1

Contrastive Loss for SiGAN

- If we directly minimize Ew(X1, X2)
  - The energy and the loss can be made zero by simply making Gw(X1) a constant function
  - We don't want to see that
- By adding a contrastive term
  - The loss function can be

CNN's parameters

The same or not (0/1)

Partial loss function for a genuine pair

$$L(\mathrm{W}) = \sum_{i=1}^{P} L(\boxed{W}, \boxed{(Y,} \boldsymbol{x}_1, \boldsymbol{x}_2)^i)$$

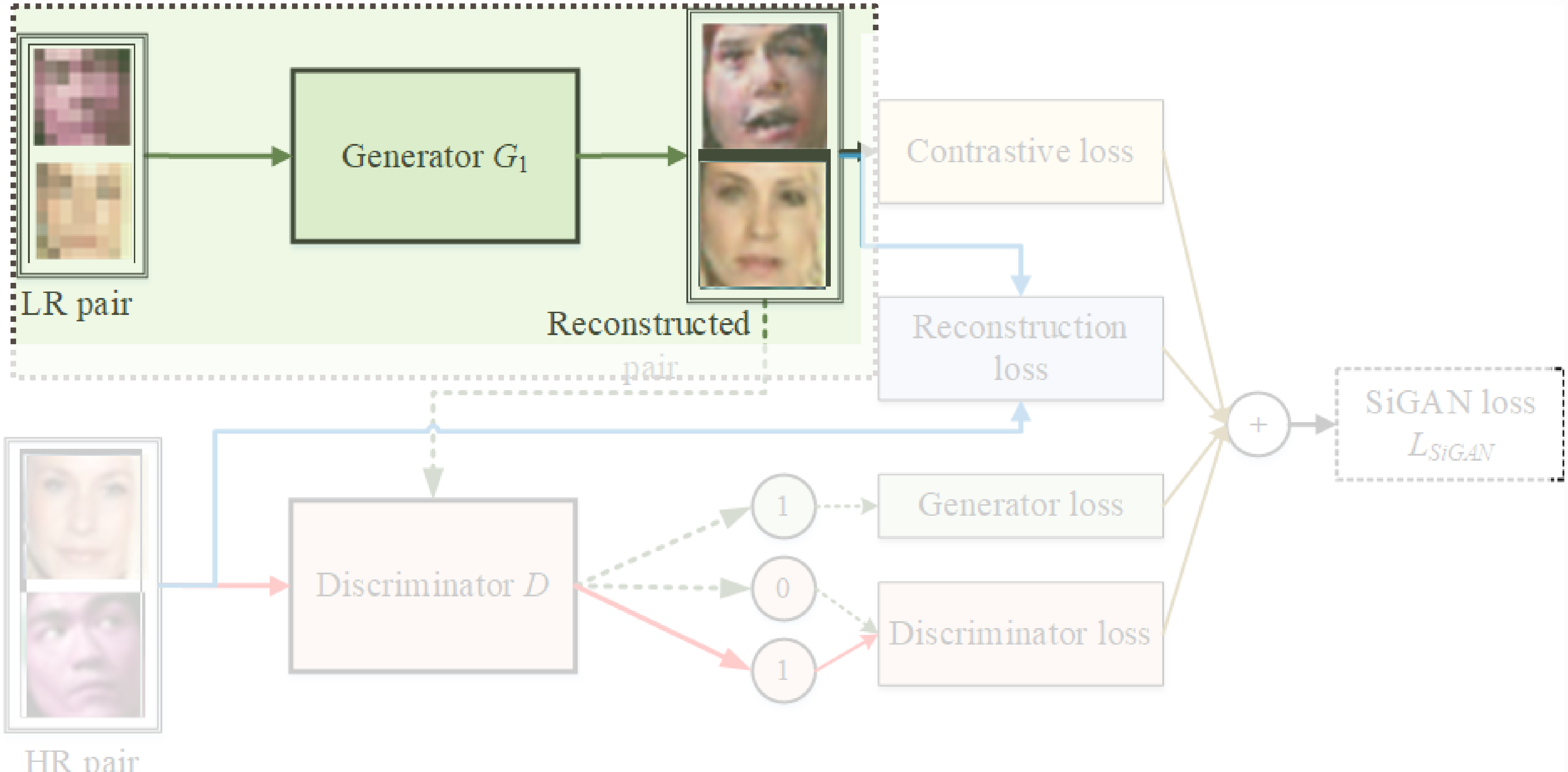
$$L(W, (Y, \boldsymbol{x}_1, \boldsymbol{x}_2)^i) = y \boxed{L_G} \left(E_W(\boldsymbol{x}_1, \boldsymbol{x}_2)\right) + (1 - y) \boxed{L_I} \left(E_W(\boldsymbol{x}_1, \boldsymbol{x}_2)\right)$$

$$L_{\mathrm{G}} = \frac{1}{2}(E_w)^2$$

$$L_{\mathrm{I}} = \frac{1}{2}[\max(0, margin - E_w)]^2$$

Partial loss function for an impostor pair

Test Stage of The Proposed SiGAN



A simple forward process

Experiment Settings

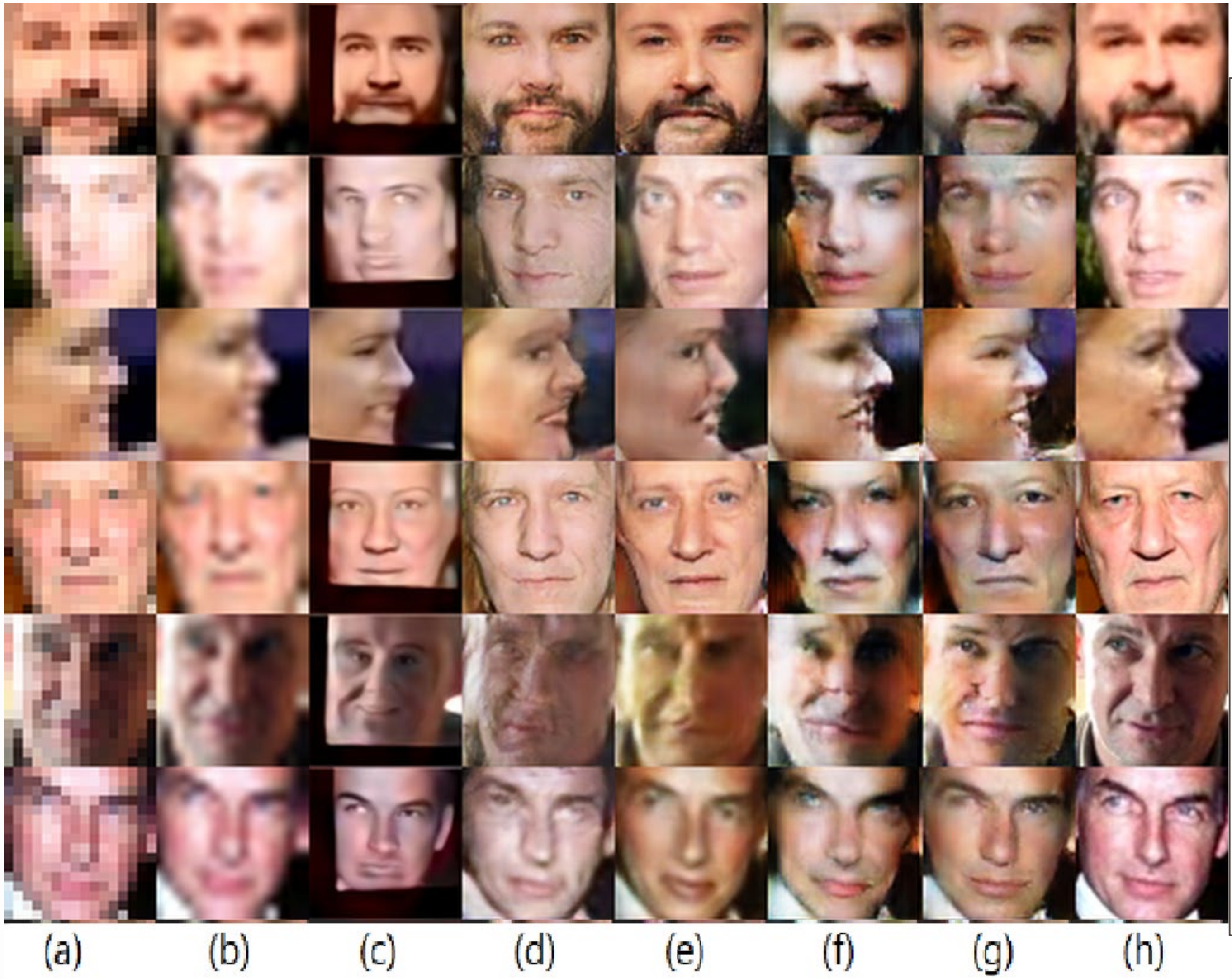
- LR: 8x8
- HR: 32x32 (4x upscaling factor)
- #Identities of training set: 10,575
- #Training images: 491,131
- #Test images: 3,283
- Face recognition engine: FACENET (State-of-the-art)

Subjective Result (8x8→32x32)

- Face hallucination: Identity-recognizable reconstruction

(a) LR face image
(b) Bicubic interpolation
(c) ECCVCNN
(d) GAN
(e) UR-DGN
(f) Pixel-Recurrent (Google)
(g) Ours w/o label
(h) Ours
(i) Original HR face image

Subjective Result (16x16→64x64)



(a) LR face image
(b) Bicubic interpolation
(c) ECCVCNN
(d) GAN
(e) UR-DGN
(f) Pixel-Recurrent (Google)
(g) Ours
(h) Original HR face image

