[A2013] 시각지능심화응용 C

Single Image Super-Resolution

시각지능연구실 김종희 선임연구원

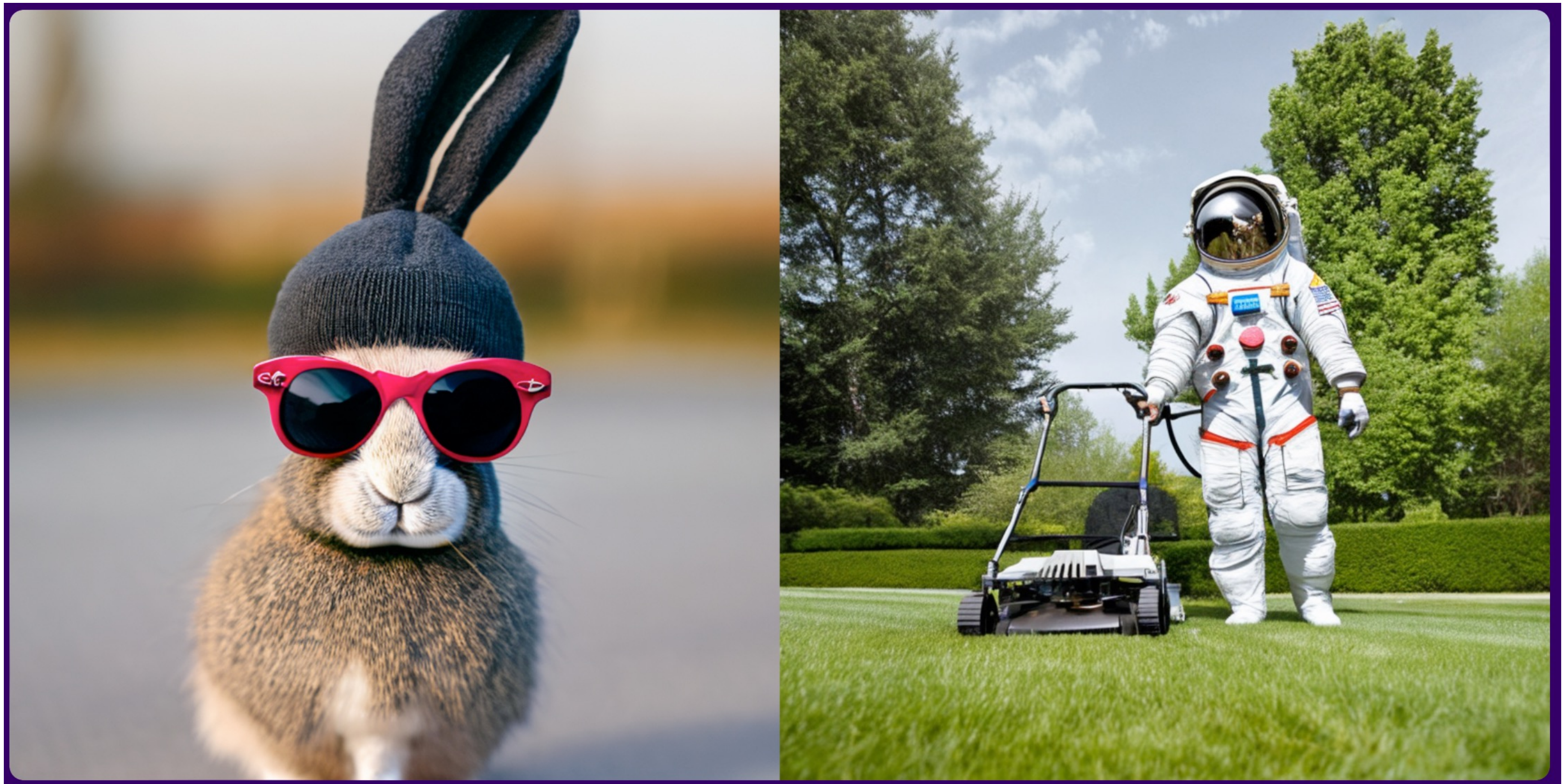
Contents

Introduction

- Single image super-resolution
  - Reconstructing high-resolution image from a low-resolution image.
  - Image processing > Image restoration > Super-resolution

Introduction

- Super-resolution in Stable Diffusion 2
  - Stable Diffusion 2.0 also includes an Upscaler Diffusion model that enhances the resolution of images by a factor of 4.



Examples of images produced using Stable Diffusion 2.0, at 768x768 image resolution.

Introduction

- Super-resolution in Stable Diffusion 2
  - Stable Diffusion 2.0 also includes an Upscaler Diffusion model that enhances the resolution of images by a factor of 4.

Left: 128x128 low-resolution image. Right: 512x512 resolution image produced by Upscaler.

Introduction

- Low-resolution (LR) image $I_x = D\left(I_y; \delta\right)$
  - Corresponding high-resolution (HR) image I_y
  - A degradation mapping function D
    - The parameters of the degradation process δ
    - Most works directly model the degradation as a single downsampling operation $D\left(I_y; \delta\right) = \left(I_y\right) \downarrow_s, \{s\} \subset \delta$
      → Supervised SR
    - There are other works modeling the degradation as a combination of several operations $D\left(I_y; \delta\right) = \left(I_y \otimes k\right) \downarrow_s, +\eta_\zeta, \{k, s, \zeta\} \subset \delta$
      → Unsupervised SR

- The objective of SR
  - $\hat{\theta} = \arg\min_{\theta} L\left(\hat{I}_y, I_y\right) + \lambda \Phi(\theta)$
    - $\hat{I}_y = F(I_x; \theta)$
    - F: super-resolution networks

Introduction

- Single image super-resolution (SISR)
  - Supervised / Unsupervised
  - PSNR-oriented / Perceptual-oriented (GAN)
  - Paired / Unpaired (Real-world)
  - SOTA / Efficient

- Video super-resolution
- Face hallucination (super-resolution)
- Depth map super-resolution
- Hyperspectral image super-resolution
- Light field super-resolution
- MRI super-resolution
- Reference-based super-resolution

Introduction

- Recent trends (Workshops and Awesome github)
  - NTIRE 2023 with CVPR (New Trends in Image Restoration and Enhancement)

    - Single Image Super-Resolution (X4) Bicubic
    - Real-Time Image Super-Resolution - Track 1
    - Real-Time Image Super-Resolution - Track 2
    - Efficient Image Super-Resolution
    - Light Field Image Super-Resolution Challenge
    - Stereo Image Super-Resolution - Track 1 Fidelity & Bicubic
    - Stereo Image Super-Resolution - Track 2 Perceptual & Bicubic
    - Stereo Image Super-Resolution - Track 3 Fidelity & Realistic

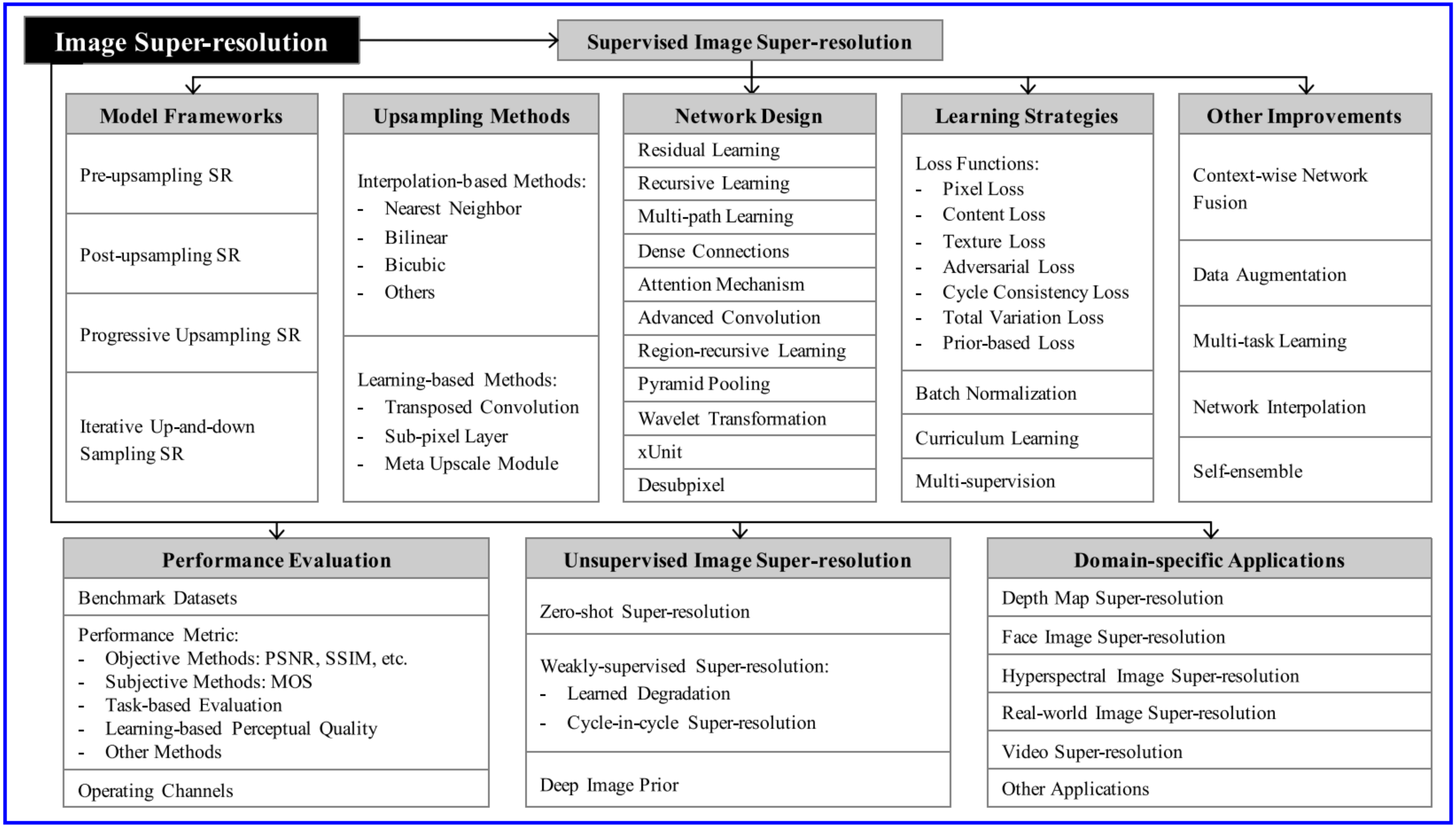
  - AIM 2022 with ICCV/ECCV (Advances in Image Manipulation)

    - Mobile AI & AIM: Real-Time Image Super-Resolution
    - Mobile AI & AIM: Real-Time Video Super-Resolution
    - AIM: Compressed Input Super-Resolution - Track 1 Image
    - AIM: Compressed Input Super-Resolution - Track 2 Video

  - Awesome super-resolution
    - https://github.com/ChaofWang/Awesome-Super-Resolution

