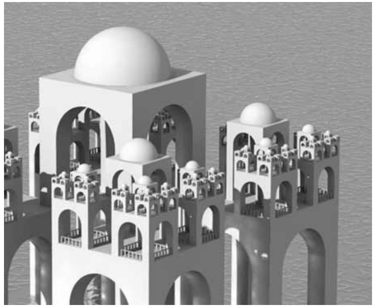
Super-resolution

*For queries contact : Subeesh Vasu (subeeshvasu@gmail.com)

Image super-resolution

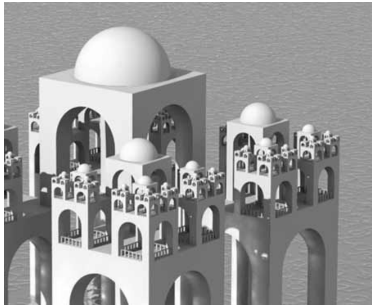
n 

m

*Slide courtesy, Super Resolution From a Single Image, iccv 2009, Glasner

Image super-resolution

n



m

αn

αm

α - scale factor

*Slide courtesy, Super Resolution From a Single Image, iccv 2009, Glasner

Degradation model

I_{HR}

αn

αm

Degradation model

I_{HR}

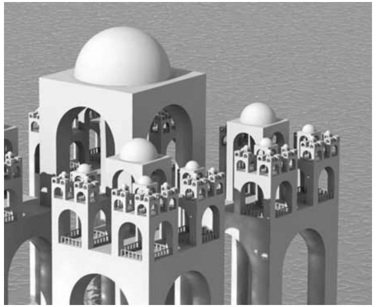
αn

αm

$$\mathrm{f}_\alpha : \mathbb{R}^{\alpha m \times \alpha n} \longrightarrow \mathbb{R}^{m \times n}$$

$$I_{\mathrm{LR}} = \mathrm{f}_\alpha(I_{\mathrm{HR}})$$

n



m

Degradation model

I_{HR}

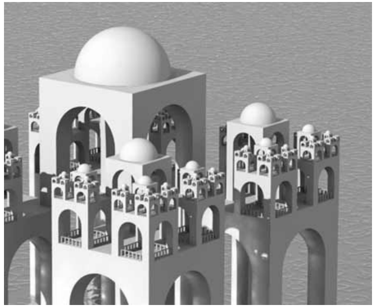
αn

αm

$\mathrm{f}_\alpha : \mathbb{R}^{\alpha m \times \alpha n} \longrightarrow \mathbb{R}^{m \times n}$

n

$I_{\mathrm{LR}} = \mathrm{f}_\alpha(I_{\mathrm{HR}})$



m

f_α - can be a combination of decimation, noise, and blur

Single image super-resolution (Hallucination)

$$\mathrm{g}_\alpha : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}^{\alpha m \times \alpha n}$$

*Slide courtesy, Single image super-resolution, Cs129 Computational Photography James Hays, Brown, fall 2012; Slides of Libin “Geoffrey” Sun and James Hays

Single image super-resolution (Hallucination)

$$g_\alpha : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}^{\alpha m \times \alpha n}$$

Main goals

- Be faithful to the low resolution input image $I_{\mathrm{HR}}^{est} = g_\alpha(I_{\mathrm{LR}})$

*Slide courtesy, Single image super-resolution, Cs129 Computational Photography James Hays, Brown, fall 2012; Slides of Libin “Geoffrey” Sun and James Hays

Single image super-resolution (Hallucination)

$$g_\alpha : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}^{\alpha m \times \alpha n}$$

Main goals

- Produce a detailed, realistic output image

*Slide courtesy, Single image super-resolution, Cs129 Computational Photography James Hays, Brown, fall 2012; Slides of Libin “Geoffrey” Sun and James Hays

Single image super-resolution (Hallucination)

$$g_\alpha : \mathbb{R}^{m\times n} \longrightarrow \mathbb{R}^{\alpha m\times \alpha n}$$

Main goals

- Be faithful to the low resolution input image $I_{\mathrm{HR}}^{est} = g_\alpha(I_{\mathrm{LR}})$
- Produce a detailed, realistic output image

*Slide courtesy, Single image super-resolution, Cs129 Computational Photography James Hays, Brown, fall 2012; Slides of Libin “Geoffrey” Sun and James Hays

Single Image SR via Deep Learning

- SRCNN (Super-Resolution Convolutional Neural Networks) : ECCV 2014, PAMI 2016 : 3 layer CNN, MSE loss, input - bicubic interpolated image

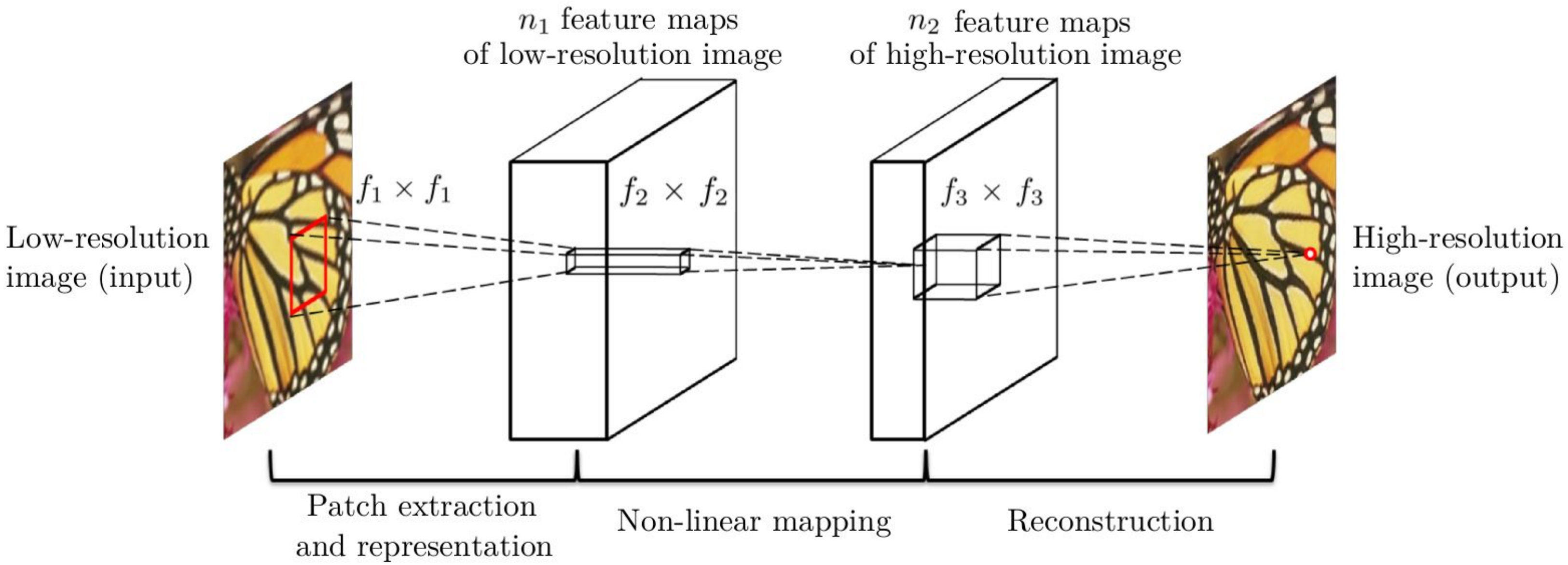
- Follow up works => focused on improvement in run-time, accuracy, perceptual quality, and extensions to videos