Objective Results

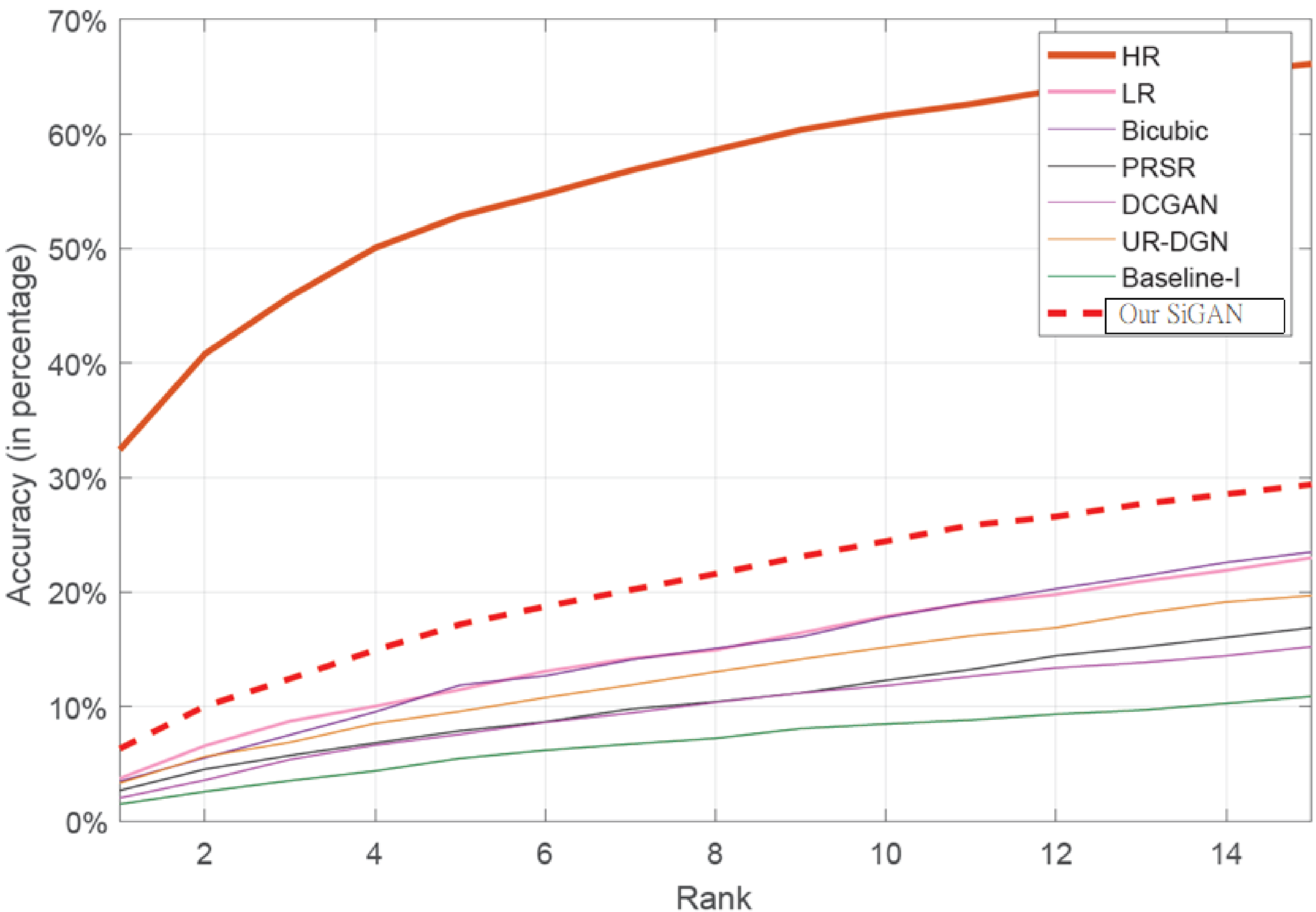
Method	Top-1	Top-5	Top-10
HR (32×32)	30.4%	51.2%	59.6%
LR (8×8)	10.7%	19.5%	33.1%
Bicubic	10.8%	20.1%	34.4%
DFCG [11]	9.3%	17.7%	21.4%
UR-DGN [9]	9.9%	18.6%	22.7%
DCGAN [22]	4.6%	10.9%	16.8%
PRSR [25]	10.8%	18.8%	24.4%
SR-GAN [15]	8.8%	11.1%	19.4%
Wavelet-SRNet [17]	12.8%	20.2%	30.3%
SiGAN (ResNet)	15.8%	27.5%	40.4%
SiGAN (DenseNet)	15.1%	26.8%	40.3%

Face recognition rate comparison
LR=8x8
HR=32x32

Method	Top-1	Top-5	Top-10
HR (64×64)	36.8%	55.9%	63.8%
LR (16×16)	12.4%	27.4%	37.1%
Bicubic	11.6%	27.5%	37.6%
DFCG [11]	9.6%	23.7%	34.8%
UR-DGN [9]	12.2%	29.0%	38.7%
DCGAN [22]	9.3%	24.9%	33.9%
PRSR [25]	13.3%	29.7%	40.1%
SR-GAN [15]	11.6%	23.2%	36.3%
Wavelet-SRNet [17]	12.0%	25.5%	38.8%
SiGAN (ResNet)	17.9%	32.9%	48.1%
SiGAN (DenseNet)	18.3%	33.5%	50.0%

Face recognition rate comparison
LR=16x16
HR=64x64

Objective Result (8x8)



Summary of Our SiGAN

- Contributions
  - Label information is embedded in the generator of GAN
    - A Guider for the generator
  - High visual quality and identity-recognizable reconstruction
  - Faster hallucination process

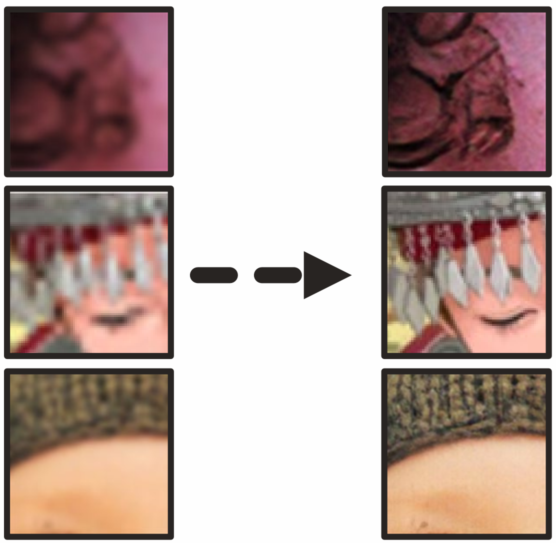
Outline

- Overview of Deep Learning
  - Supervised – Unsupervised
- Deep super-resolution
  - Traditional super-resolution
  - Structured image super-resolution
    - Face hallucination
  - 2-D image super-resolution (generic images)
  - *N*-D image super-resolution (Hyperspectral images)
- Summary