Introduction

- Overview

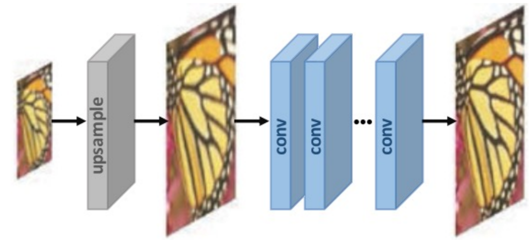


* Z. Wang *et al.*, "Deep Learning for Image Super-Resolution: A Survey," in IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 10, pp. 3365-3387, 1 Oct. 2021

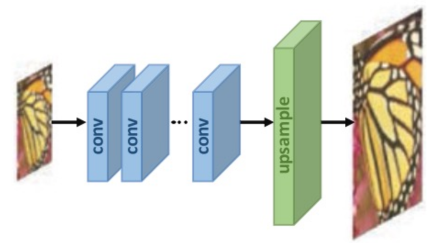
Supervised Super-Resolution - SR Frameworks

Supervised Super-Resolution

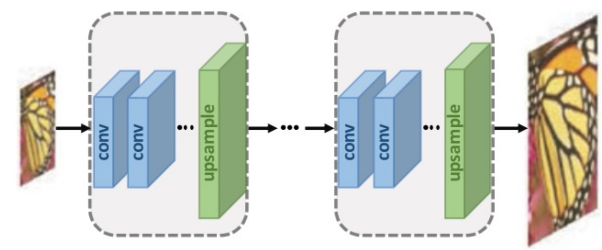
- Super-resolution frameworks
  - Pre-upsampling super-resolution
    - Apply interpolation and then refine
  - Post-upsampling super-resolution
    - Learnable upsampling layer
  - Progressive upsampling super-resolution
    - e.g. $\times 2 \rightarrow \times 4$
  - Iterative up-and-down sampling super-resolution
    - Upsample → downsample → …



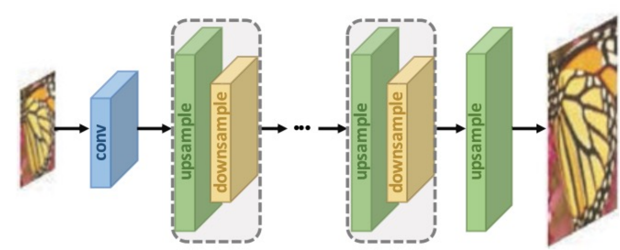
(a) Pre-upsampling SR



(b) Post-upsampling SR



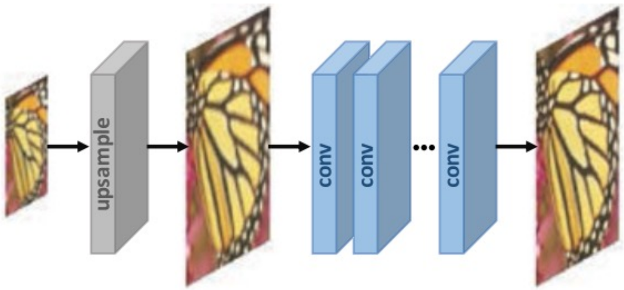
(c) Progressive upsampling SR



(d) Iterative up-and-down Sampling SR

Supervised Super-Resolution

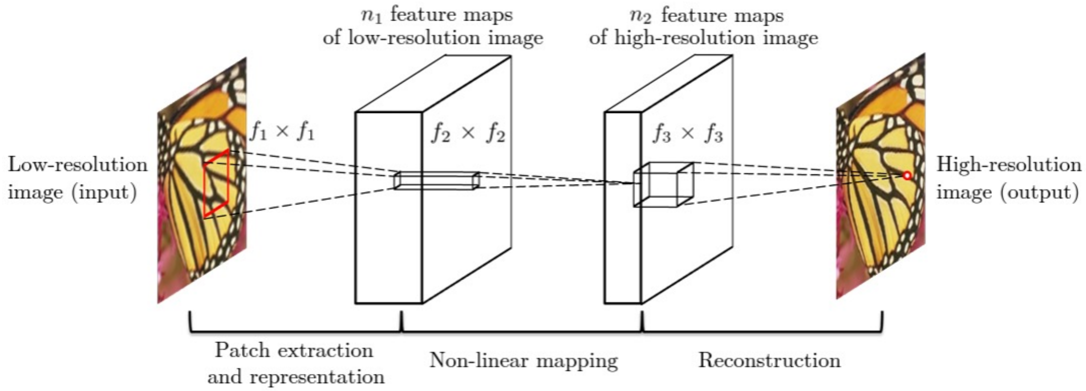
- Super-resolution frameworks
  - Pre-upsampling super-resolution
    - Directly learning the mapping from LR to HR → difficult
    - Apply Interpolation to obtain HR images and then refine them.
      → reduces the learning difficulty.
      - Since the most difficult upsampling operation has been completed, CNNs only need to refine the coarse images.
    - However, the predefined upsampling often introduce side effects.
      - e.g., noise amplification and blurring
    - Since most operations are performed in HR space,
      the cost of time and space is much higher than other frameworks.



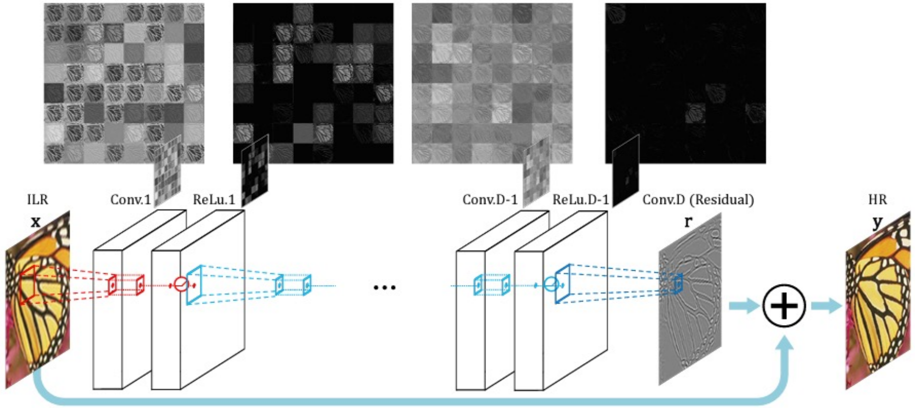
(a) Pre-upsampling SR

Supervised Super-Resolution

- Super-resolution frameworks
  - Pre-upsampling super-resolution
    - SRCNN*: First CNN-based SR, 3-layer networks



    - VDSR**: Deep CNN-based SR, 20-layer networks

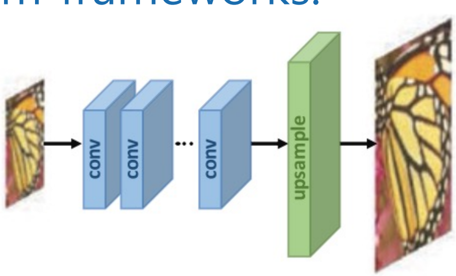


* C. Dong *et al.*, "Image super-resolution using deep convolutional networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 2, pp. 295–307, Feb. 2016.
** J. Kim *et al.*, "Accurate image super-resolution using very deep convolutional networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 1646–1654.

Supervised Super-Resolution

- Super-resolution frameworks
  - Post-upsampling super-resolution
    - Disadvantages of pre-upsampling SR
      - Feature extraction in HR space → the cost of time and space ↑
      - The predefined upsampling often introduce side effects.
    - In order to improve the computational efficiency and make full use of deep learning technology to increase resolution.
    - The LR input images are fed into deep CNNs without increasing resolution, and end-to-end learnable upsampling layers are applied at the end of the network.
      - Feature extraction process in LR space reduces the computation and spatial complexity.

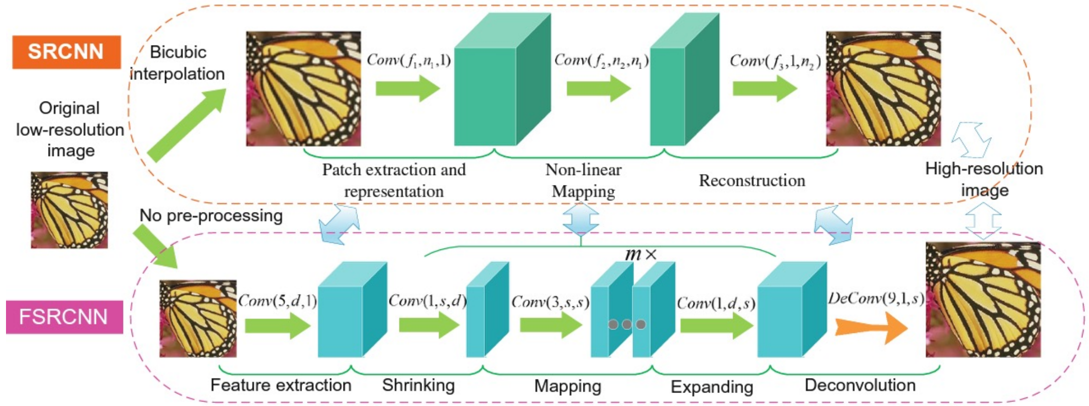
    → The most mainstream frameworks.



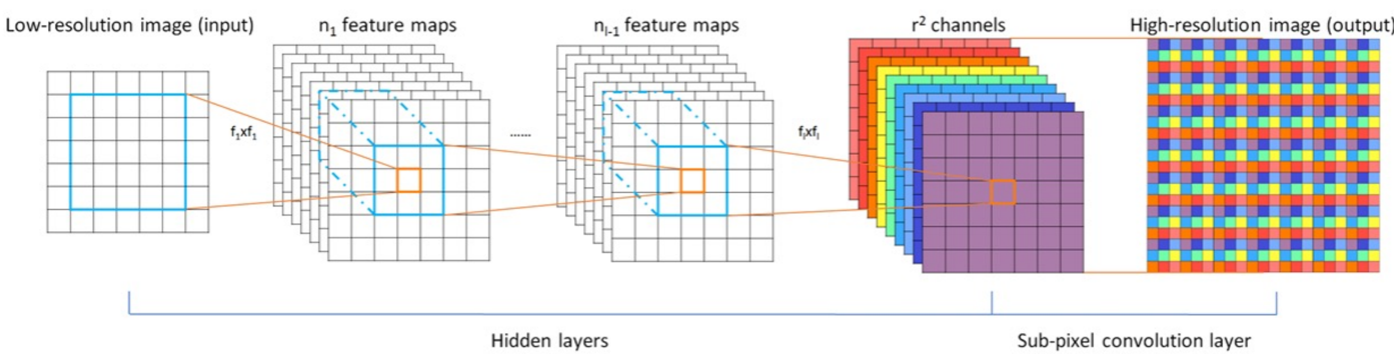
(b) Post-upsampling SR

Supervised Super-Resolution

- Super-resolution frameworks
  - Post-upsampling super-resolution
    - F-SRCNN*: LR feature extraction → Deconvolution



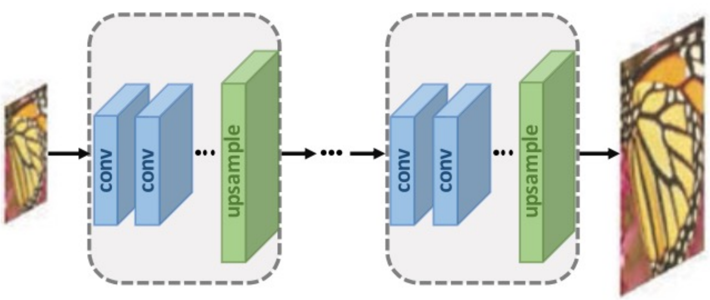
    - ESPCN**: LR feature extraction → Sub-pixel convolution (PixelShuffle)



* C. Dong *et al.*, "Accelerating the super-resolution convolutional neural network," in Proc. Eur. Conf. Comput. Vis., 2016, pp 391–407.
** W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 1874–1883.