- Some of the key findings

    - larger context, training tricks, architectural modifications=> better performance
    - feature extraction in LR dimension => improvement in speed
    - VGG Loss and GAN loss => better perceptual quality
    - Recurrent networks => for videos

*ECCV - European Conference on Computer Vision, PAMI - IEEE Transactions on Pattern Analysis and Machine Intelligence

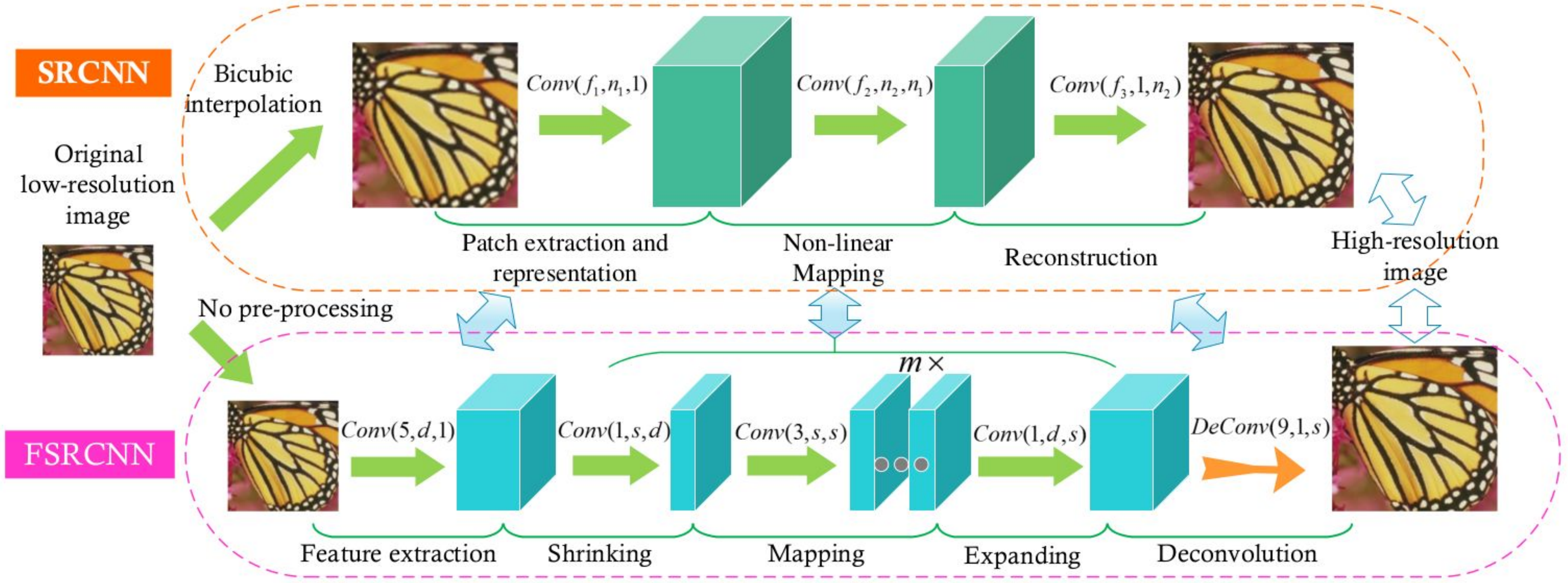
SRCNN (ECCV 2014, PAMI 2016) - net arch



SRCNN details

- Loss function $L(\Theta) = \frac{1}{n}\sum_{i=1}^{n} ||F(\mathbf{Y}_i;\Theta) - \mathbf{X}_i||^2$
- input is the bicubic interpolated form of LR image
- number of filters : n1 = 64, and n2 = 32
- filter size (f1 - f2 - f3) : 9-1-5, 9-3-5, and 9-5-5
- 9-5-5 performs best => utilizing neighborhood information in the mapping stage is beneficial
- experiments with more layers didn’t succeed => “the deeper the better” doesn’t hold true with this deep model for super-resolution