RESOLUTION-AWARE ADVERSARIAL LEARNING

IEEE SAM 2020, Oral

GAN based Super Resolution



Reconstruction Error

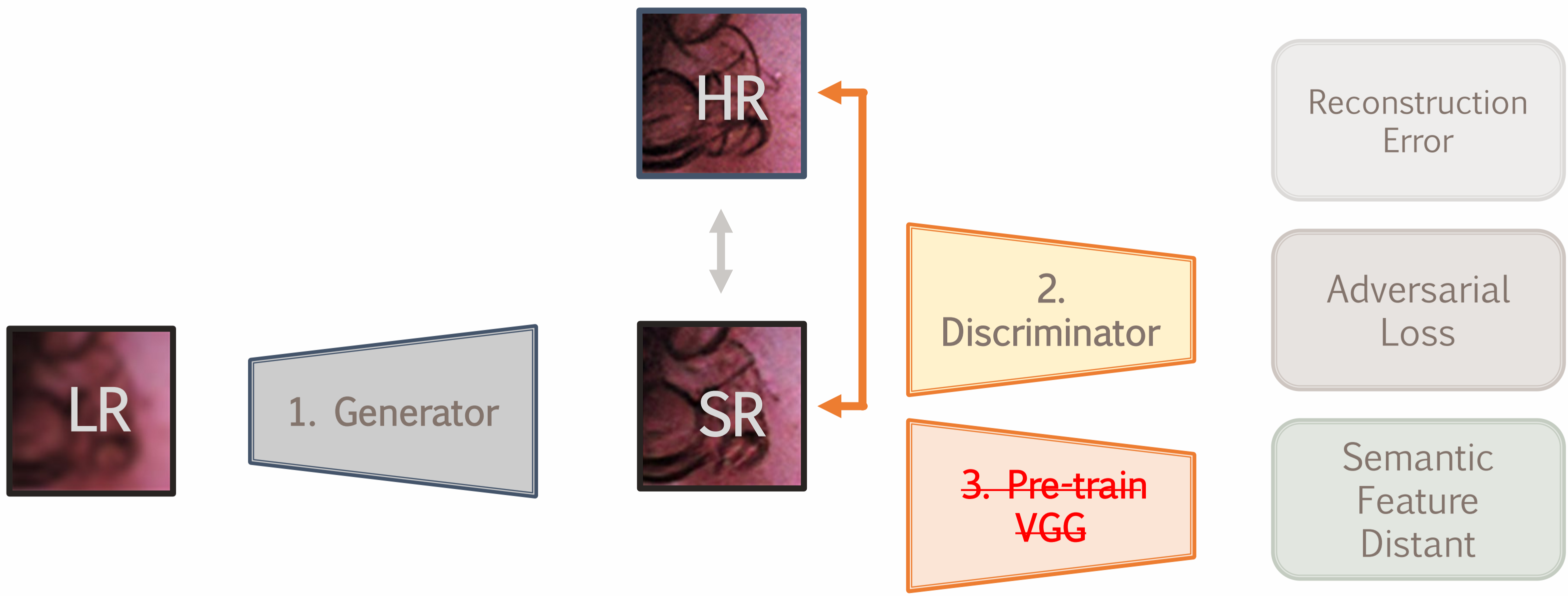
Adversarial Loss

Semantic Feature Distant

fidelity

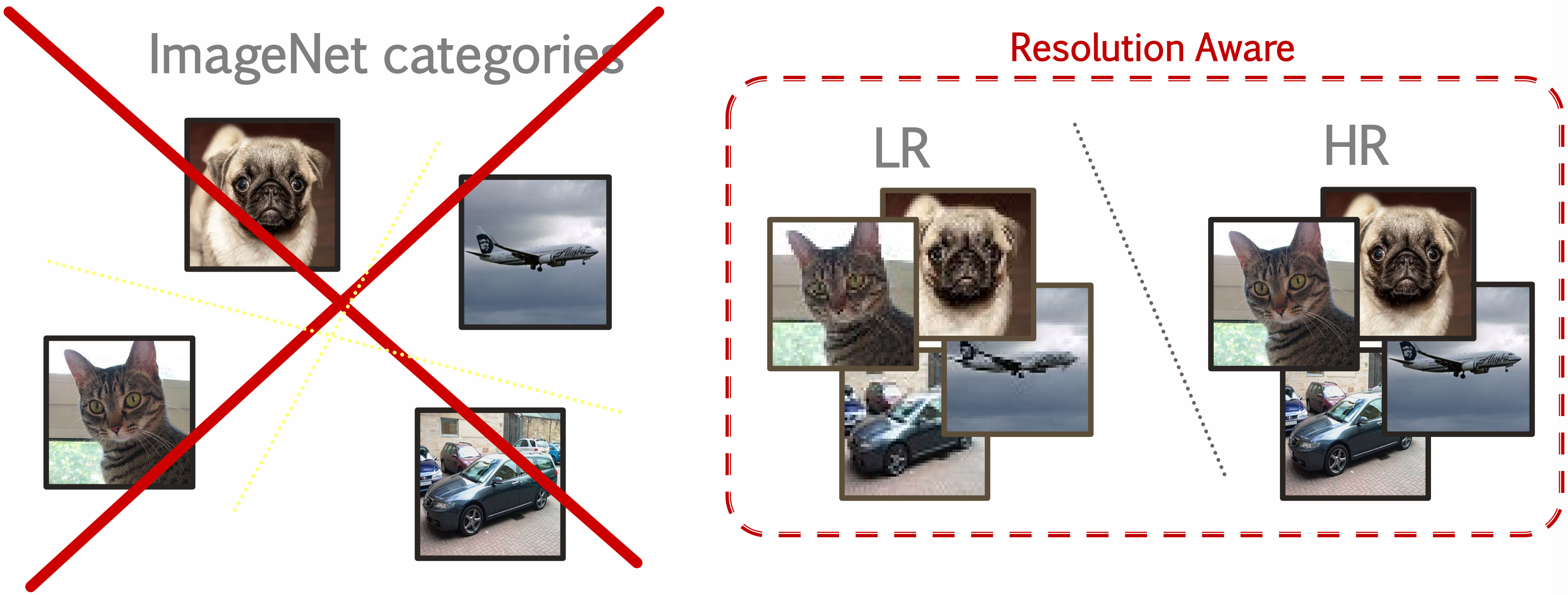
visual quality

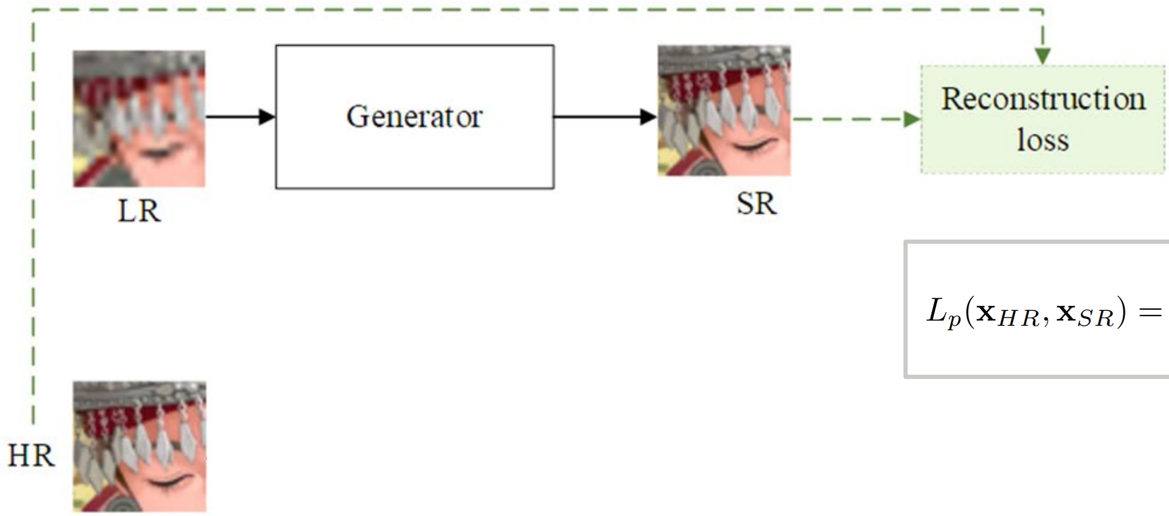
GAN based Super Resolution



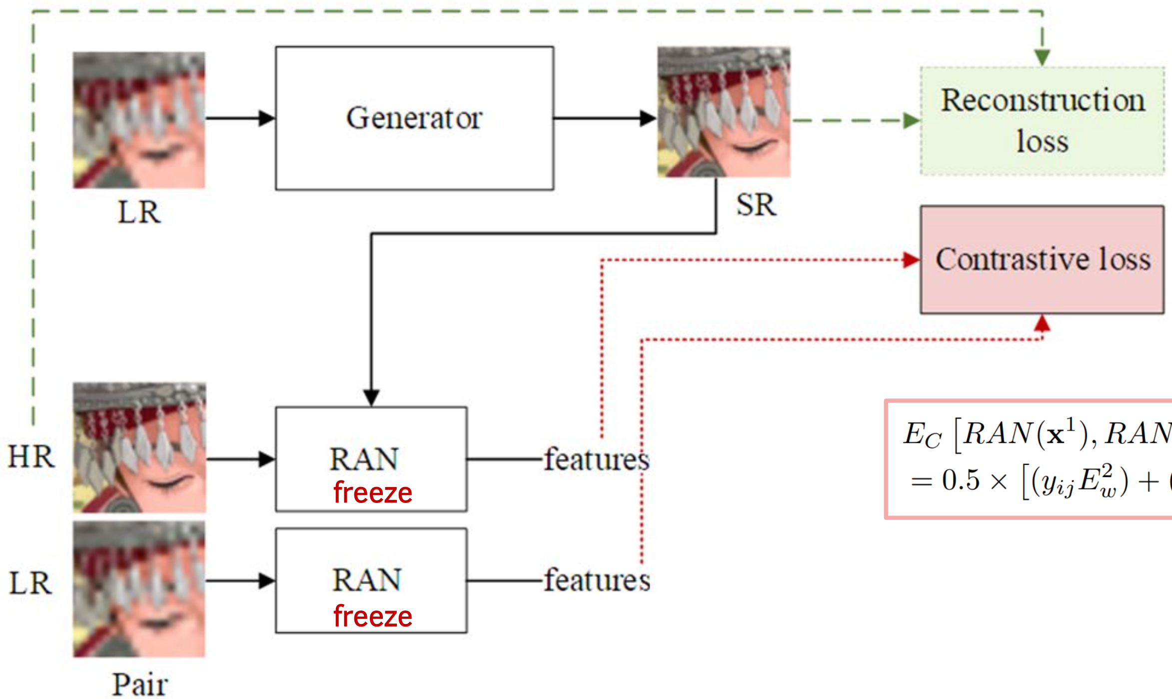
Not for measuring the features of the HR and LR

Resolution Aware feature Network (RAN)

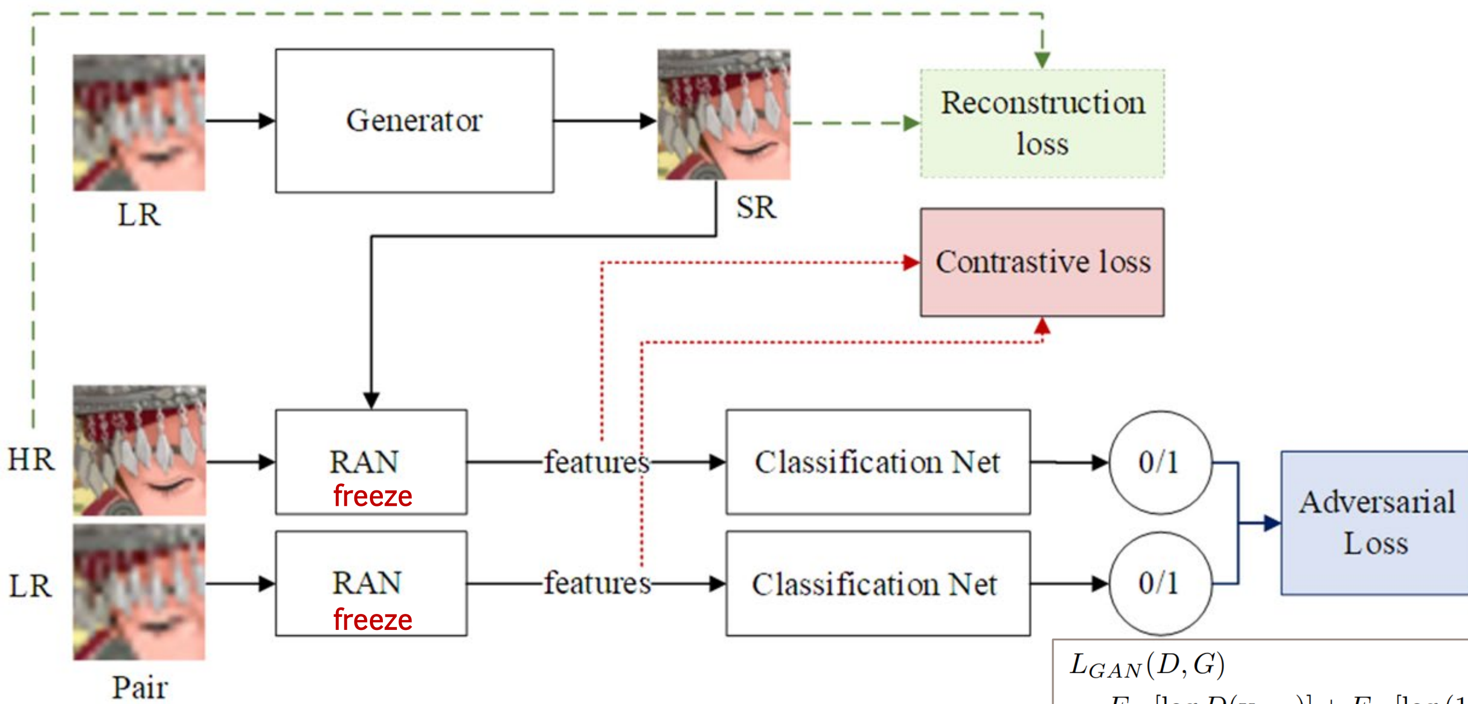




$$L_p(\mathbf{x}_{HR}, \mathbf{x}_{SR}) = \frac{1}{N^2 - 1}\sum_{i=1}^{N^2} |\mathbf{x}_{SR} - \mathbf{x}_{HR}|$$