Supervised Super-Resolution

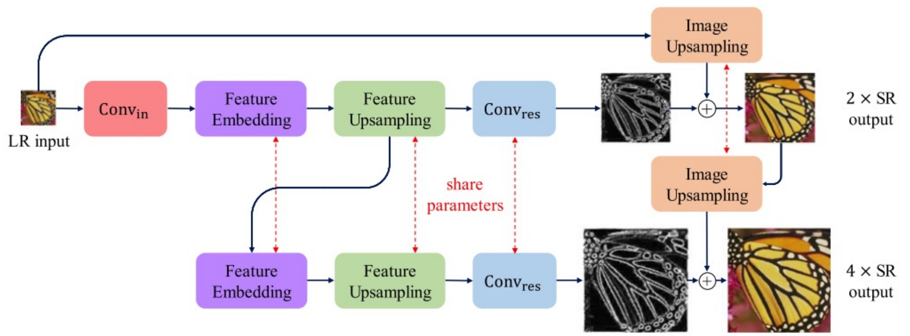
- Super-resolution frameworks
  - Progressive upsampling super-resolution
    - Disadvantages of post-upsampling SR
      - Direct upsampling greatly increases the learning difficulty for large scaling factors (e.g., 4, 8).
      - Each scaling factor requires training an individual SR model, which cannot cope with the need for multi-scale SR.
    - To address these, a progressive upsampling framework is adopted.
      - A cascade of CNNs progressively reconstructs HR images.
      - At each stage, images are upsampled to HR and refined by CNNs.
      - e.g. $\times 2 \rightarrow \times 4 \rightarrow \times 8$



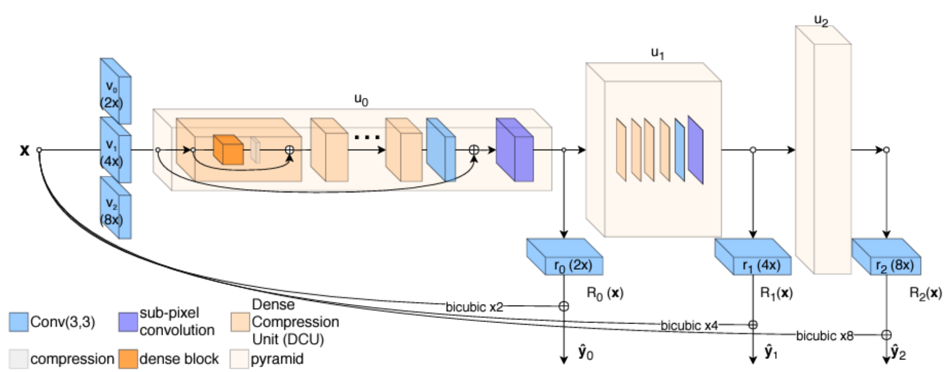
(c) Progressive upsampling SR

Supervised Super-Resolution

- Super-resolution frameworks
  - Progressive upsampling super-resolution
    - LapSRN*: a cascade of CNNs
    - MS-LapSRN**: a cascade of shared CNNs



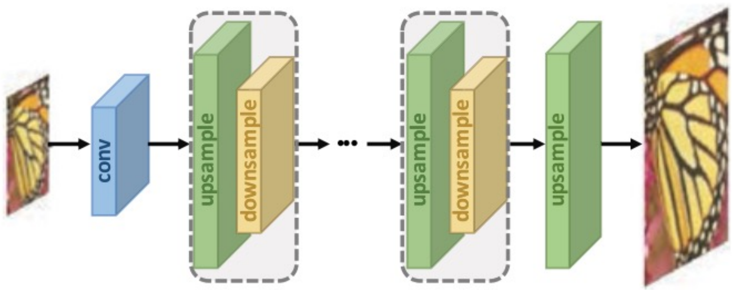
    - ProSR***: Asymmetric cascade + scale-wise convolution for 1st layer



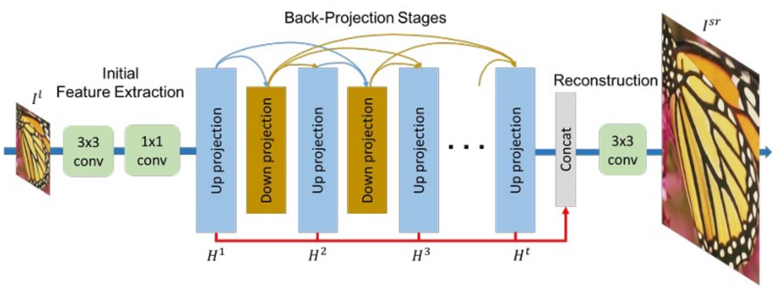
* W.-S. Lai *et al.*, "Deep laplacian pyramid networks for fast and accurate super-resolution," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 5835–5843.
** W.-S. Lai *et al.*, "Fast and accurate image super-resolution with deep laplacian pyramid networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 11, pp. 2599–2613, Nov. 2018.
*** Y. Wang *et al.*, "A fully progressive approach to single-image super-resolution," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, 2018, pp. 9770–9779.

Supervised Super-Resolution

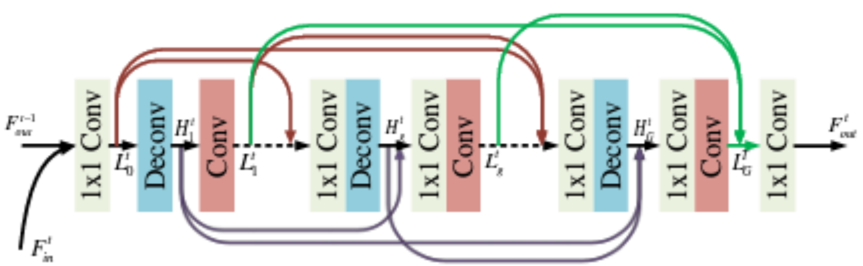
- Super-resolution frameworks
  - Iterative up-and-down sampling super-resolution
    - Apply the concept of back-projection (HR → LR)



(d) Iterative up-and-down Sampling SR



DBPN*

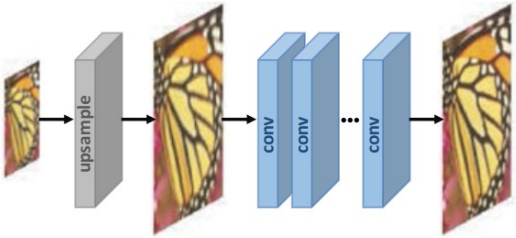


SRFBN**

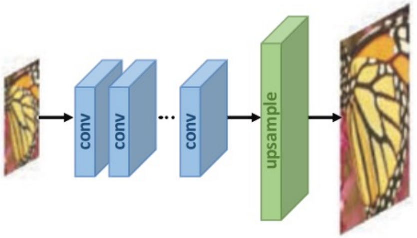
* M. Haris *et al.*, "Deep back-projection networks for super-resolution," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 1664–1673.
** Z. Li *et al.*, "Feedback network for image super-resolution," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2019, pp. 3862–3871.

Supervised Super-Resolution

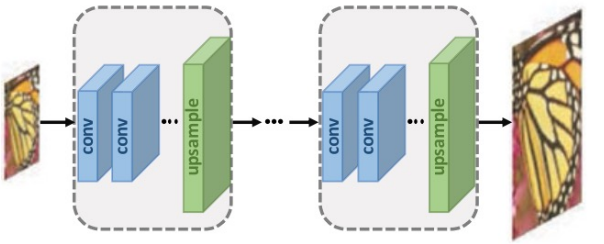
- Super-resolution frameworks
  - Pre-upsampling super-resolution
    - Apply interpolation and then refine
  - Post-upsampling super-resolution
    - Learnable upsampling layer
  - Progressive upsampling super-resolution
    - e.g. ×2 → ×4
  - Iterative up-and-down sampling super-resolution
    - Upsample → downsample → …



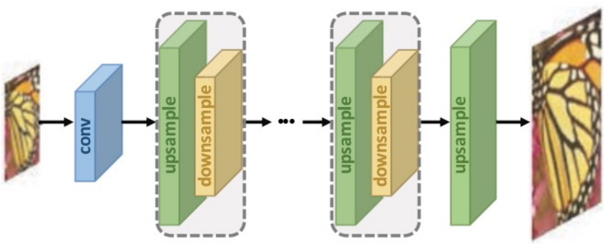
(a) Pre-upsampling SR



(b) Post-upsampling SR



(c) Progressive upsampling SR



(d) Iterative up-and-down Sampling SR

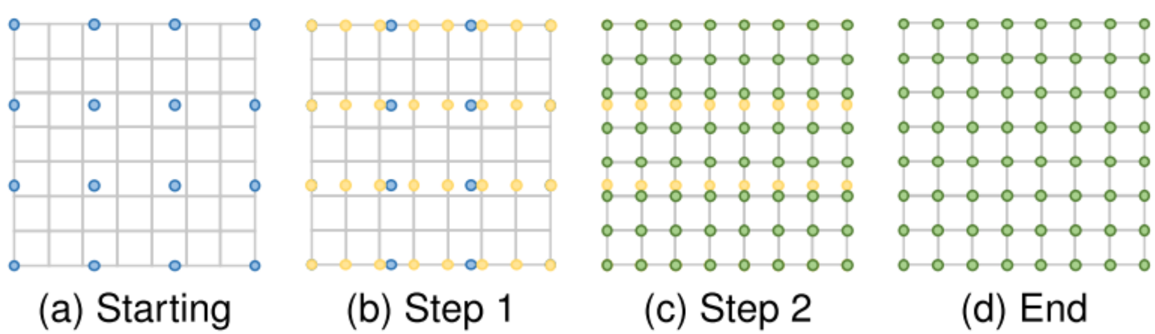
Supervised Super-Resolution - Upsampling Methods

Supervised Super-Resolution

- Upsampling methods
  - Interpolation-based upsampling
    - Nearest-neighbor interpolation
    - Bilinear interpolation
    - Bicubic interpolation
  - Learning-based upsampling
    - Transposed convolution layer
    - Sub-pixel layer
    - Meta upscale module

Supervised Super-Resolution

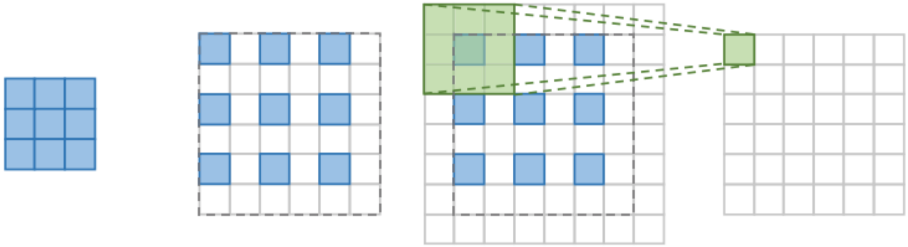
- Upsampling methods
  - Interpolation-based upsampling
    - Nearest-neighbor interpolation, Bilinear interpolation, Bicubic interpolation
    - The interpolation-based upsampling methods improve the image resolution only based on its own image signals, without bringing any more information.
    - Instead, they often introduce some side effects.
      - computational complexity, noise amplification, blurring results
    - Therefore, the current trend is to replace the interpolation-based methods with learnable upsampling layers.