FSRCNN (Fast Super-Resolution Convolutional Neural Network : ECCV 2016) net arch



FSRCNN details

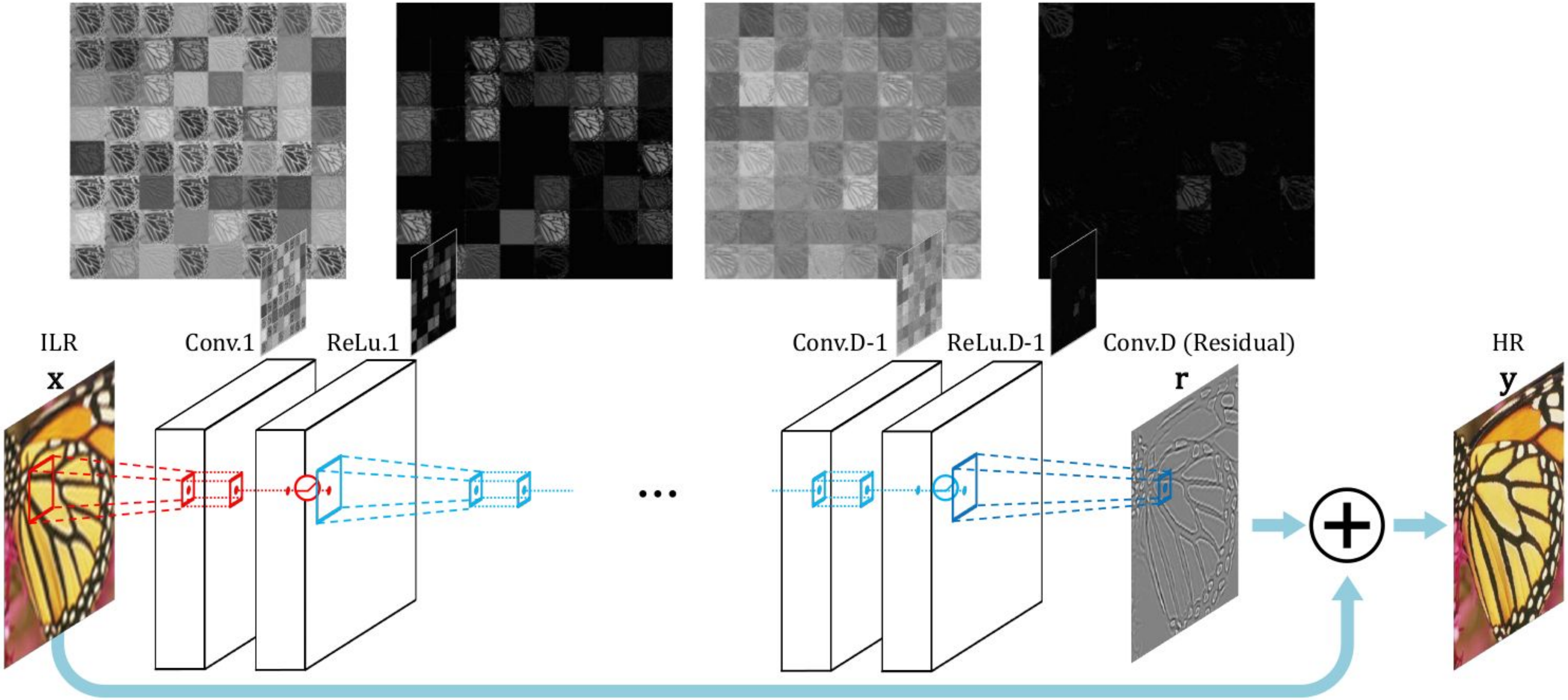
- Faster + better performance as compared to SRCNN

- 8 layer CNN, MSE loss, input - original LR image

- Tricks for speed up - All layers except last works on LR dimension, last layer performs upsampling using a deconv layer, middle layer filters are of size 3 x 3, and contain lesser number of filters

- total number of parameters reduced by 4 and speedup of 40 x as compared to SRCNN

VDSR (Accurate Image Super-Resolution Using Very Deep Convolutional Networks : CVPR 2016) net arch

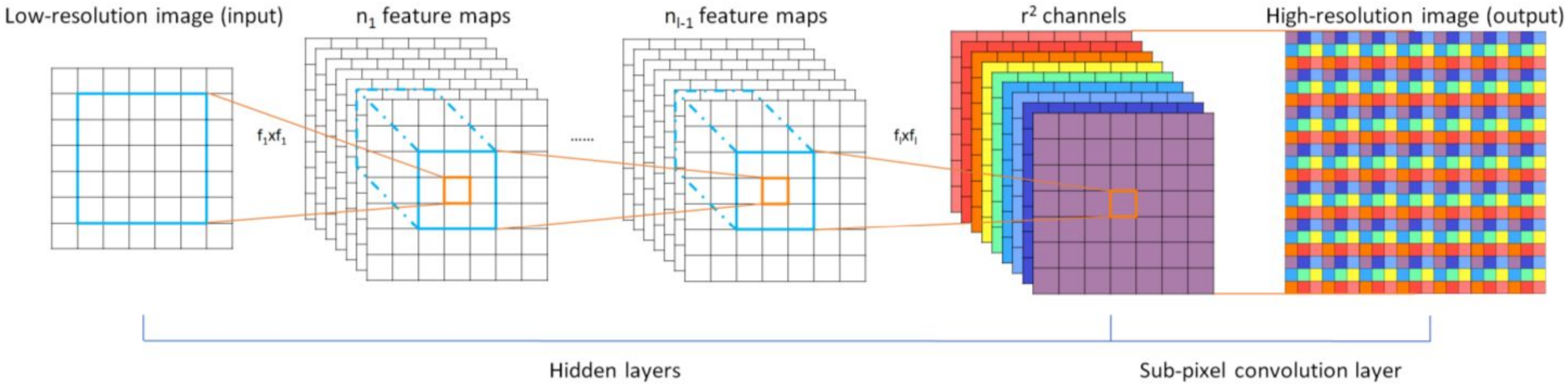


*CVPR - IEEE Conference on Computer Vision and Pattern Recognition

VDSR details

- very deep CNN, 20 layers
- to improve convergence they train for residual + use higher learning rate and adjustable gradient clipping
- input - bicubic interpolated form of LR image; LOSS - MSE between residual and network output
- context : utilize contextual information spread over very large image regions

ESPCN (Efficient Sub-Pixel Convolutional Neural Network : CVPR 2016) net arch



ESPCN details

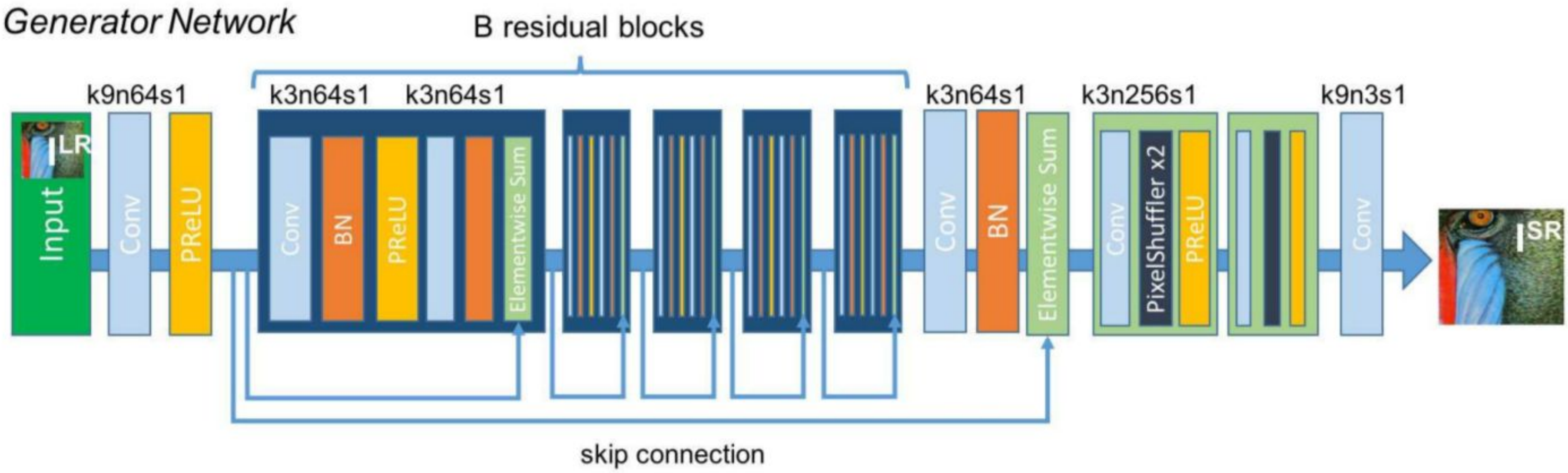
- Faster + better performance as compared to SRCNN