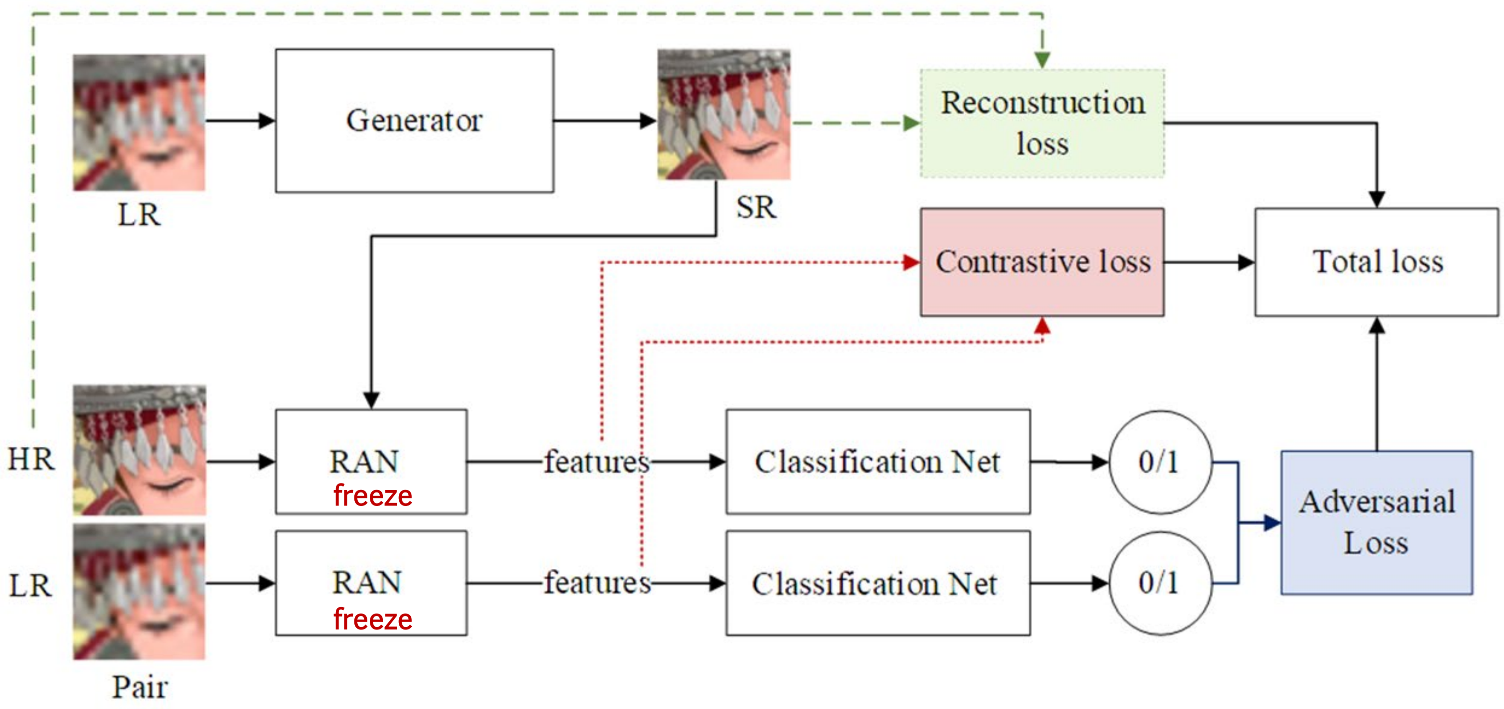


$$
\begin{aligned}
&E_C\left[RAN(\mathbf{x}^1), RAN(\mathbf{x}^2)\right] \\
&= 0.5 \times \left[(y_{ij}E_w^2) + (1 - y_{ij}) \times max(0, (m - E_w)_2^2)\right]
\end{aligned}
$$



$$L_{GAN}(D,G) = E_D\left[\log D(\mathbf{x}_{HR})\right] + E_G\left[\log\left(1 - D(G(\mathbf{x}_{LR}))\right)\right]$$

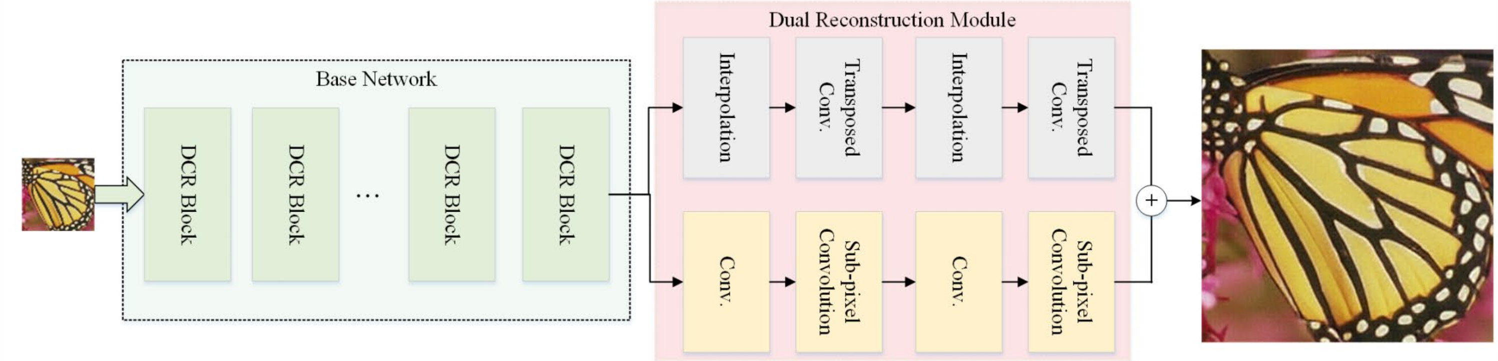
Couple Adversarial Training (CAT)



Network Structure

- RAN / Discriminator (VGG16)

- Generator (DRSR)
  - Hswish -> Swish



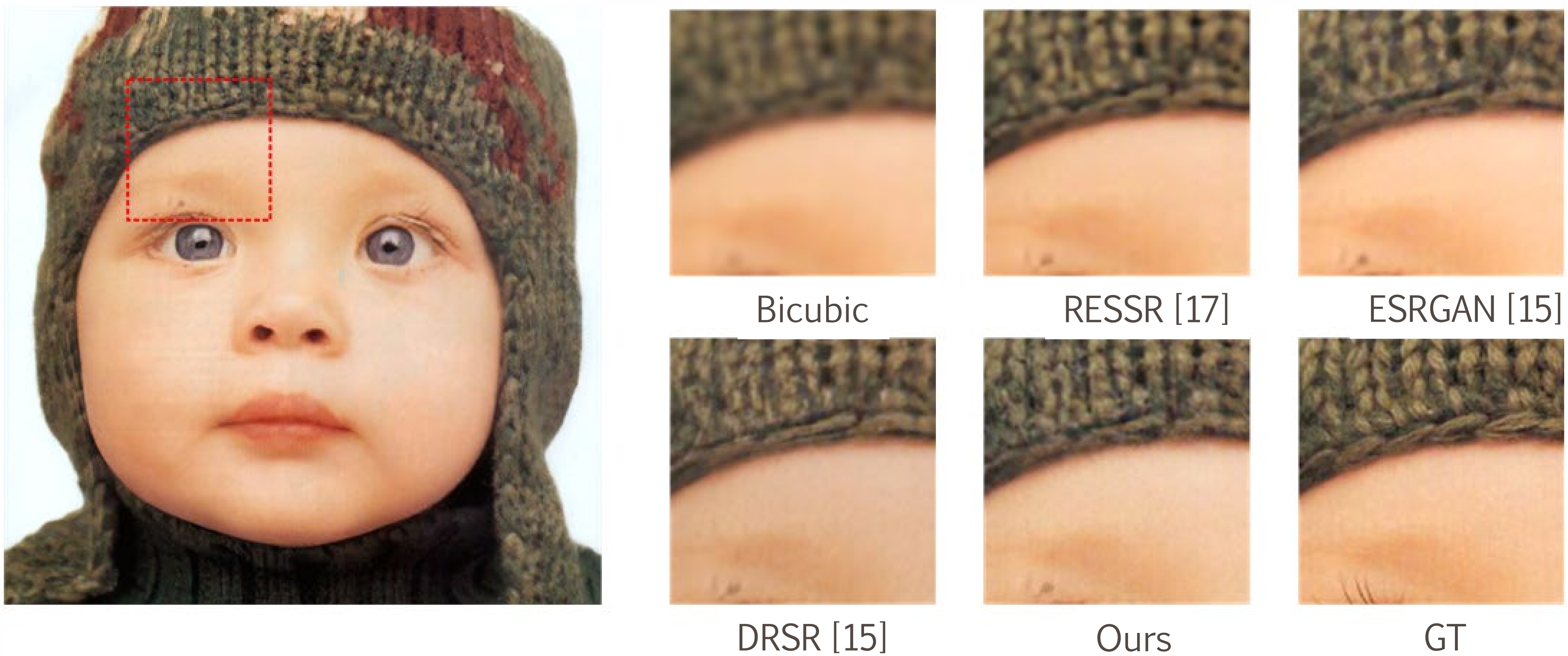
RESULTS

Objective Quality Comparison

TABLE I
PERFORMANCE COMPARISON AMONG THE DIFFERENT SR METHODS EVALUATED ON SET5 [9], BSD100, [11] AND URBAN100 [9].