Bicubic interpolation

Supervised Super-Resolution

- Upsampling methods
  - Learning-based upsampling
    - Transposed convolution layer (Deconvolution layer)
      - It increases the image resolution by expanding the image by inserting zeros and performing convolution.
      - It enlarges the image size in an end-to-end manner while maintaining a connectivity pattern compatible with vanilla convolution.
      - However, it can easily create a checkerboard-like pattern of varying magnitudes and thus hurt the SR performance.
      - F-SRCNN*, SRDenseNet**, DBPN***



(a) Starting (b) Expanding (c) Convolution

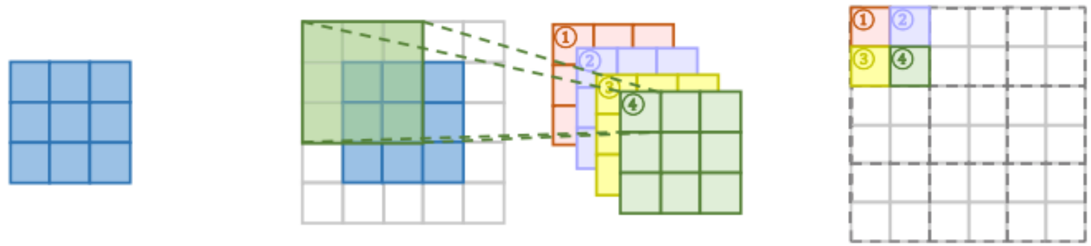
Transposed convolution layer

Deconvolution and checkerboard artifacts****

* C. Dong *et al.*, "Accelerating the super-resolution convolutional neural network," in Proc. Eur. Conf. Comput. Vis., 2016, pp 391–407.
** T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 4809–4817.
*** M. Haris *et al.*, "Deep back-projection networks for super-resolution," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 1664–1673.
**** A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," Distill, 2016,

Supervised Super-Resolution

- Upsampling methods
  - Learning-based upsampling
    - Sub-pixel layer (PixelShuffle)
      - It performs upsampling by generating a plurality of channels by convolution and then reshaping them.
        - It has a larger receptive field, which provides more contextual information to help generate more realistic details.
      - ESPCN*, SRGAN**, CARN***, RDN****



(a) Starting (b) Convolution (c) Reshaping

Sub-pixel layer

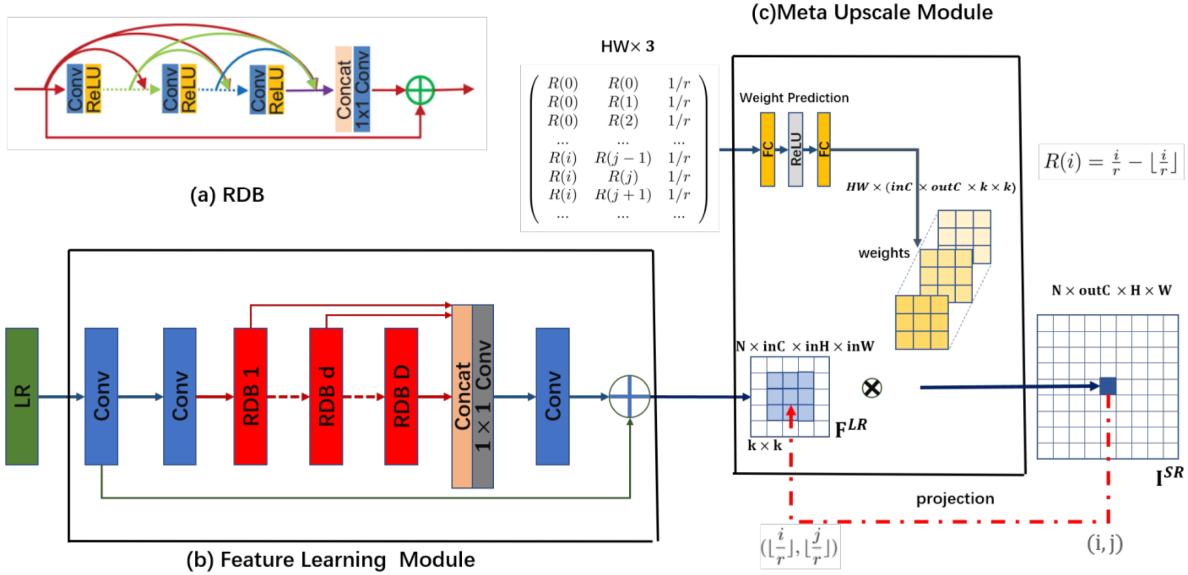
* C. Dong *et al.*, "Accelerating the super-resolution convolutional neural network," in Proc. Eur. Conf. Comput. Vis., 2016, pp 391–407.
** T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 4809–4817.
*** M. Haris *et al.*, "Deep back-projection networks for super-resolution," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 1664–1673.
**** H. Gao, H. Yuan, Z. Wang, and S. Ji, "Pixel transposed convolutional networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 5, pp. 1218–1227, May, 2020.

Supervised Super-Resolution

- Upsampling methods
  - Learning-based upsampling
    - Meta upscale module
      - The previous methods need to predefine the scaling factors.
        - Training different upsampling modules for different factors.
        - Inefficient and not in line with real needs.
          - e.g. SD, HD, FHD, etc. → UHD
      - It can continuously zoom in it with arbitrary factors by a single model.
      - Due to the large amount of training data (multiple factors are simultaneously trained), the module can exhibit comparable or even better performance on fixed factors.

Supervised Super-Resolution

- Upsampling methods
  - Learning-based upsampling
    - Meta upscale module
      - For each target position on the HR images, this module project it to a small patch on the LR feature maps (i.e., $k \times k \times c_{in}$), predicts convolution weights (i.e., $k \times k \times c_{in} \times c_{out}$) according to the projection offsets and the scaling factor by dense layers and perform convolution.



MetaSR*

* X. Hu, H. Mu, X. Zhang, Z. Wang, T. Tan, and J. Sun, "MetaSR: A magnification-arbitrary network for super-resolution," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2019, pp. 1575–1584.

Supervised Super-Resolution

- Upsampling methods
  - Learning-based upsampling
    - Meta upscale module
      - Although this module needs to predict weights during inference, the execution time of the upsampling module only accounts for about 1 percent of the time of feature extraction.
      - However, this method predicts a large number of convolution weights for each target pixel based on several values independent of the image contents.
        → the prediction result may be unstable and less efficient when faced with larger magnifications.

Supervised Super-Resolution

- Upsampling methods
  - Interpolation-based upsampling
    - Nearest-neighbor interpolation
    - Bilinear interpolation
    - Bicubic interpolation
  - Learning-based upsampling
    - Transposed convolution layer
    - Sub-pixel layer
    - Meta upscale module

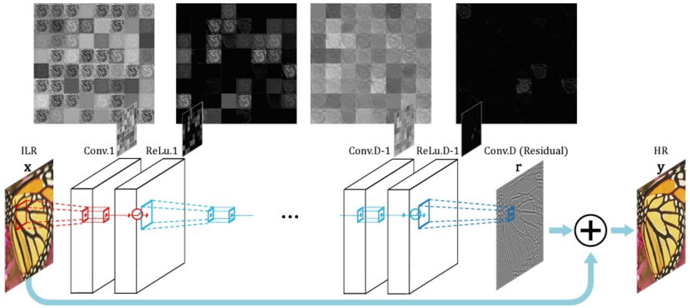
Supervised Super-Resolution - Network Design

Supervised Super-Resolution

- Network Design
  - Residual learning
  - Recursive learning
  - Multi-path learning
  - Dense connections
  - Attention mechanism
  - Advanced convolution
  - Wavelet transformation

Supervised Super-Resolution

- Network Design
  - Residual learning (skip connection)
    - Global residual learning
      - Since the image SR is an image-to-image translation task where the input image is highly correlated with the target image, researchers try to learn only a residual map to restore the missing high-frequency details.
        - Since the residuals in most regions are close to zero, the model complexity and learning difficulty are greatly reduced.
      - VDSR*, MemNet**, DRRN***, IDN****

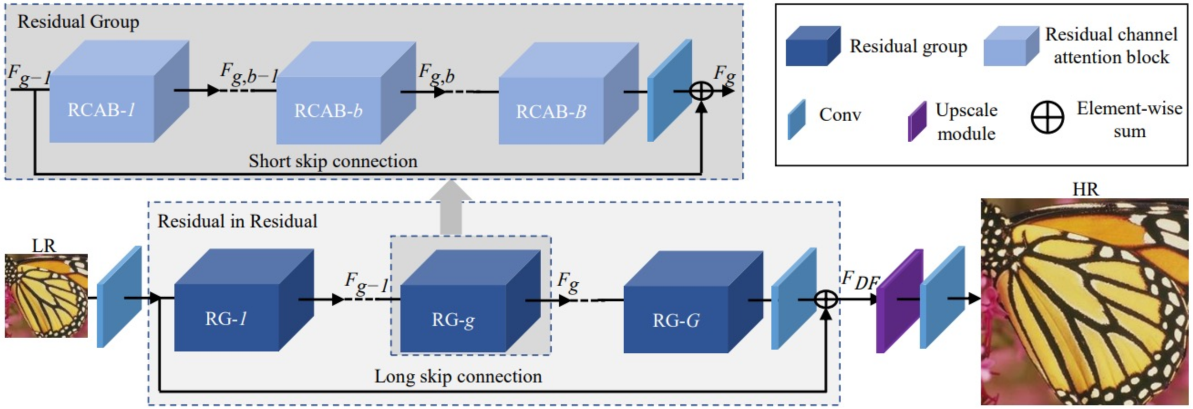


VDSR*

* J. Kim *et al.*, "Accurate image super-resolution using very deep convolutional networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 1646–1654.
** Y. Tai, J. Yang, X. Liu, and C. Xu, "Memnet: A persistent memory network for image restoration," in Proc. IEEE Int. Conf. Comput. Vis., 2017, 4539–4547.
*** Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 2790–2798.
**** Z. Hui *et al.*, "Fast and accurate single image super-resolution via information distillation network," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 723–731.