- 3 layer CNN (similar to SRCNN), MSE loss, input - original LR image

- Tricks for speed up - All layers except last works on LR dimension, last layer performs upsampling using an efficient sub-pixel convolution layer

- They perform video SR by applying single image SR in a frame by frame fashion. Speed up achieved by ESPCN allowed them to perform real-time SR of 1080p videos

SRResNet (super resolution residual network : CVPR 2017) net arch (photo-realistic SR)



SRResNet details

- deep CNN, with 16 residual blocks
- Modified form of ResNet is used to build network architecture
- Used as the generator network for the photo-realistic SR work by Ledig et al (CVPR 2017)
- input - LR image; LOSS - MSE
- Significant improvement in performance as compared to SRCNN, ESPCN, DRCN

Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network : SRGAN (CVPR 2017) net arch

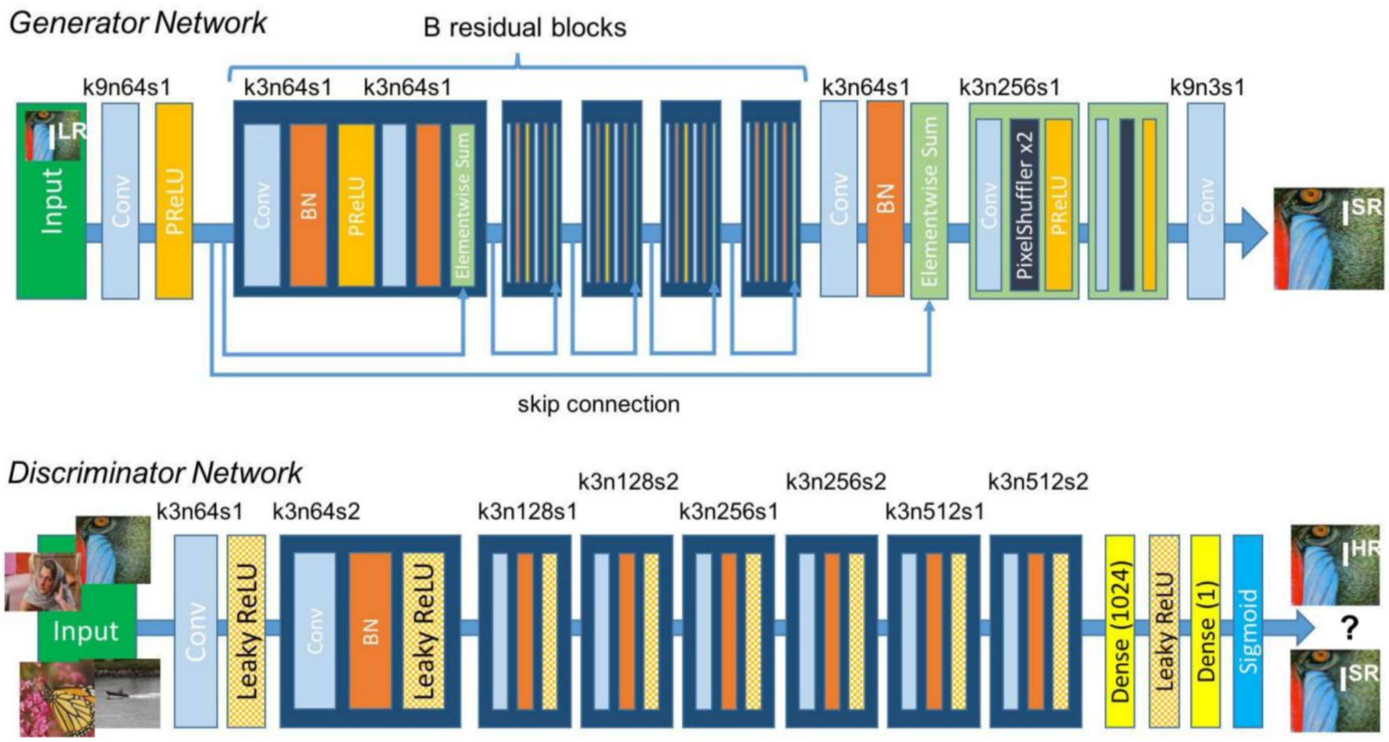


Figure 4: Architecture of Generator and Discriminator Network with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer.