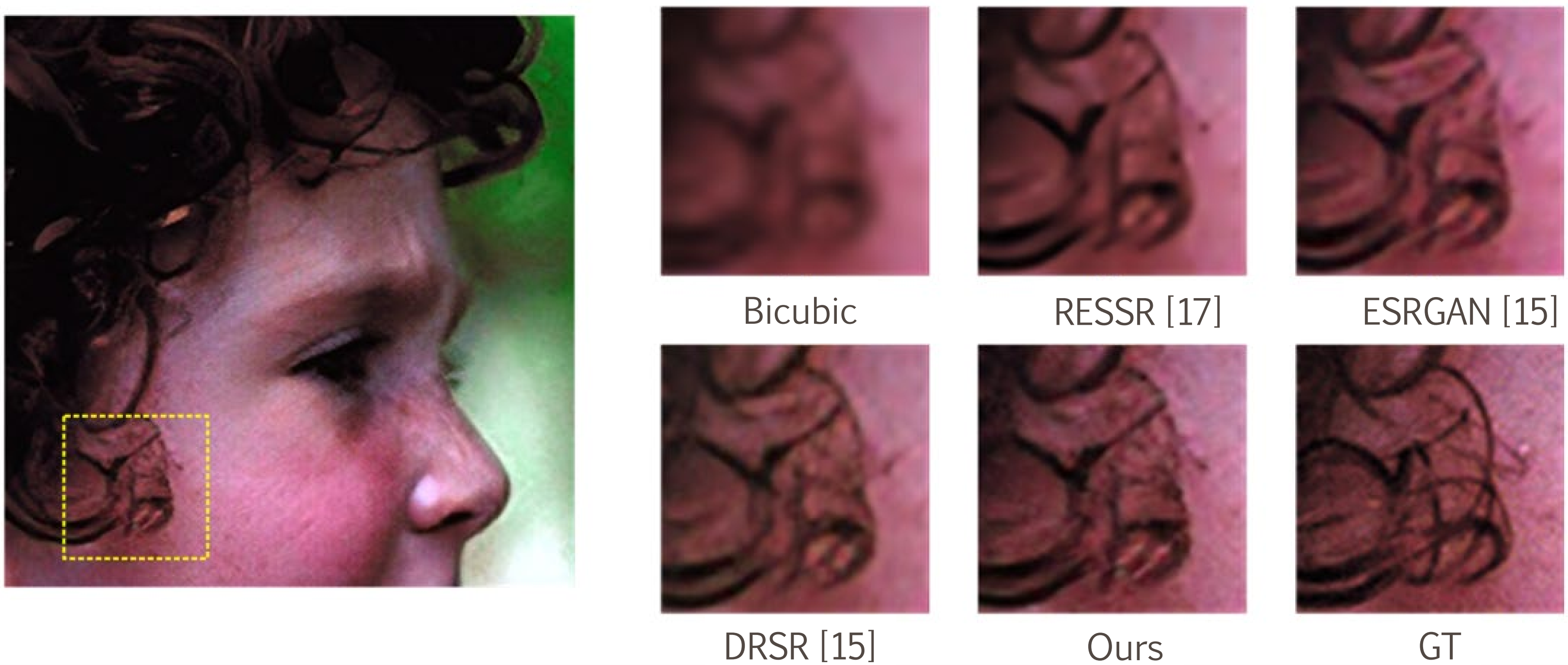
Method	Set5		BSD100		Urban100	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
MSRResNet [15]	**30.28**	**0.864**	26.27	0.712	24.62	0.766
ESRGAN [15]	29.06	0.814	25.57	0.682	24.15	0.712
DRSR [6]	29.18	0.823	25.86	0.705	24.22	0.726
RESSR [17]	30.11	0.860	26.22	0.709	24.65	0.766
Baseline (ours)	29.25	0.858	**27.76**	**0.779**	**24.99**	**0.802**
Proposed	29.66	0.848	26.51	0.723	24.54	0.759

Subjective Quality Comparison



Subjective Quality Comparison

Subjective Quality Comparison



Conclusion

- Resolution Aware feature Network (RAN)
  - Get the resolution-aware information to the deep neural network

- Combined contrastive loss to learn the discriminative features to "Resolution"

- Excellent both visual and objective quality of the reconstructed images

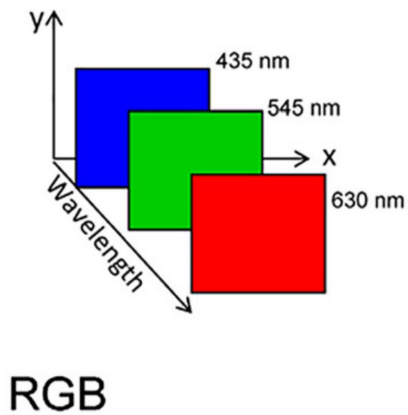
Outline

- Overview of Deep Learning
  - Supervised – Unsupervised
- Deep super-resolution
  - Traditional super-resolution
  - Structured image super-resolution
    - Face hallucination
  - 2-D image super-resolution (generic images)
  - *N*-D image super-resolution (Hyperspectral images)
- Summary

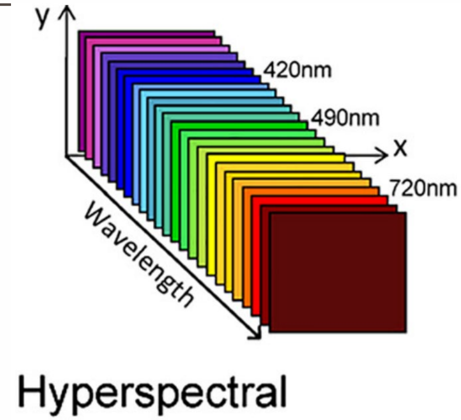
HYPERSPECTRAL IMAGE SR + COMPRESSION

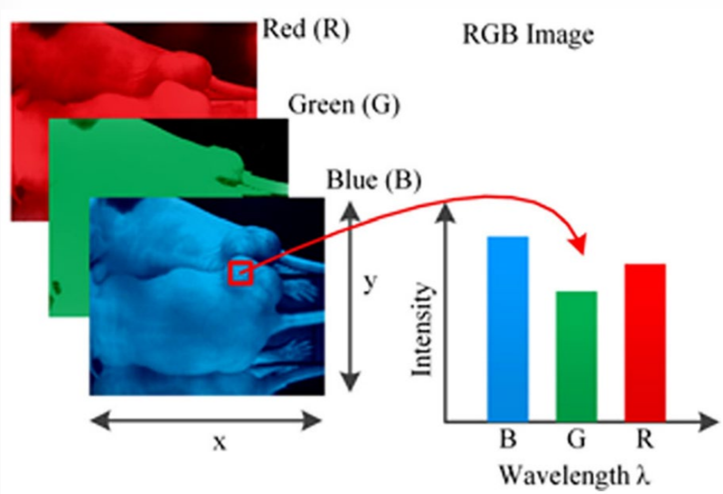
IEEE Transactions on Geoscience and Remote Sensing (TGRS), 2021

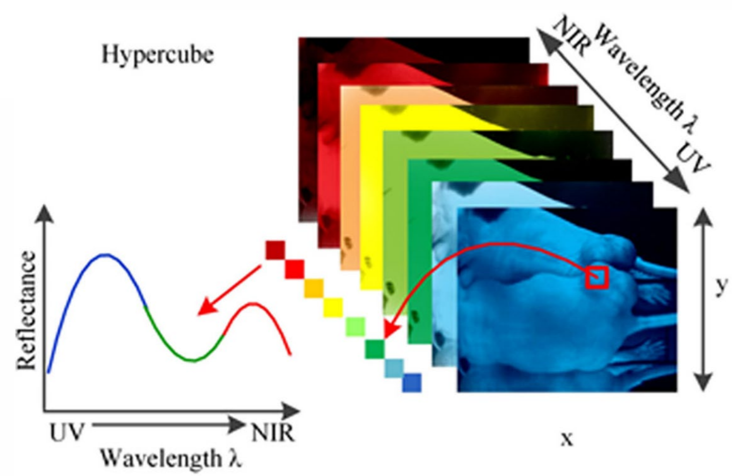
Hyperspectral Image (HSI)