Supervised Super-Resolution

- Network Design
  - Residual learning (skip connection)
    - Local residual learning
      - The local residual learning is similar to the residual learning in ResNet and used to alleviate the degradation problem caused by ever increasing network depths, reduce training difficulty and improve the learning ability.
      - RCAN*, RED-Net**, DSRN***, MSRN****



RCAN*

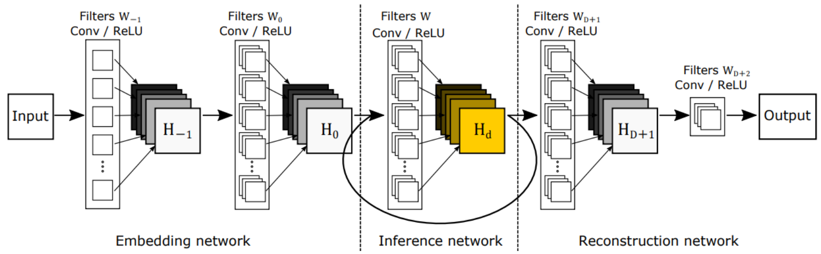
* Y. Zhang *et al.*, "Image super-resolution using very deep residual channel attention networks," in Proc. Eur. Conf. Comput. Vis., 2018, pp. 294–310.
** X. Mao *et al.*, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in Proc. 30th Int. Conf. Neural Inf. Process. Syst., 2016, pp. 2810–2818
*** W. Han *et al.*, "Image super-resolution via dual-state recurrent networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 1654–1663
**** J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in Proc. Eur. Conf. Comput. Vis., 2018, pp 527–542.

Supervised Super-Resolution

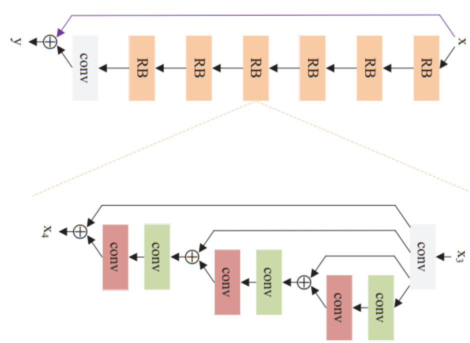
- Network Design
  - Recursive learning
    - In order to learn higher-level features without introducing overwhelming parameters,
      - applies the same modules multiple times in a recursive manner.
    - It can indeed learn more advanced representations without introducing excessive parameters, but still can't avoid high computational costs.
      - parameter efficient → computation efficient X
    - And it inherently brings vanishing or exploding gradient problems.

Supervised Super-Resolution

- Network Design
  - Recursive learning
    - DRCN*: 16× a single convolutional layer
      - reaches a receptive field of 41×41, which is much larger than 13×13 of SRCNN, without over many parameters.



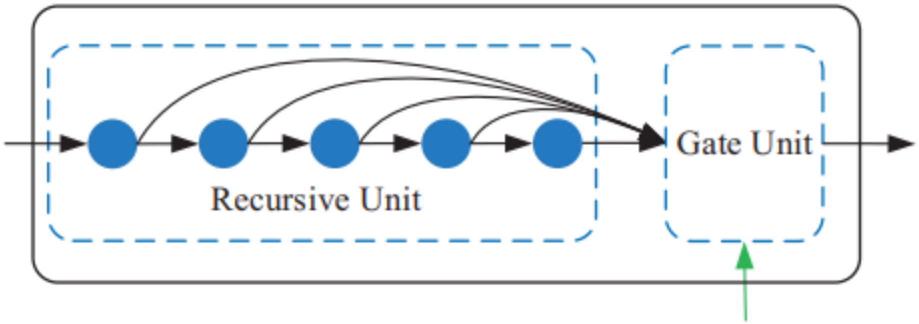
    - DRRN**: 25× a ResBlock
      - obtains even better performance than the 17-ResBlock baseline.



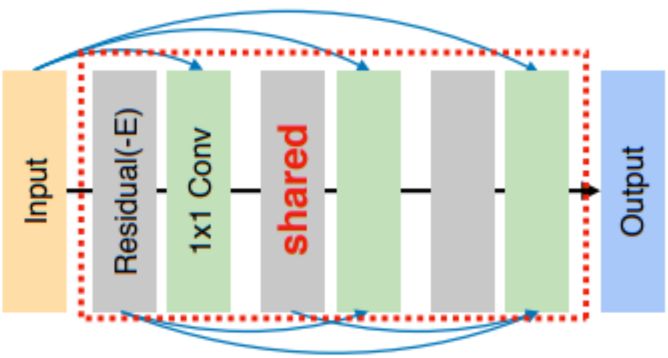
* J. Kim *et al.*, "Deeply-recursive convolutional network for image super-resolution," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 1637–1645.
** X. Mao *et al.*, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in Proc. 30th Int. Conf. Neural Inf. Process. Syst., 2016, pp. 2810–2818

Supervised Super-Resolution

- Network Design
  - Recursive learning
    - MemNet*: a 6-recursive ResBlock where the outputs of every recursion are concatenated and go through an extra 1×1 convolution for memorizing and forgetting.



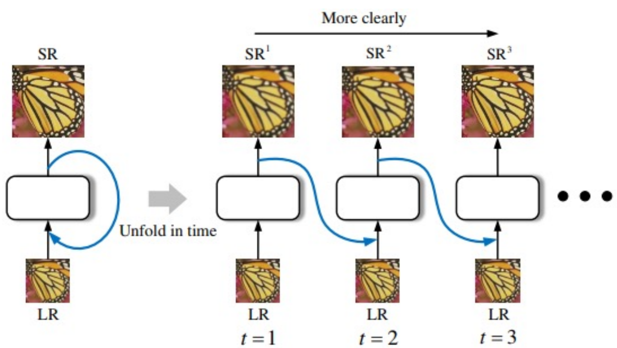
    - CARN**: a similar recursive unit including several ResBlocks.



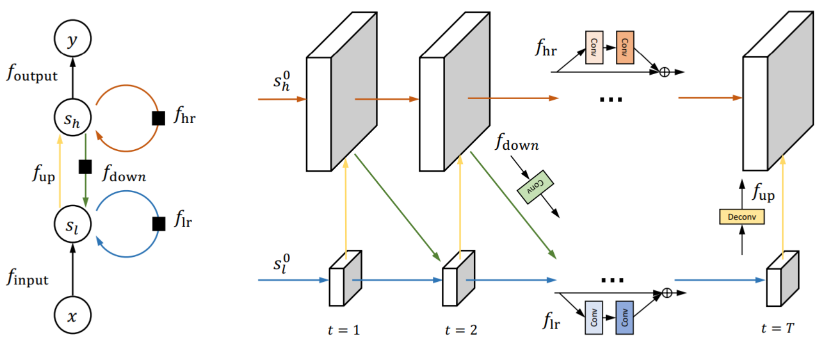
* Y. Tai, J. Yang, X. Liu, and C. Xu, "Memnet: A persistent memory network for image restoration," in Proc. IEEE Int. Conf. Comput. Vis., 2017, 4539–4547.
** N. Ahn, B. Kang, and K.-A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," in Proc. Eur. Conf. Comput. Vis., 2018, pp. 256–272.

Supervised Super-Resolution

- Network Design
  - Recursive learning
    - SRFBN*: a feedback network based on recursive learning, where the weights of the entire network are shared across all recursions.



    - DSRN**: to exchange signals between the LR and HR states. At each time step (i.e.,recursion), the representations of each branch are updated and exchanged for better exploring LR-HR relationships.

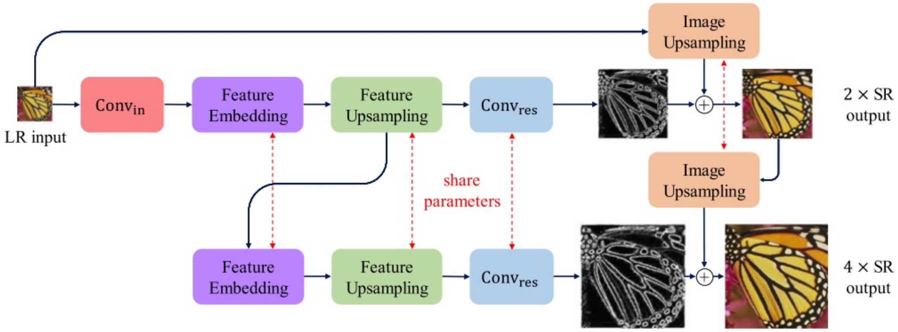


(a) DSRN (b) Unfolding DSRN T times

* Z. Li *et al.*, "Feedback network for image super-resolution," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2019, pp. 3862–3871.
** W. Han *et al.*, "Image super-resolution via dual-state recurrent networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 1654–1663.

Supervised Super-Resolution

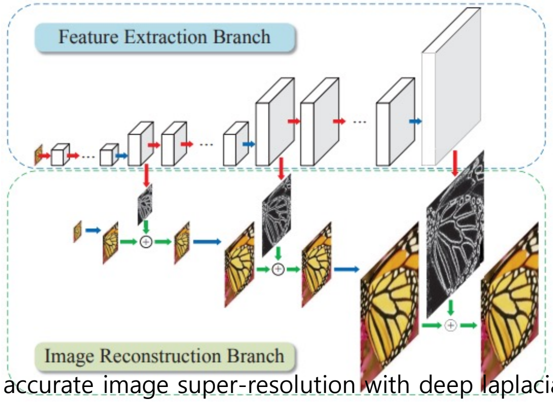
- Network Design
  - Recursive learning
    - LapSRN*: employ the embedding and upsampling modules as recursive units, and thus much reduce the model size at the expense of little performance loss.



* W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Fast and accurate image super-resolution with deep laplacian pyramid networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 11, pp. 2599–2613, Nov. 2018.

Supervised Super-Resolution

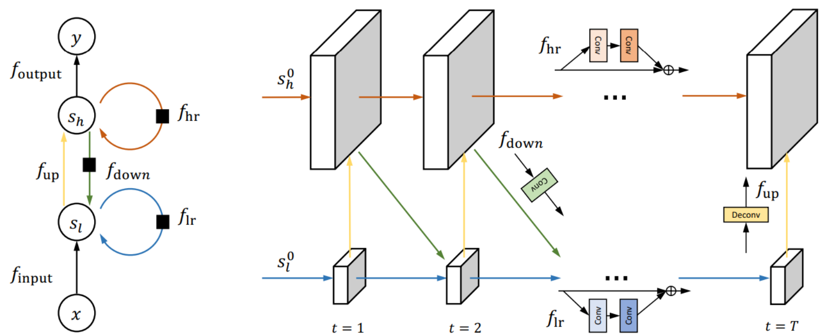
- Network Design
  - Multi-Path learning
    - Global Multi-Path Learning
      - Multiple paths to extract features of different aspects of the images.
      - These paths can cross each other in the propagation and thus greatly enhance the learning ability.
      - LapSRN*
        - A feature extraction path predicting the sub-band residuals in a coarse-to-fine fashion
        - A reconstruction path HR images based on the signals from both paths.



* W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Fast and accurate image super-resolution with deep laplacian pyramid networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 11, pp. 2599–2613, Nov. 2018.

Supervised Super-Resolution

- Network Design
  - Multi-Path learning
    - Global Multi-Path Learning
      - DSRN* utilizes two paths to extract information in LR and HR space, respectively, and continuously exchanges information for further improving learning ability.