SRGAN details

- SRResNet as generator
- Loss = combinations of (MSE, VGG loss, Adversarial loss)
- Input - LR image; efficient sub-pixel convolution layer for upsampling
- Better MOS (Mean Opinion Score) as compared to existing works

The Perception-Distortion Tradeoff (CVPR 2018)

- distortion and perceptual quality are at odds with each other

The Perception-Distortion Tradeoff (CVPR 2018)

- distortion and perceptual quality are at odds with each other
- distortion metrics - PSNR, SSIM, IFC

The Perception-Distortion Tradeoff (CVPR 2018)

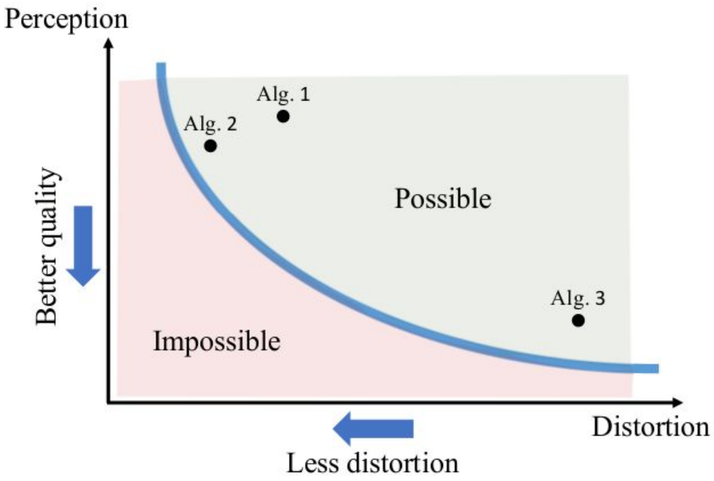
- distortion and perceptual quality are at odds with each other

- distortion metrics - PSNR, SSIM, IFC

- perceptual quality metrics - mean opinion score (MOS), no-reference quality measures such as NIQE

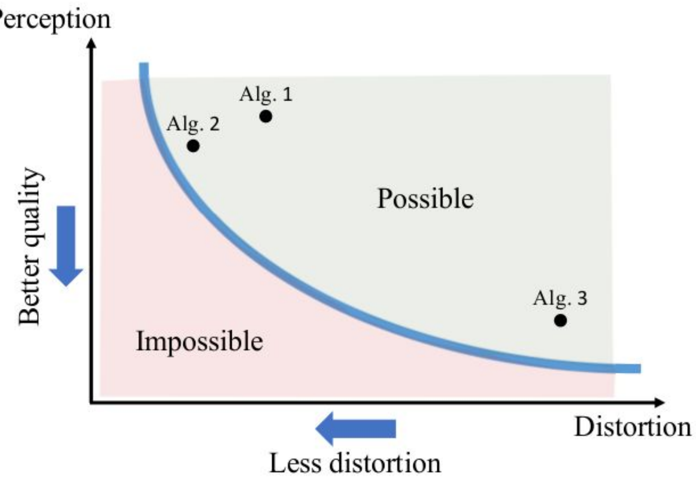
The Perception-Distortion Tradeoff (CVPR 2018)

- distortion and perceptual quality are at odds with each other



The Perception-Distortion Tradeoff (CVPR 2018)

- distortion and perceptual quality are at odds with each other



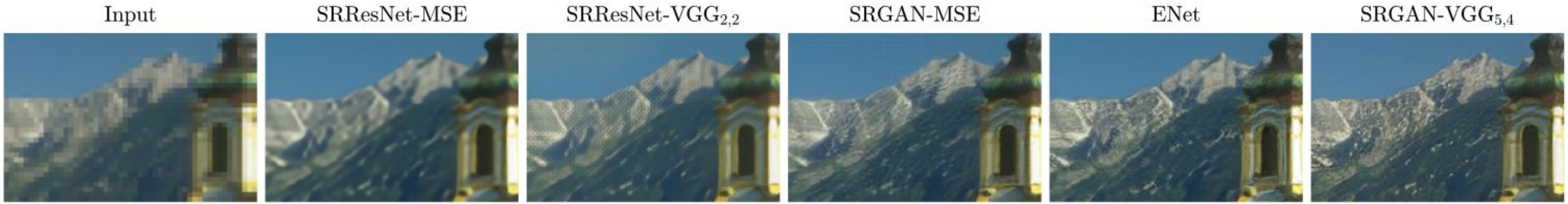
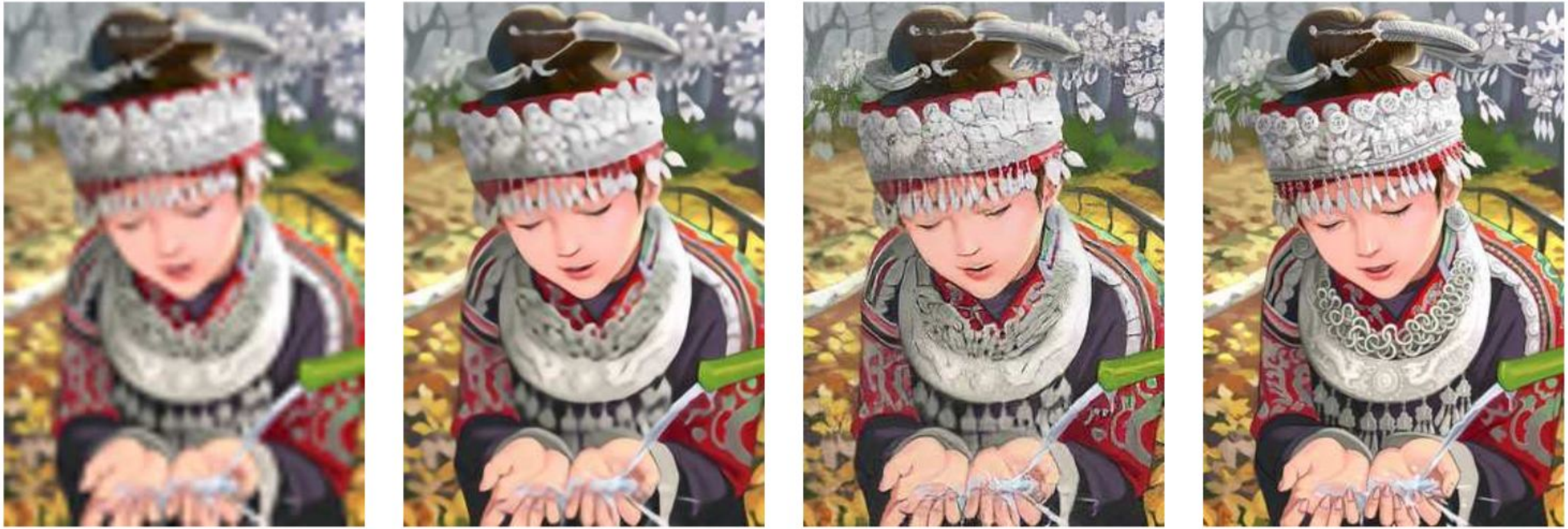