Usually used in satellite





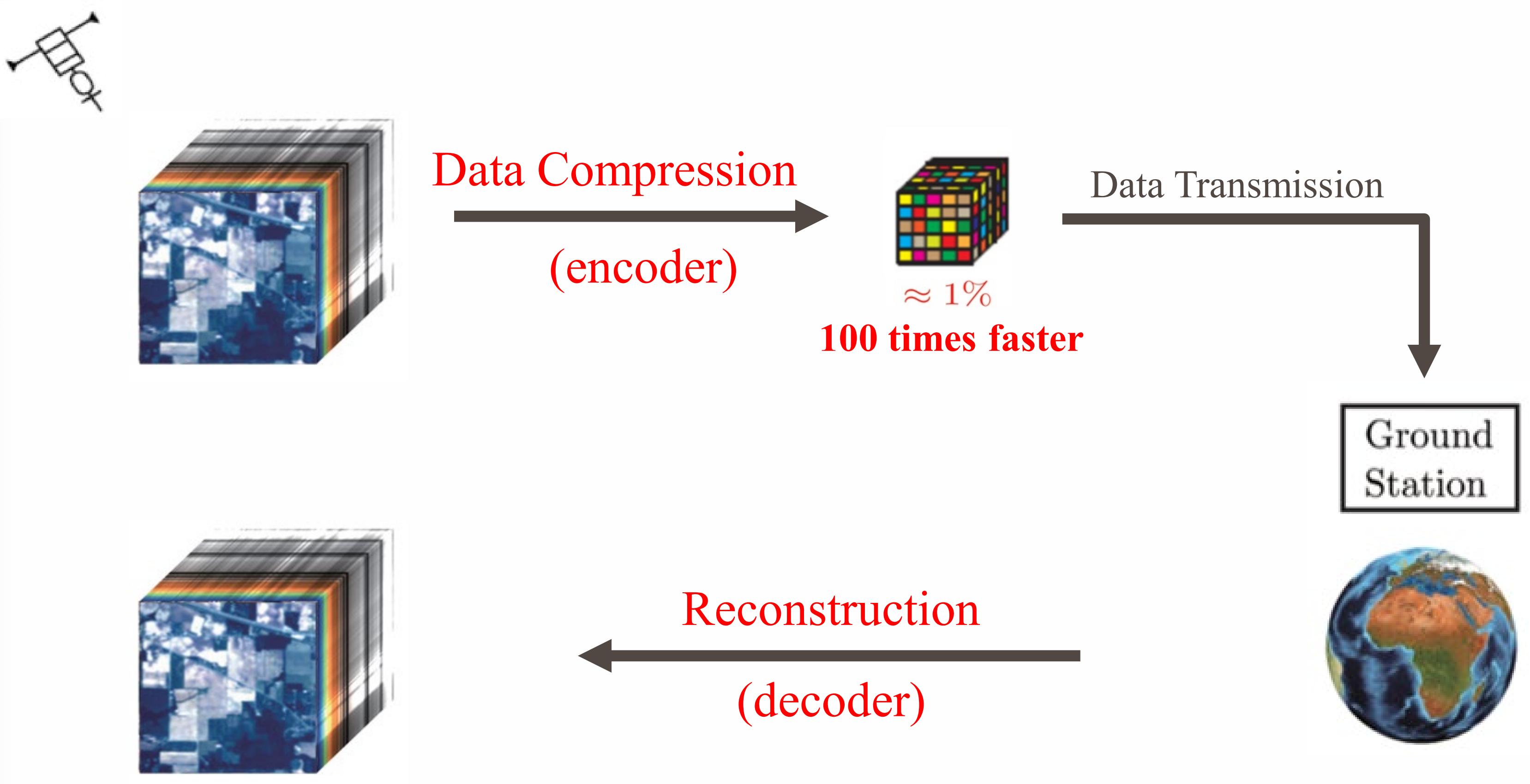


[Metha'18] N. Mehta et al., "Single-Cell Analysis Using Hyperspectral Imaging Modalities," ASME Journal of Biomechanical Engineering, vol.140, Feb, 2018

What issues in HSI

- Storage requirement:
  - Hyperspectral data contains abundant spectral information but also need more storage device
- Data throughput:
  - Transmit whole hyperspectral data is redundant, our lightweight encoder achieve low sampling rate (1%)
- We provide
  - Compress HIS (efficient transmission) first + super-resolution (recover signal) in ground station.
  - Our SR (Super Resolution)-aware decoder reconstructs the hyperspectral data well only with 1% information as input