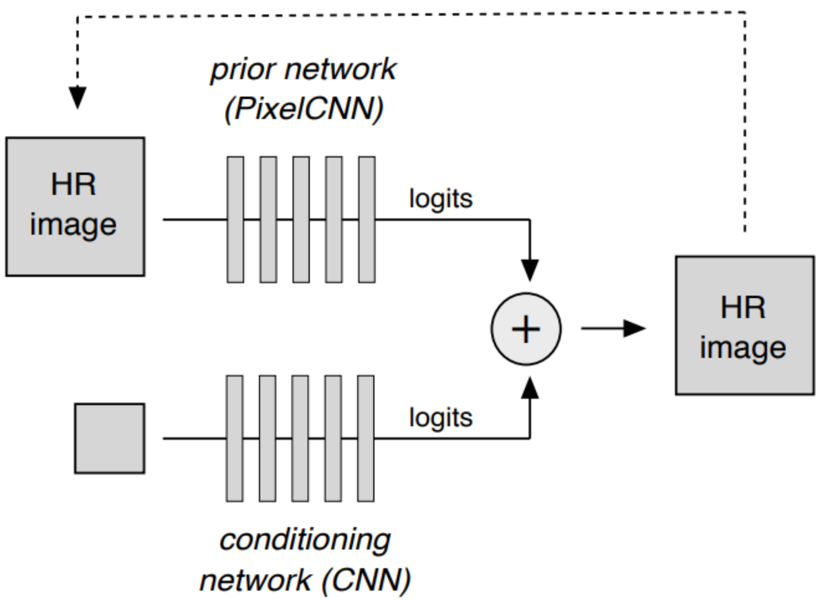


(a) DSRN (b) Unfolding DSRN T times

* W. Han *et al.*, "Image super-resolution via dual-state recurrent networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 1654–1663.

Supervised Super-Resolution

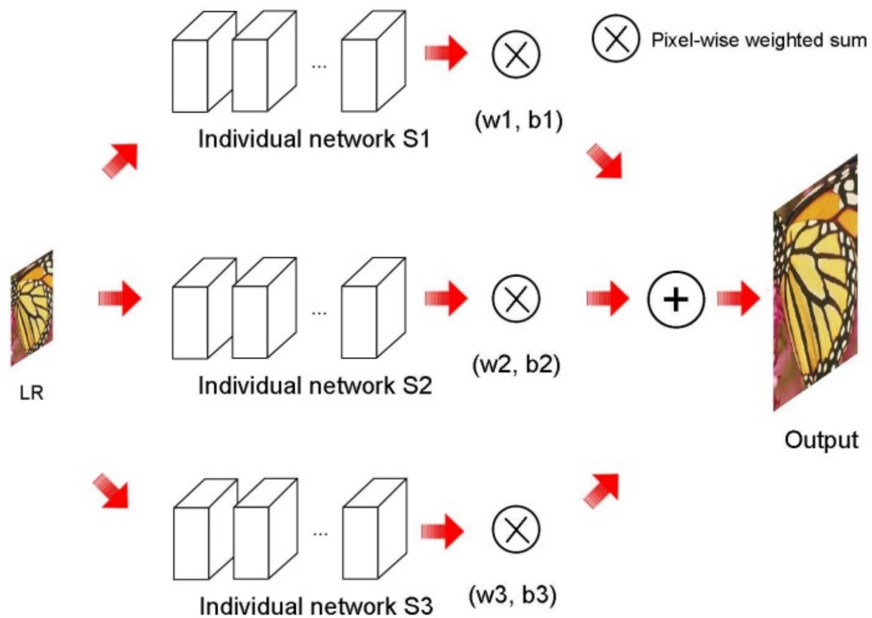
- Network Design
  - Multi-Path learning
    - Global Multi-Path Learning
      - PixelCNN* adopts a conditioning path to capture the global structure of images, and a prior path to capture the serial dependence of generated pixels.



* R. Dahl, M. Norouzi, and J. Shlens, "Pixel recursive super resolution," in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 5449–5458

Supervised Super-Resolution

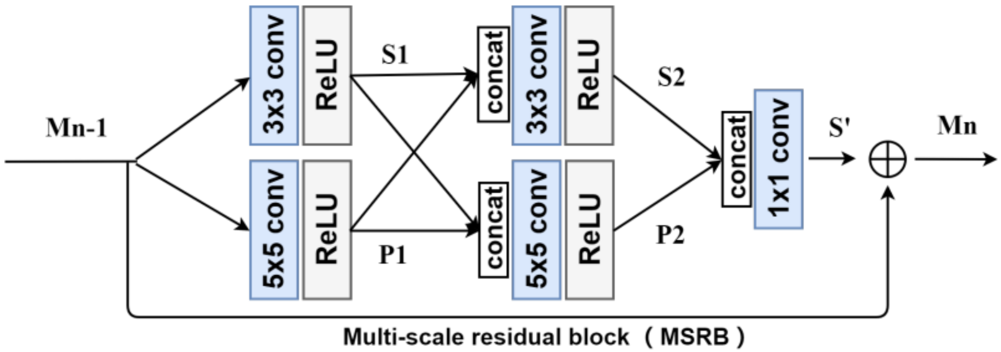
- Network Design
  - Multi-Path learning
    - Global Multi-Path Learning
      - CNF* employs multiple paths with unbalanced structures to perform upsampling and fuse them at the end of the model.



* H. Ren, M. El-Khamy, and J. Lee, "Image super resolution based on fusing multiple convolution neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2017, pp. 1050–1057.

Supervised Super-Resolution

- Network Design
  - Multi-Path learning
    - Local Multi-Path Learning
      - MSRN * adopts a new block for multiscale feature extraction inspired by the inception module.

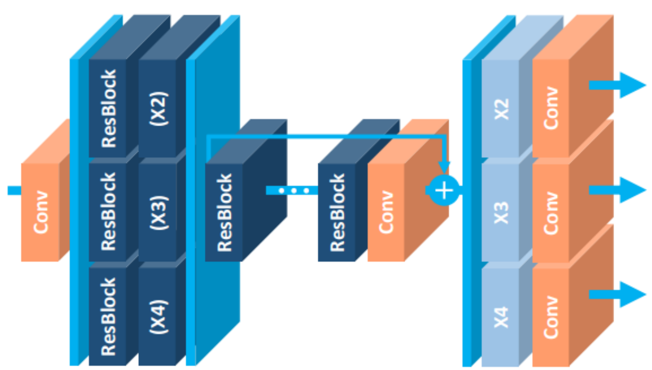


* J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in Proc. Eur. Conf. Comput. Vis., 2018, pp 527–542.

Supervised Super-Resolution

- Network Design
  - Multi-Path learning
    - Scale-Specific Multi-Path Learning
      - MDSR*
        - The principal components of the model are shared.
          - i.e., the intermediate layers for feature extraction
        - Scale-specific pre-processing paths and upsampling paths at the beginning and the end of the network.

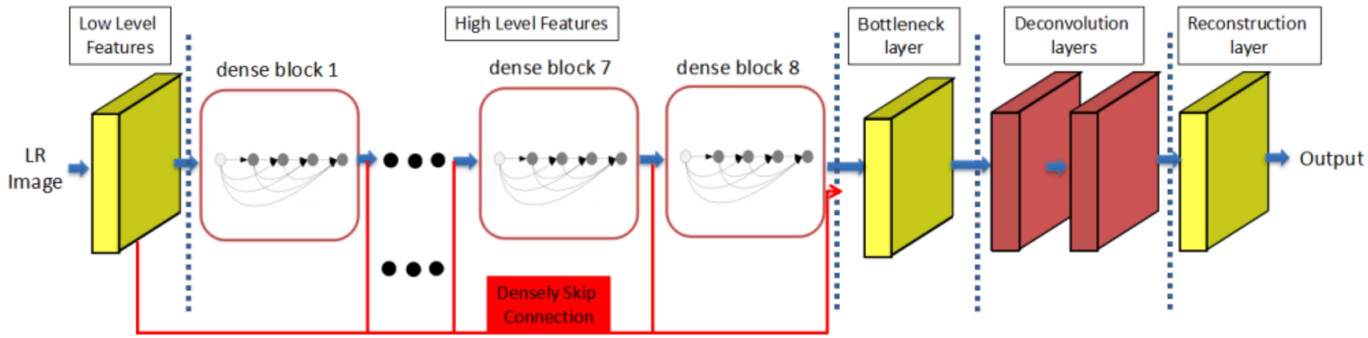
        → model size reduction with comparable performance
      - The similar approach is also adopted by CARN and ProSR.



* B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2017, pp. 1132–1140.

Supervised Super-Resolution

- Network Design
  - Dense Connections
    - For the sake of fusing low-level and high-level features to provide richer information, dense connections are introduced into the SR field.
    - SRDenseNet* not only adopt dense blocks to construct a 69-layer SRDenseNet, but also insert dense connections between different dense blocks.

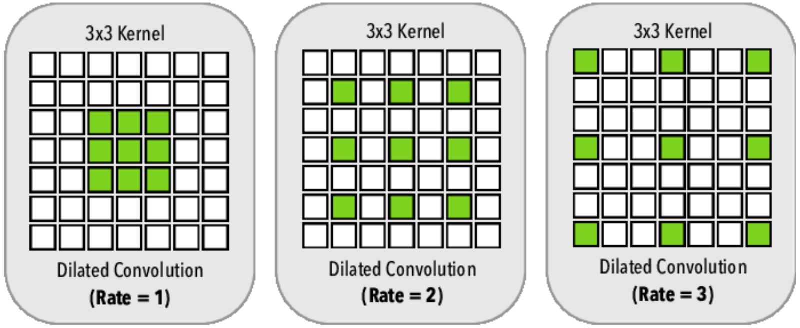


    - These layer-level and block-level dense connections are also adopted by MemNet, CARN, RDN, ESRGAN, and DBPN.

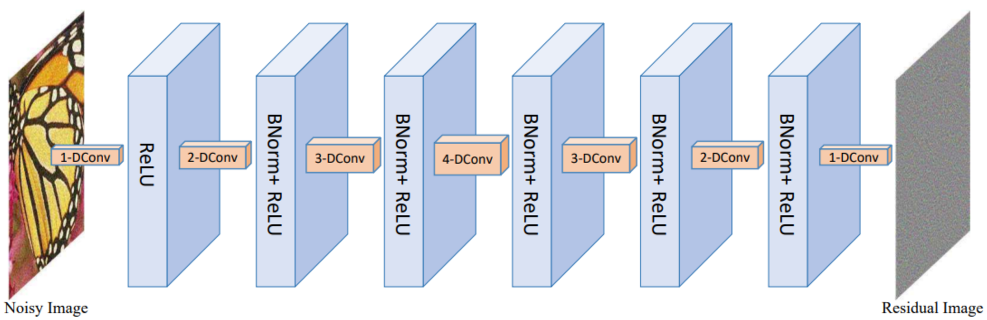
* T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 4809–4817.