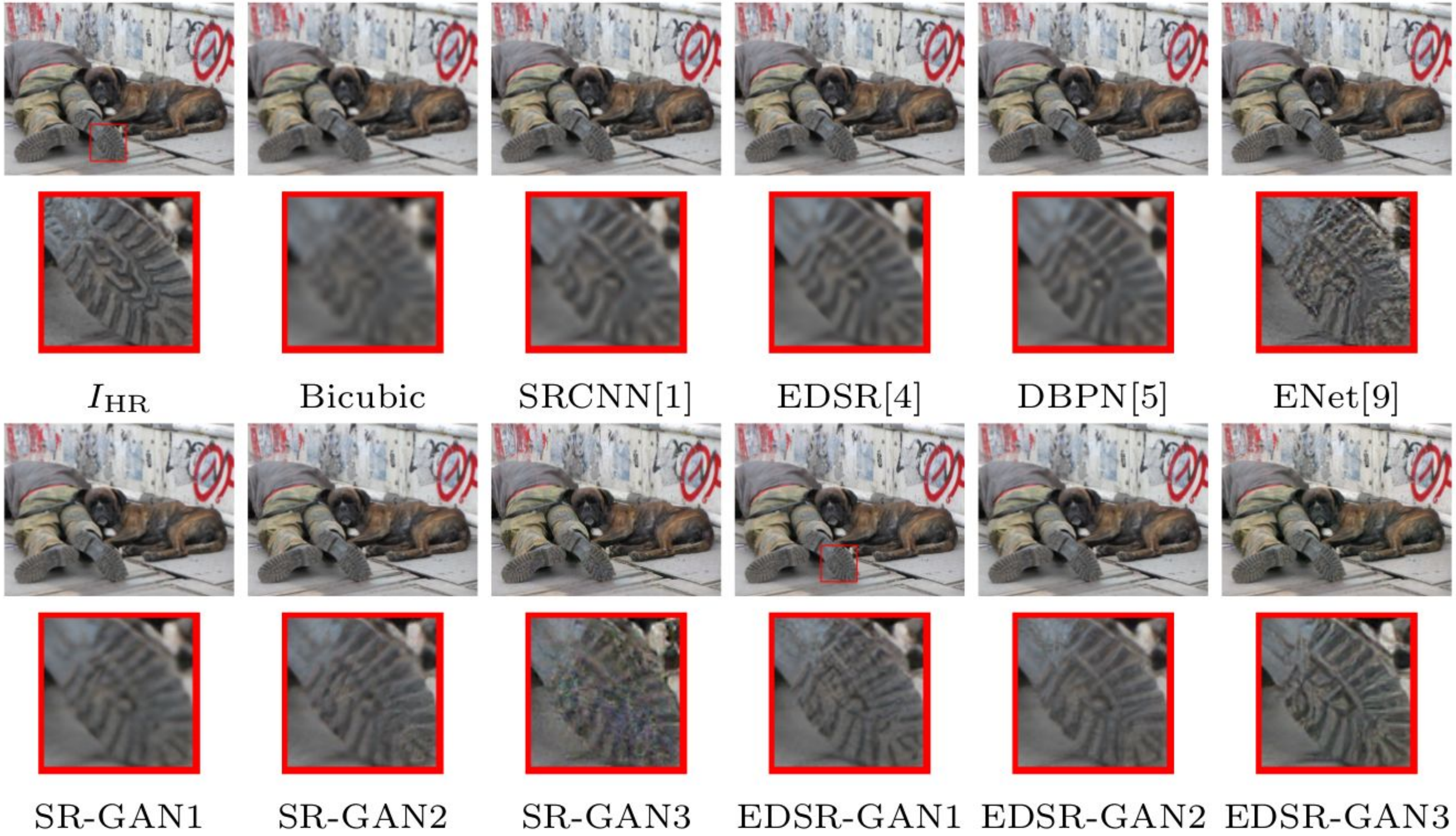
Figure 9. **Visual comparison of algorithms closest to the perception-distortion bound.** The algorithms are ordered from low to high distortion (evaluated by IFC). Notice the co-occurring increase in perceptual quality.