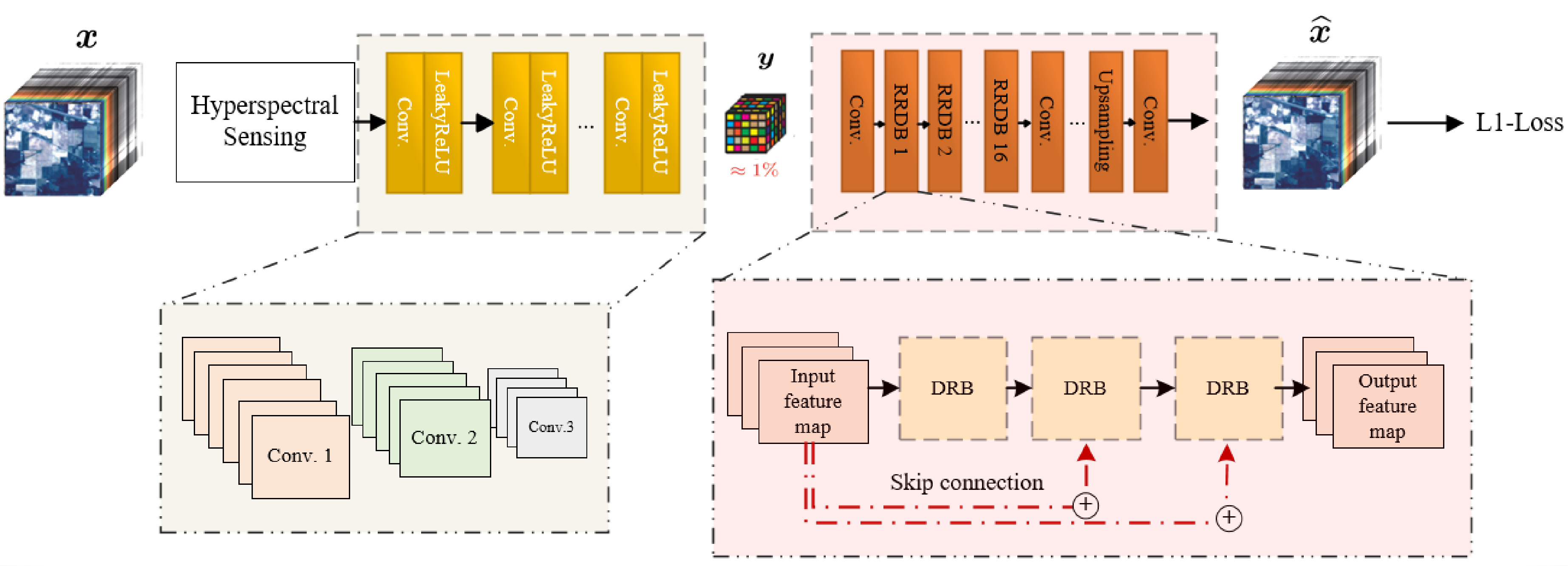
Introduction



Proposed HCSN

Hyperspectral **C**ompression **S**uper-resolution **N**etwork

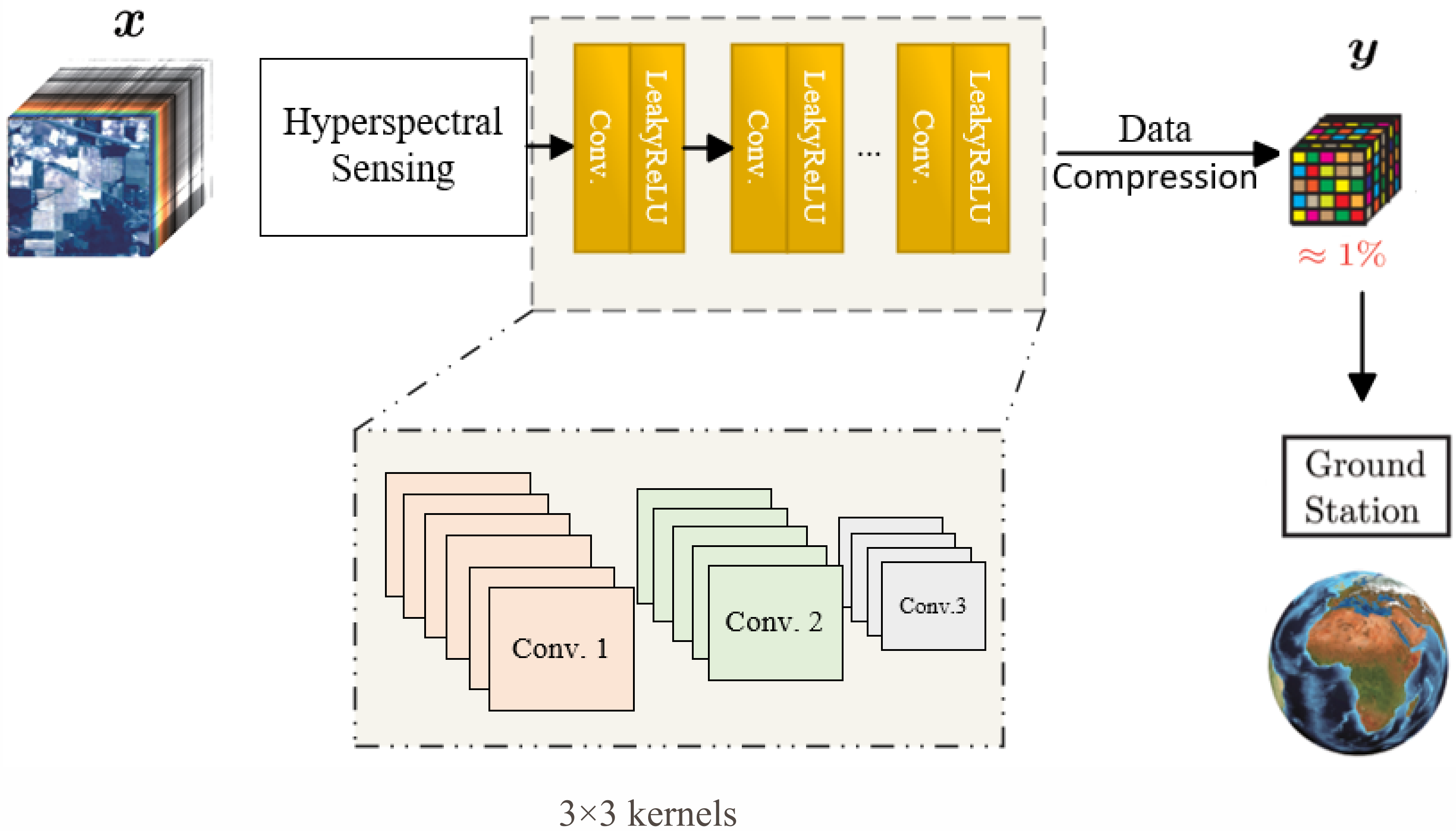
Consider "spectral" and "spatial" info



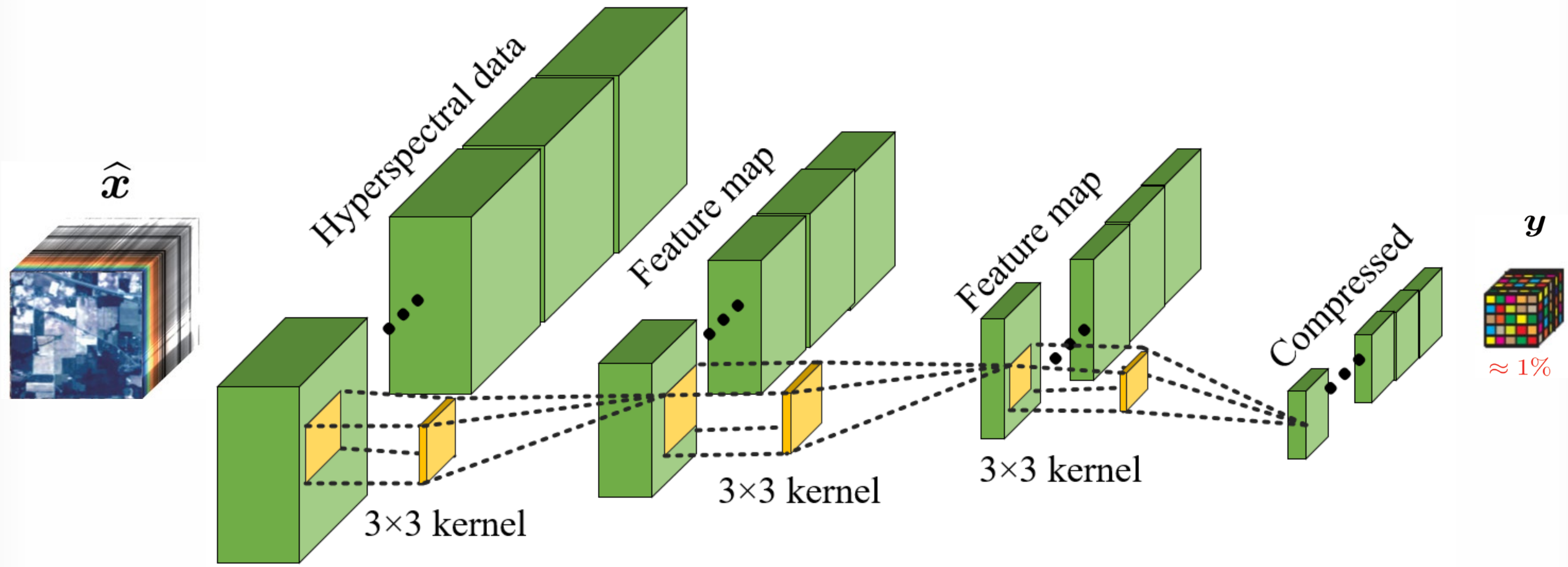
lightweight encoder

SR-aware decoder

Lightweight Encoder

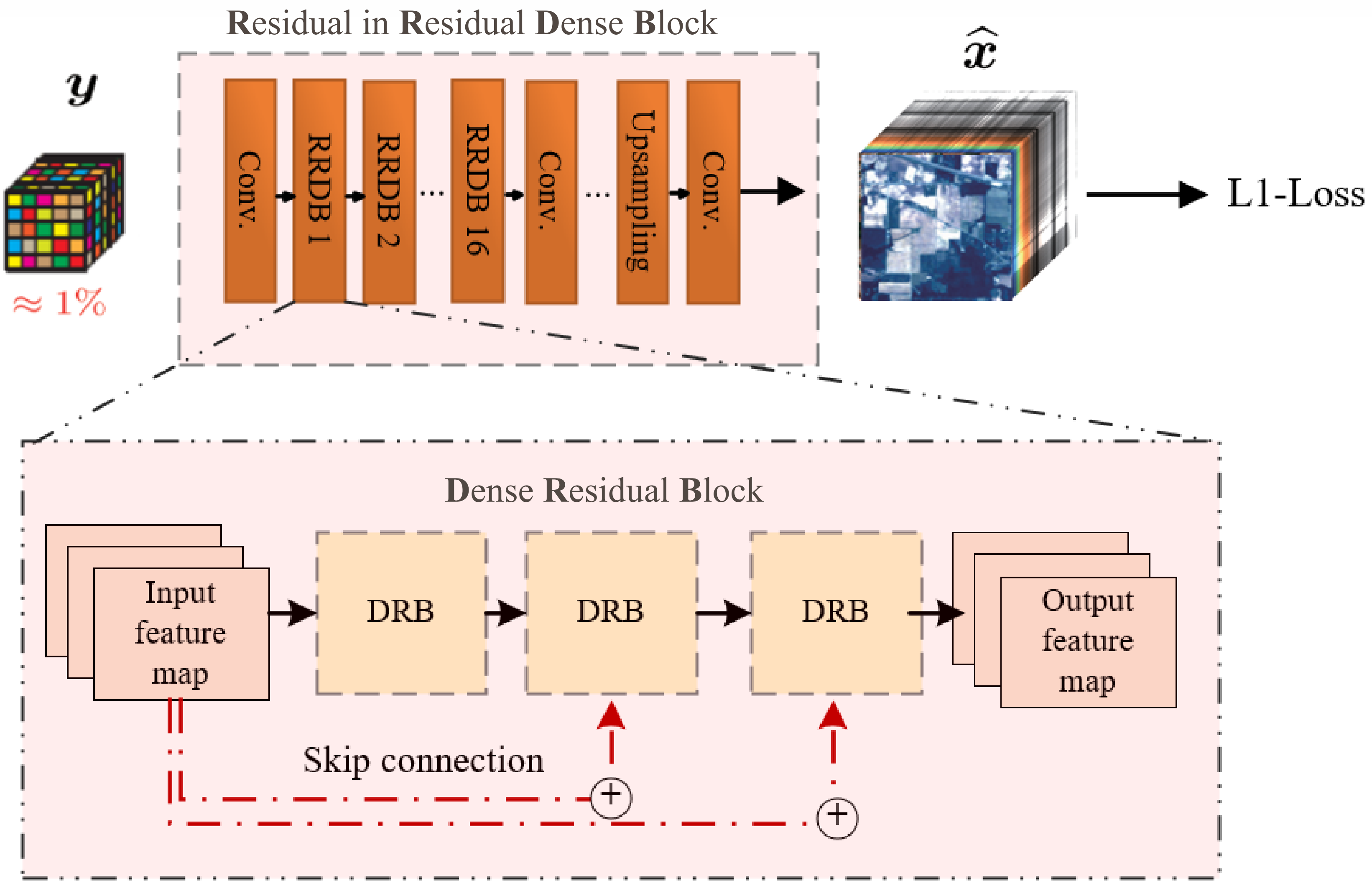


Lightweight Encoder

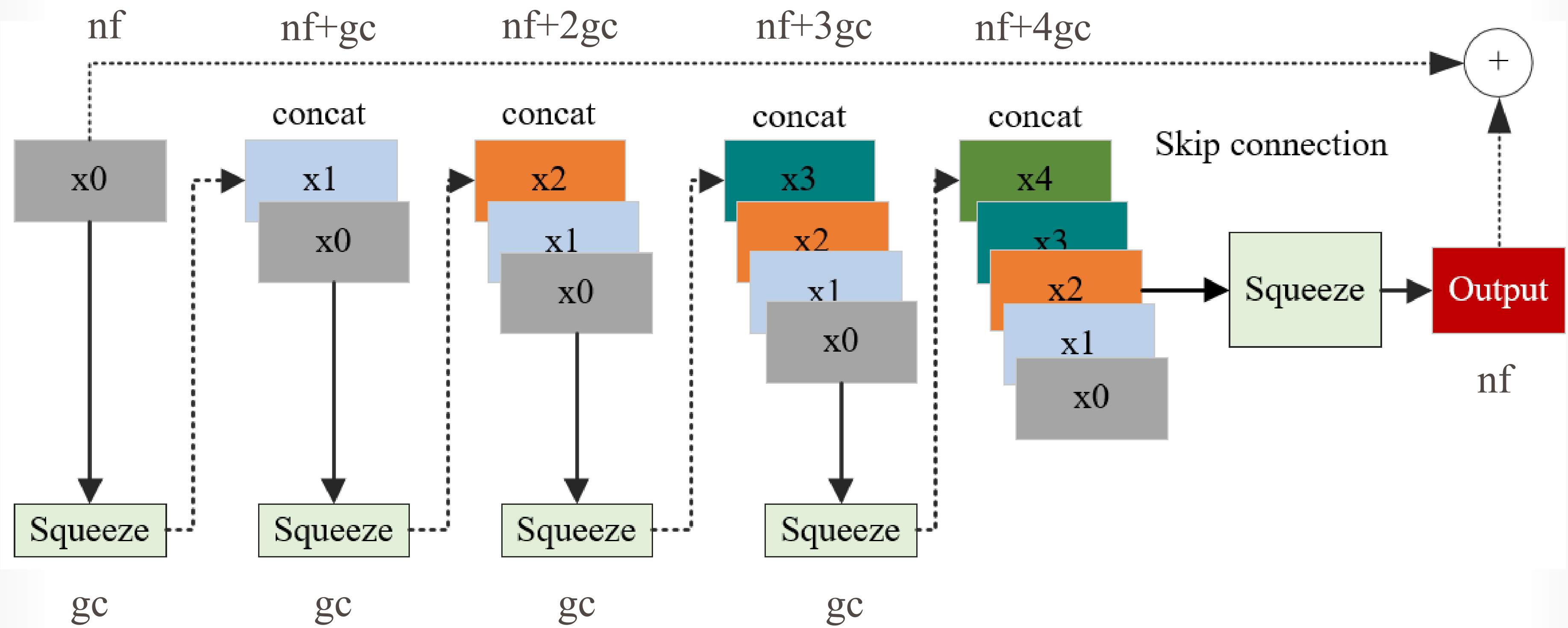


Only use three 3×3 kernel conv layers

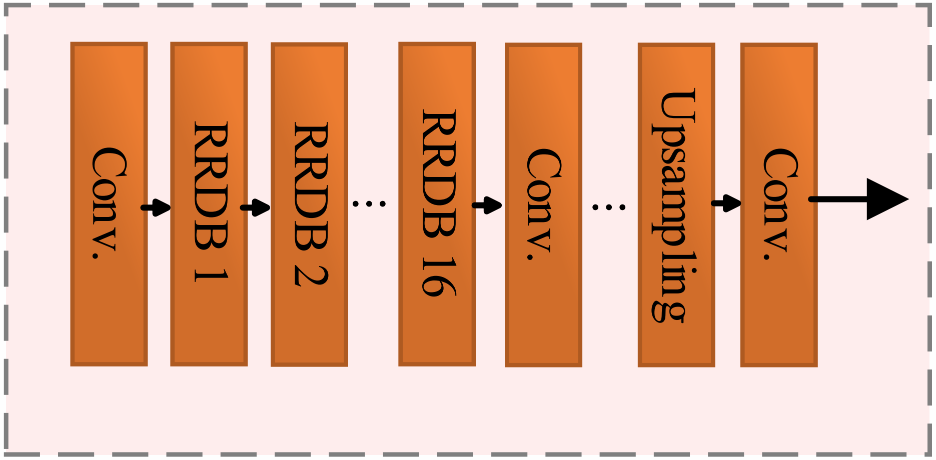
SR-aware Decoder

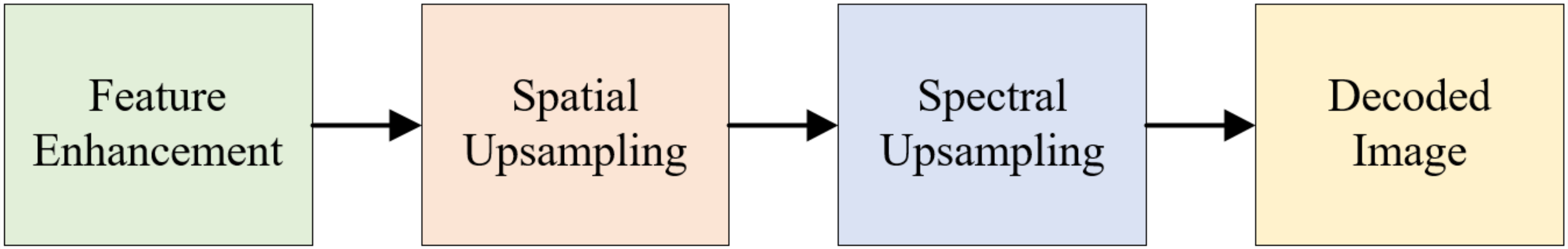


Dense Residual Block (DRB)



SR-aware Decoder





Experiment

- We train the proposed HCSN with 2,537 sub-image sized of 256×256×172
- 2,537 sub-images acquired by AVIRIS sensor:
- - 102 images for city areas (C-type)
- - 1,870 images for mountain areas (M-type)
- - 272 images for farm/grass areas (F-type)
- - 293 images for lake/coastline areas (L-type)
- Randomly selected 90%, 10% for training set and testing set

Aviris Data Portal. [Online]. Available: https://aviris.jpl.nasa.gov/dataportal/.

Experiment

- Spectral compressive acquisition (SpeCA) [Martín'16]
- Spatial/spectral compressed encoder (SPACE) [Lin'20]
- Locally similar sparsity-based hyperspectral unmixing compressive sensing (LSS) [Zhang'16]
- Compressive sensing via joint tensor Tucker decomposition and weighted 3-D total variation regularization (TenTV) [Wang'17]

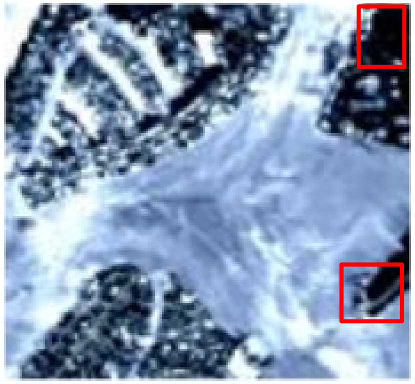
[Martín'16] G. Martín and J. M. Bioucas-Dias, "Hyperspectral blind reconstruction from random spectral projections," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 9, no. 6, pp. 2390–2399, June 2016.
[Lin'20] C.-H. Lin, J. M. Bioucas, T.-H. Lin, Y.-C. Lin, and C.-H. Kao, "A new hyperspectral compressed sensing method for efficient satellite communications," in Proceedings of the 11th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), Hangzhou, China, Jun. 2020. *(Special Session: Unsupervised Computing and Large-Scale Optimization for Multi-dimensional Data Processing)*
[Zhang'16] L. Zhang, W. Wei, Y. Zhang, H. Yan, F. Li, and C. Tian,"Locally similar sparsity-based hyperspectral compressive sensing using unmixing," IEEE Transactions on Computational Imaging, vol. 2, no. 2, pp. 86–100, June 2016.
[Wang'17] Y. Wang, L. Lin, Q. Zhao, T. Yue, D. Meng, and Y. Leung, "Compressive sensing of hyperspectral images via joint tensor Tucker decomposition and weighted total variation regularization," IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 12, pp. 2457–2461, Dec 2017.