Supervised Super-Resolution

- Network Design
  - Advanced convolution
    - Dilated convolution



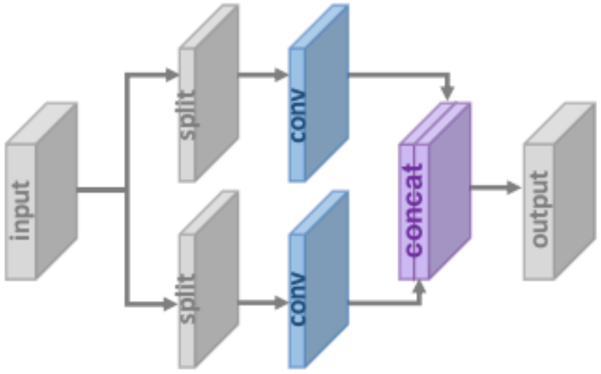
      - Zhang et al.* replace the common convolution by dilated convolution in SR models, increase the receptive field over twice and achieve much better performance.
        - s-DConv: s-dilated convolution



* K. Zhang *et al.*, "Learning deep CNN denoiser prior for image restoration," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 2808–2817.

Supervised Super-Resolution

- Network Design
  - Advanced convolution
    - Group convolution
      - Motivated by recent advances on lightweight CNNs, IDN* and CARN-M** replace the vanilla convolution by group convolution.
        - The group convolution much reduces the number of parameters and operations at the expense of a little performance loss.

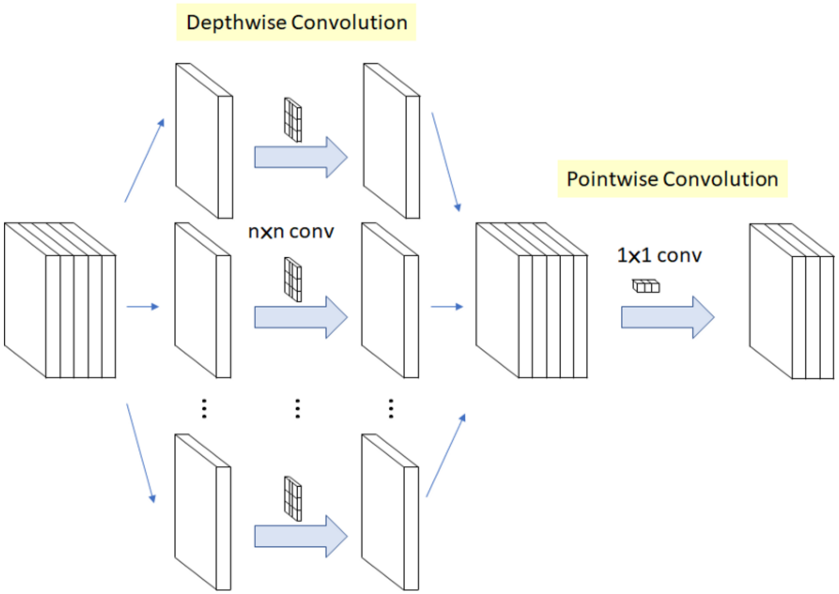


(g) Group convolution

* Z. Hui, X. Wang, and X. Gao, "Fast and accurate single image super-resolution via information distillation network," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 723–731.
** N. Ahn, B. Kang, and K.-A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," in Proc. Eur. Conf. Comput. Vis., 2018, pp. 256–272

Supervised Super-Resolution

- Network Design
  - Advanced convolution
    - Depthwise separable convolution
      - It consists of a factorized depthwise convolution and a pointwise convolution (i.e., 1×1 convolution), and thus reduces plenty of parameters and operations at only a small reduction in accuracy.

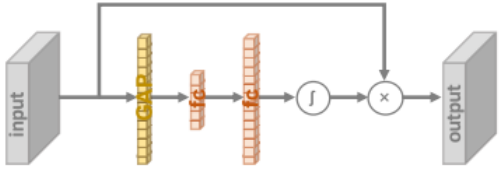


      - Nie et al. * employ the depthwise separable convolution and much accelerate the SR architecture.

* A. Ignatov et al., "PIRM challenge on perceptual image enhancement on smartphones: Report," in Proc. Eur. Conf. Comput. Vis., 2018, pp 315–333.

Supervised Super-Resolution

- Network Design
  - Attention Mechanism
    - Channel attention
      - Each input channel is squeezed into a channel descriptor (i.e., a constant) using global average pooling (GAP), then these descriptors are fed into two dense layers to produce channel-wise scaling factors for input channels.



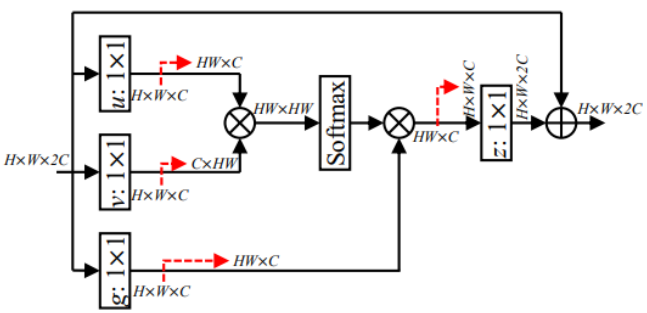
(c) Channel attention

      - RCAN* incorporates the channel attention mechanism, which markedly improves the SR performance.
      - SAN** further propose a second-order channel attention (SOCA) module to better learn the feature correlations.
        - The SOCA adaptively rescales the channel-wise features by using second-order feature statistics instead of GAP.

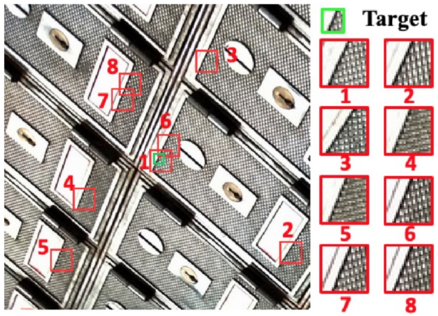
* Y. Zhang *et al.*, "Image super-resolution using very deep residual channel attention networks," in Proc. Eur. Conf. Comput. Vis., 2018, pp. 294–310.
** T. Dai *et al.*, "Second-order attention network for single image super-resolution," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2019, pp. 11057–11066.

Supervised Super-Resolution

- Network Design
  - Attention Mechanism
    - Non-local attention
      - Most existing SR models have very limited local receptive fields. However, some distant objects or textures may be very important for local patch generation.



Non-local attention



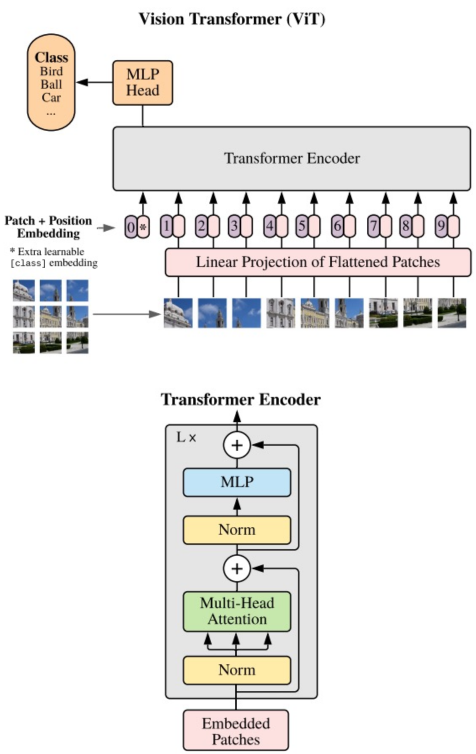
NL attention for SR***

      - RNAN* propose local and non-local attention blocks to extract features that capture the long-range dependencies between pixels.
      - SAN** also incorporate the non-local attention mechanism to capture long-distance spatial contextual information.

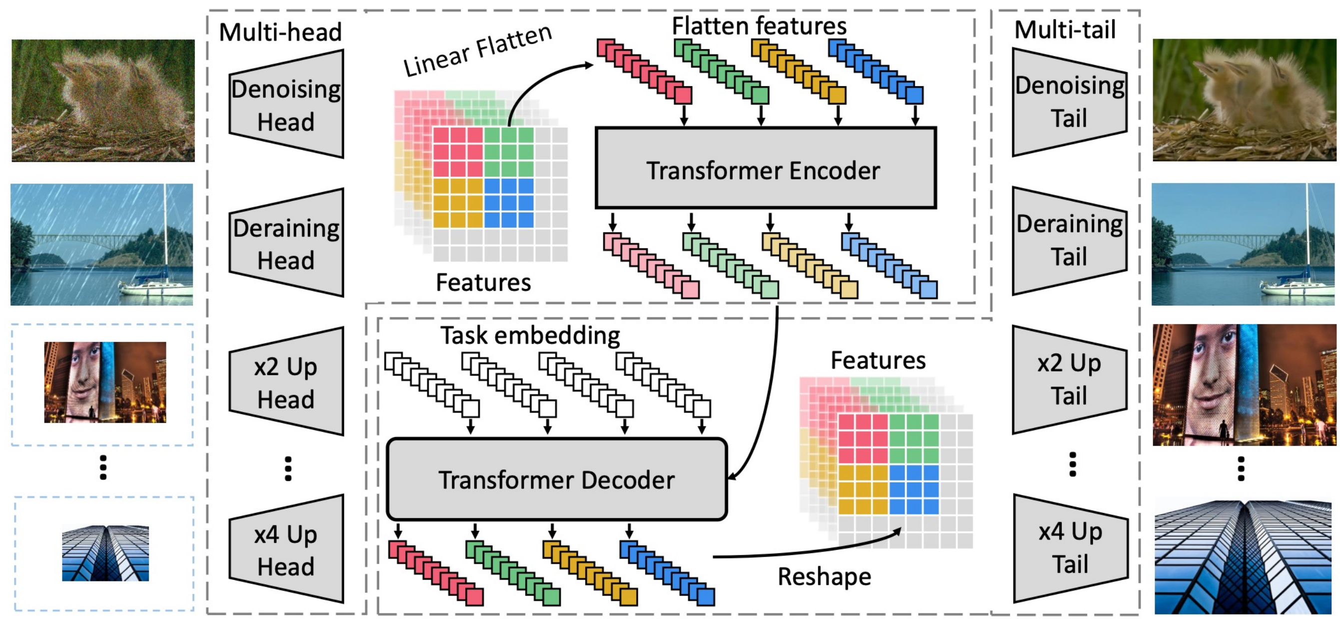
* Y. Zhang, K. Li, K. Li, B. Zhong, and Y. Fu, "Residual non-local attention networks for image restoration," in Proc. Int. Conf. Learn. Representations, 2019
** T. Dai *et al.*, "Second-order attention network for single image super-resolution," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2019, pp. 11057–11066.
*** Y. Mei et al., "Image Super-Resolution with Cross-Scale Non-Local Attention and Exhaustive Self-Exemplars Mining," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2020.