SR Results for x4 - CVPR 2017

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

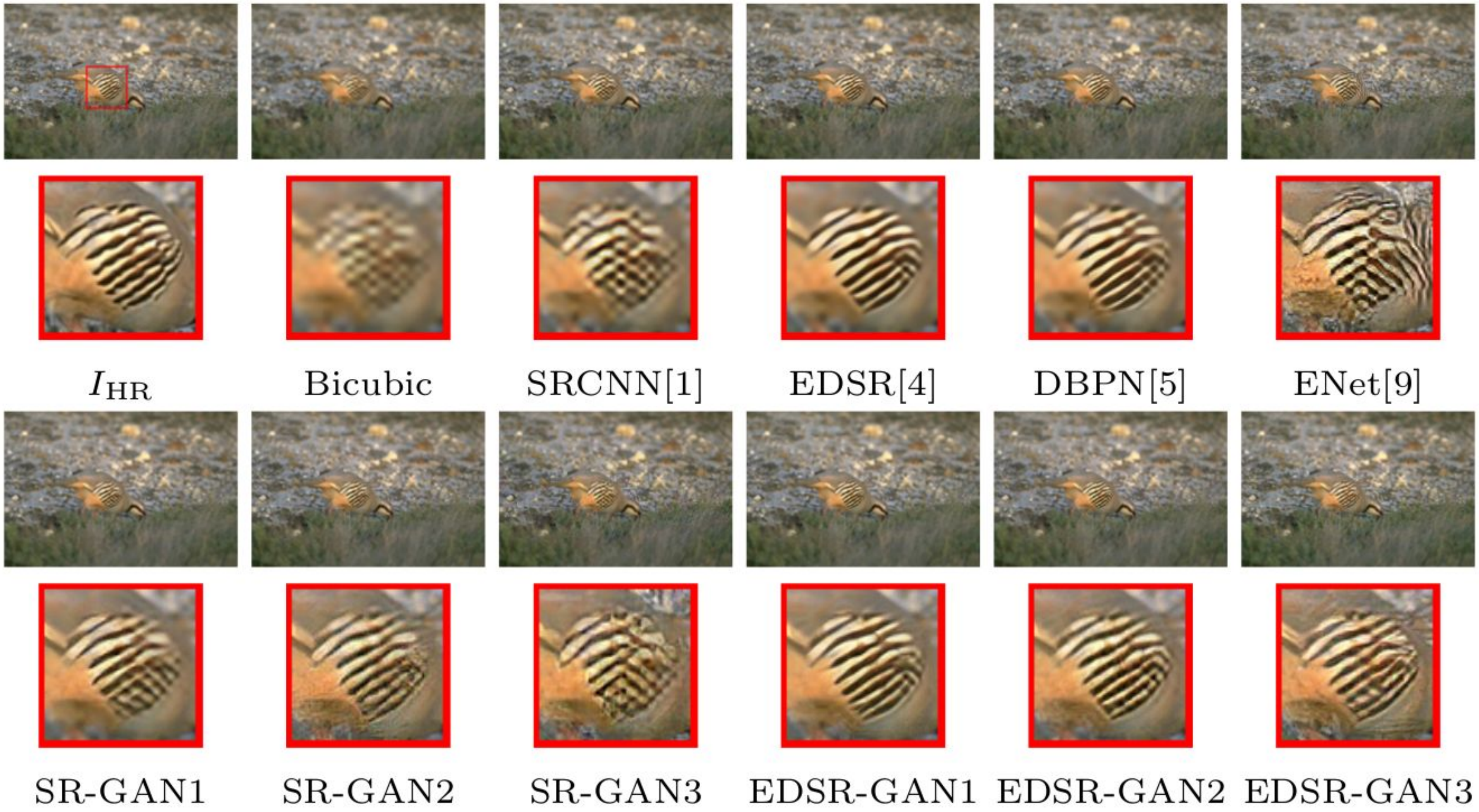


SR Results for x4 - ECCV 2018



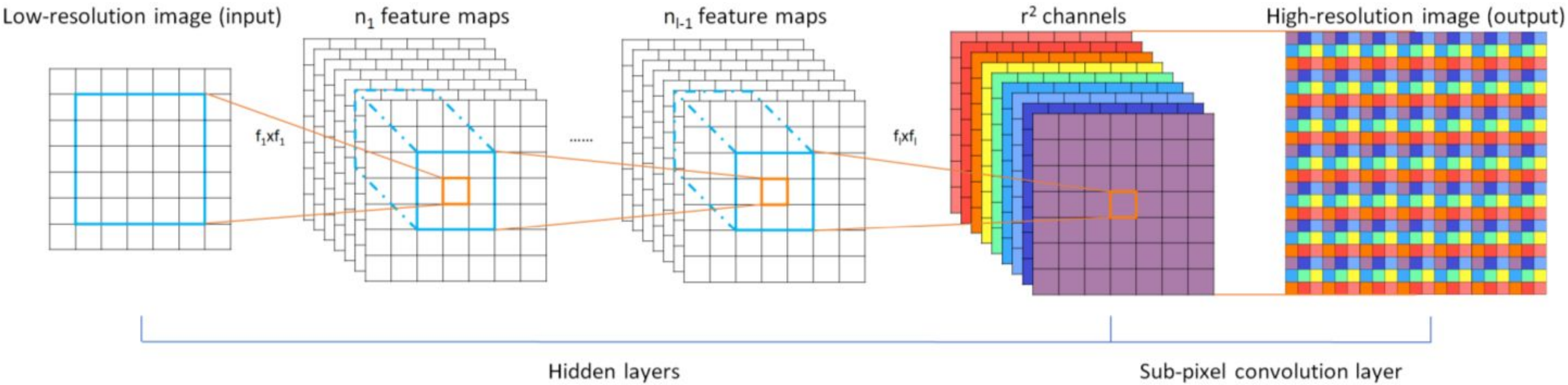
*EDSR - CVPRW 2017, DBPN - CVPR 2018, ENet - ICCV 2017, EDSR-GAN - ECCVW 2018

SR Results for x4 - ECCV 2018



Video super-resolution

Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network (CVPR 2016)



- They perform video SR by applying single image SR in a frame by frame fashion. Speed up achieved by ESPCN allowed them to perform real-time SR of 1080p videos

Real-Time Video Super-Resolution with Spatio-Temporal Networks and Motion Compensation : VESPCN (CVPR 2017)

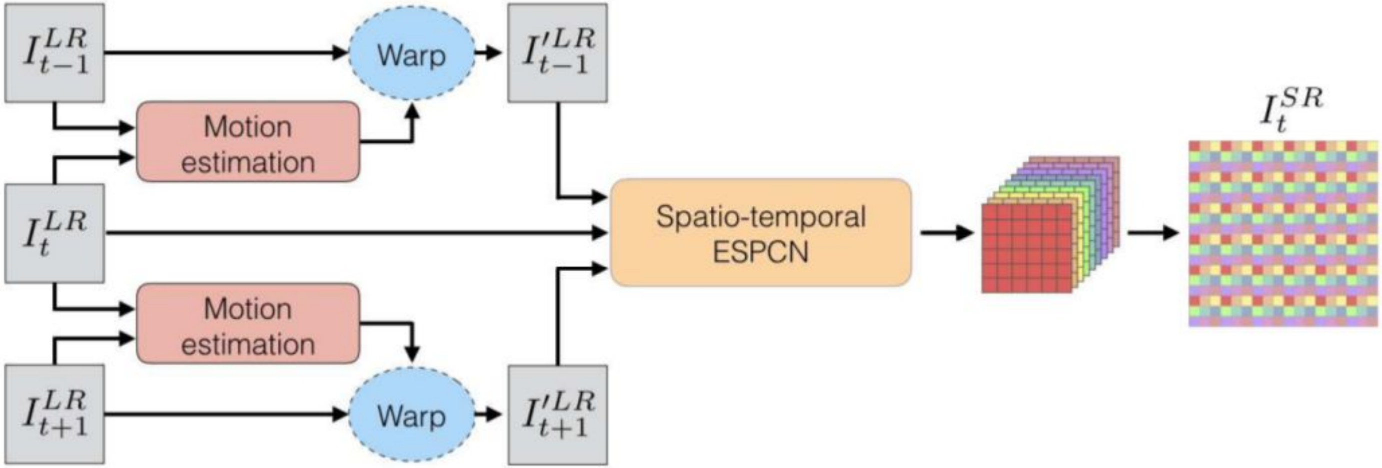


Figure 1: Proposed design for video SR. The motion estimation and ESPCN modules are learnt end-to-end to obtain a motion compensated and fast algorithm.