Experiment

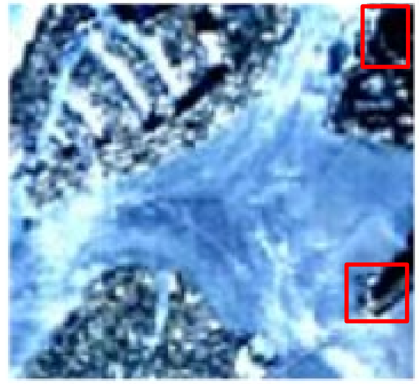
- (Spatial quality) PSNR (dB) – Peak Signal-to-Noise Ratio
- (Global quality) RMSE (degree) – Root Mean Square Error
- (Spectral quality) SAM (degree) – Spectral Angle Mapper

Test Set / Method	C-type	M-type	F-type	L-type
	PSNR↑ / RMSE↓ / SAM↓			
SPACE	24.129/613.661/7.207	29.161/140.415/3.743	29.674/64.151/3.121	27.727/209.757/4.446
SpeCA	9.299/784.867/42.863	15.377/234.735/21.510	11.701/407.530/33.036	14.024/225.772/22.006
TenTV	20.208/570.255/26.247	18.533/260.221/22.972	20.401/248.994/18.714	18.824/314.248/25.523
LSS	7.002/615.037/48.546	0.427/232.486/57.256	3.848/259.960/50.781	2.380/341.429/55.669
HyperCSI-LSS	25.078/278.263/8.704	26.146/51.421/4.907	25.943/82.299/5.732	25.897/83.626/5.779
HCSN (ours)	**34.274/65.120/2.016**	**33.729/30.620/1.631**	**35.908/17.408/1.380**	**35.566/21.558/1.408**
HCSN (C)	34.551/62.437/1.862	30.260/50.947/3.584	34.267/19.361/1.908	33.463/24.162/2.187
HCSN (M)	33.188/78.269/2.731	33.752/30.652/1.595	35.327/18.978/1.550	34.567/27.795/1.801
HCSN (F)	32.834/77.508/2.718	30.074/68.014/4.873	35.750/17.657/1.357	33.137/29.138/2.339
HCSN (L)	33.666/70.175/2.272	31.806/39.403/2.538	34.541/20.117/1.770	34.972/22.456/1.528

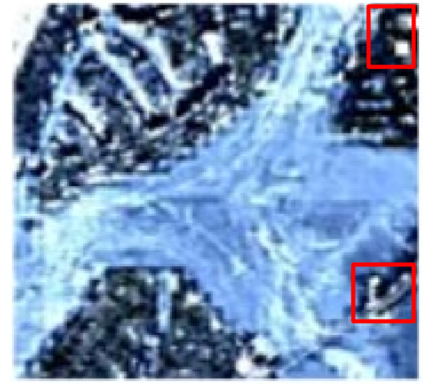
Experiment



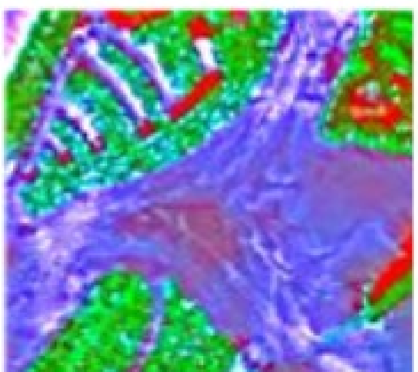
(a) Ground Truth



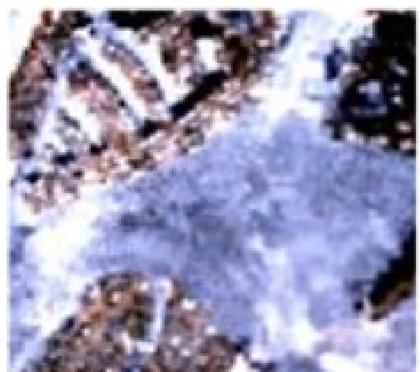
(b) HCSN
SAM: **2.958**



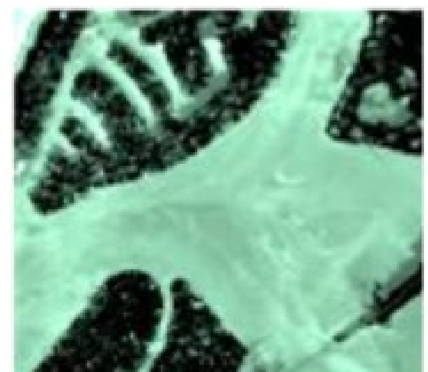
(c) SPACE
SAM: 6.019



(d) LSS
SAM: 59.563



(e) TenTV
SAM: 26.258



(f) SpeCA
SAM: 27.787

Conclusion in HIS SR

- A new deep neural network for HSI compression/reconstruction
- Fast compression by the lightweight encoder
- An efficient decoder which decode the spatial and spectral super-resolution

Outline

- Overview of Deep Learning
  - Supervised – Unsupervised
- Deep super-resolution
  - Traditional super-resolution
  - Structured image super-resolution
    - Face hallucination
  - 2-D image super-resolution (generic images)
  - *N*-D image super-resolution (Hyperspectral images)
- Summary

Summary

- Single image super-resolution still remains several issues to be overcome
  - Good metric beyond GAN loss
    - Visual quality vs math equation
  - Different types of images have different requirements
    - Network architecture design
    - Applications
  - Finding a good prior for super-resolution always works
    - Such as “face hallucination”