Frame-Recurrent Video Super-Resolution (CVPR 2018)

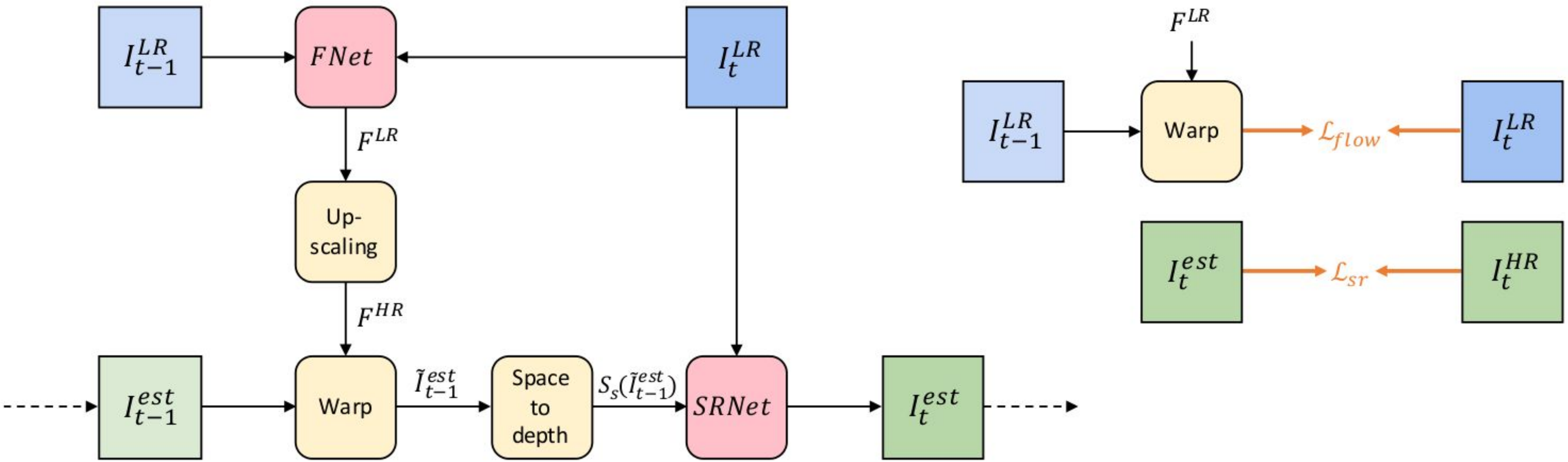


Figure 2: Overview of the proposed FRVSR framework (left) and the loss functions used for training (right). After computing the flow F^{LR} in LR space using FNet, we upsample it to F^{HR}. We then use F^{HR} to warp the HR-estimate of the previous frame I_{t-1}^{est} onto the current frame. Finally, we map the warped previous output $\tilde{I}_{t-1}^{\mathrm{est}}$ to LR-space using the space-to-depth transformation and feed it to the super-resolution network SRNet along with the current input frame I_t^{LR}. For training the networks (shown in red), we apply a loss on I_t^{est} as well as an additional loss on the warped previous LR frame to aid FNet.

References used for constructing the slides

- Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: European conference on computer vision, Springer (2014) 184–199 : **SRCNN**
- Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence 38(2) (2016) 295–307 : **SRCNN**
- Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2016) 1646–1654 : **VDSR**
- Shi, W., Caballero, J., Husz´ar, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 1874–1883 : **ESPCN**
- Dong, C., Loy, C.C., Tang, X.: Accelerating the super-resolution convolutional neural network. In: European Conference on Computer Vision, Springer (2016) 391–407 : **FSRCNN**
- Ledig, C., Theis, L., Husz´ar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A.P., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image superresolution using a generative adversarial network. In: CVPR. Volume 2. (2017) : **SRResNet**, **SRGAN**
- Blau, Y., Michaeli, T.: The perception-distortion tradeoff. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (June 2018)
- Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: The IEEE conference on computer vision and pattern recognition (CVPR) workshops. Volume 1. (2017) : **EDSR**
- Haris, M., Shakhnarovich, G., Ukita, N.: Deep backprojection networks for superresolution. In: Conference on Computer Vision and Pattern Recognition. (2018) : **DBPN**
- Sajjadi, M.S., Sch¨olkopf, B., Hirsch, M.: Enhancenet: Single image super-resolutionthrough automated texture synthesis. In: Computer Vision (ICCV), 2017 IEEE International Conference on, IEEE (2017) 4501–4510 : **ENet**
- Subeesh V., Nimisha T. M., and A. N. Rajagopalan, Analyzing Perception-Distortion Tradeoff in SISR using EDSR-GAN. In Perceptual Image Restoration and Manipulation (PIRM) Workshop and Challenge, Eurpean Conference on Computer Vision Workshops (ECCVW 2018), Munich, Germany, September 2018 : **EDSR-GAN**
- Caballero, J. and Ledig, C. and Aitken, A. P. and Acosta, A. and Totz, J. and Wang, Z. and Shi, W.:, Real-Time Video Super-Resolution with Spatio-Temporal Networks and Motion Compensatio, CVPR 2017 : **VESPCN**
- Sajjadi, M. SM and Vemulapalli, Raviteja and Brown, Matthew:, Frame-Recurrent Video Super-Resolution, CVPR 2018

Other references

- Lai, Wei-Sheng and Huang, Jia-Bin and Ahuja, Narendra and Yang, Ming-Hsuan, Deep laplacian pyramid networks for fast and accurate superresolution. CVPR 2017
- Tai, Ying and Yang, Jian and Liu, Xiaoming, Image super-resolution via deep recursive residual network, CVPR 2017
- Tong, Tong and Li, Gen and Liu, Xiejie and Gao, Qinquan, Image super-resolution using dense skip connections, ICCV 2017
- Zhang, Yulun and Tian, Yapeng and Kong, Yu and Zhong, Bineng and Fu, Yun, Residual dense network for image super-resolution, CVPR 2018
- Wang, Xintao and Yu, Ke and Dong, Chao and Loy, Chen Change, Recovering realistic texture in image super-resolution by deep spatial feature transform, CVPR 2018
- Zhang, Kai and Zuo, Wangmeng and Zhang, Lei, Learning a single convolutional super-resolution network for multiple degradation, CVPR 2018
- Zhang, Yulun and Li, Kunpeng and Li, Kai and Wang, Lichen and Zhong, Bineng and Fu, Yun, Image Super-Resolution Using Very Deep Residual Channel Attention Networks, ECCV 2018
- Tao, Xin and Gao, Hongyun and Liao, Renjie and Wang, Jue and Jia, Jiaya, Detail-revealing deep video super-resolution, ICCV 2017