Supervised Super-Resolution

- Network Design
  - Attention Mechanism
    - IPT* (Pre-Trained Image Processing Transformer)



Vision Transformer

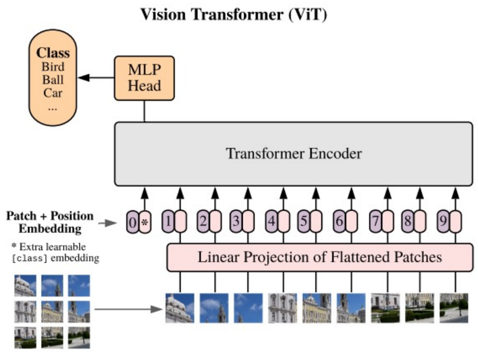


IPT architecture

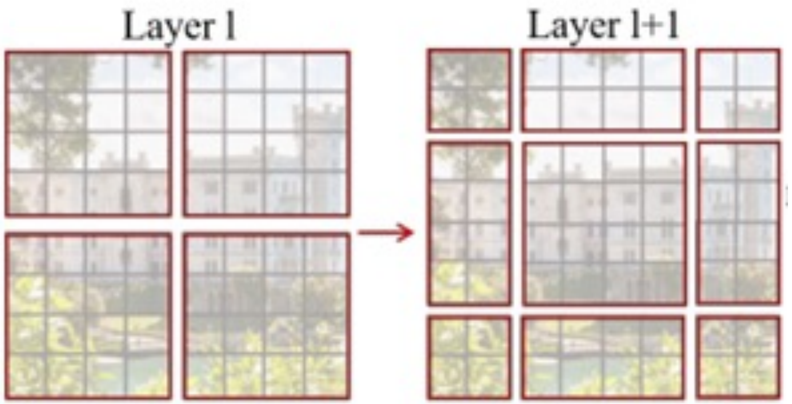
* J. Liang, J. Cao, G. Sun, K. Zhang, L. V. Gool, and R. Timofte, "SwinIR: Image Restoration Using Swin Transformer," in Proc. Int. Conf. Comput. Vis. Workshops, 2021, pp. 1833-1844.

Supervised Super-Resolution

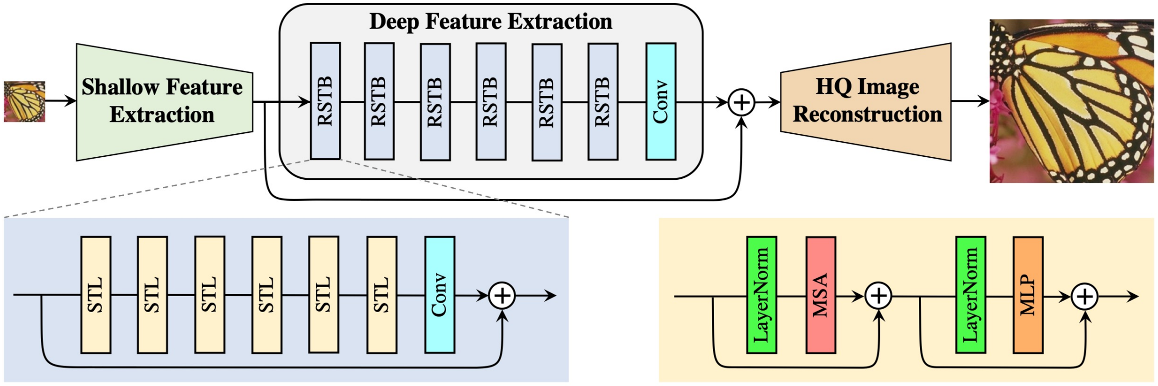
- Network Design
  - Attention Mechanism
    - SwinIR* (Image Restoration Using Swin Transformer)



Vision Transformer



Shifted window based self-attention



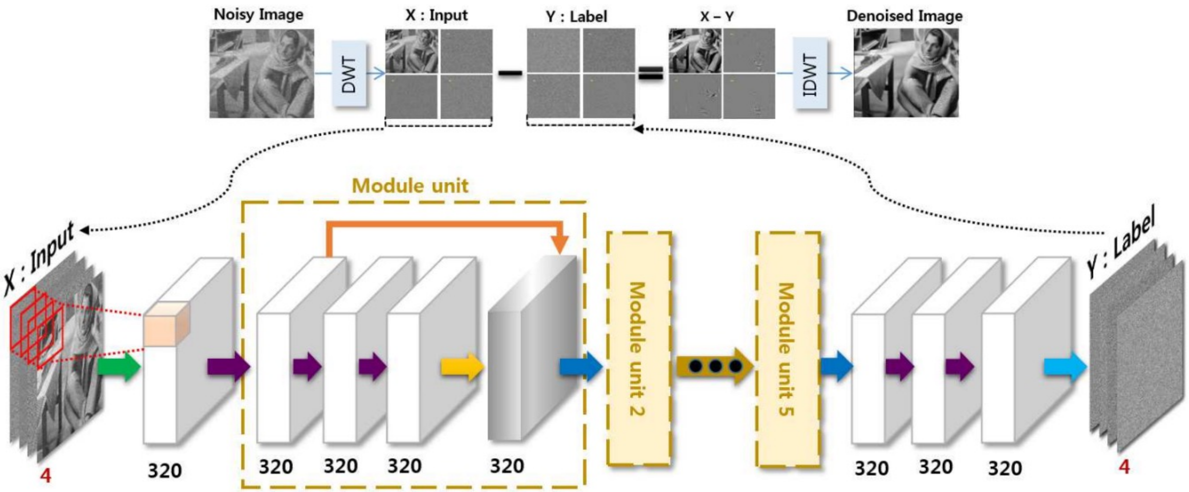
(a) Residual Swin Transformer Block (RSTB)

(b) Swin Transformer Layer (STL)

* J. Liang, J. Cao, G. Sun, K. Zhang, L. V. Gool, and R. Timofte, "SwinIR: Image Restoration Using Swin Transformer," in Proc. Int. Conf. Comput. Vis. Workshops, 2021, pp. 1833-1844.

Supervised Super-Resolution

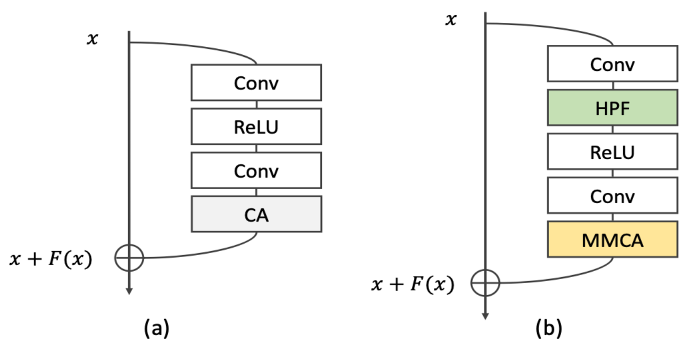
- Network Design
  - Wavelet Transformation
    - The wavelet transformation (WT) is a highly efficient representation of images by decomposing the image signal into high-frequency subbands and low-frequency subbands.
    - Bae et al.* first combine WT with deep learning based SR model, take sub-bands of interpolated LR wavelet as input and predict residuals of corresponding HR sub-bands.
      - WT and inverse WT are applied for decomposing the LR input and reconstructing the HR output, respectively.



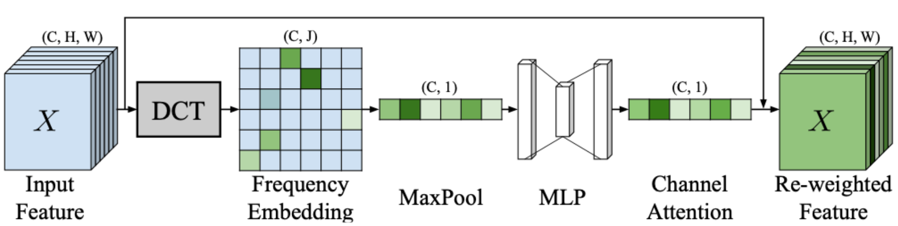
* W. Bae, J. J. Yoo, and J. C. Ye, "Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2017, pp. 1141–1149.

Supervised Super-Resolution

- Network Design
  - Wavelet Transformation
    - The DWSR*, Wavelet-SRNet**, and MWCNN*** also perform SR in the wavelet domain.
    - Due to the efficient representation by wavelet transformation, the models using this strategy often much reduce the model size and computational cost, while maintain competitive performance.
  - Discrete Cosine Transform****



Comparison to RCAN



MMCA

* T. Guo, H. S. Mousavi, T. H. Vu, and V. Monga, "Deep wavelet prediction for image super-resolution," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2017, pp. 1100–1109.
** H. Huang et al., "Wavelet-SRnet: A Wavelet-based CNN for multi-scale face super resolution," in Pro. IEEE Int. Conf. Comput. Vis., 2017, pp. 1698–1706.
*** P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo, "Multi-level Wavelet-CNN for image restoration," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, 2018, pp. 886–88609.
**** S. A. Magid et al., "Dynamic High-Pass Filtering and Multi-Spectral Attention for Image Super-Resolution," in Pro. IEEE Int. Conf. Comput. Vis., 2021.

Supervised Super-Resolution - Learning Strategies

Supervised Super-Resolution

- Learning strategies
  - Loss functions
  - Batch normalization
  - Curriculum learning
  - Multi-supervision

Supervised Super-Resolution

- Learning strategies
  - Loss functions
    - Pixel loss
      - Pixel-wise difference between two images
      - L1 loss (i.e., mean absolute error), L2 loss (i.e., mean square error)

$$\mathcal{L}_{\text{pixel\_l1}}(\hat{I}, I) = \frac{1}{hwc}\sum_{i,j,k} |\hat{I}_{i,j,k} - I_{i,j,k}| \qquad \mathcal{L}_{\text{pixel\_l2}}(\hat{I}, I) = \frac{1}{hwc}\sum_{i,j,k} (\hat{I}_{i,j,k} - I_{i,j,k})^2,$$

      - L2 loss penalizes larger errors but is more tolerant to small errors → smooth results.
        - In practice, the L1 loss shows improved performance and convergence over L2 loss.
        - Since the definition of PSNR is highly correlated with pixel-wise difference and minimizing pixel loss directly maximize PSNR.
      - However, since the pixel loss actually doesn’t take image quality (e.g., perceptual quality, textures) into account, the results often lack high-frequency details and are perceptually unsatisfying with oversmooth textures.

Supervised Super-Resolution

- Learning strategies
  - Loss functions
    - Content loss
      - The semantic differences between images using a pre-trained image classification network.

$$\mathcal{L}_{\text{content}}(\hat{I}, I; \phi, l) = \frac{1}{h_l w_l c_l} \sqrt{\sum_{i,j,k} (\phi^{(l)}_{i,j,k}(\hat{I}) - \phi^{(l)}_{i,j,k}(I))^2}$$

      - The content loss encourages the output image to be perceptually similar to the target image instead of forcing them to match pixels exactly. → visually more perceptible results.

Supervised Super-Resolution

- Learning strategies
  - Loss functions
    - Adversarial loss
      - In the GAN framework, we only need to treat the SR model as a generator, and define an extra discriminator to judge whether the input image is generated or not.

$$\mathcal{L}_{\text{gan\_ce\_g}}(\hat{I}; D) = -\log D(\hat{I})$$
$$\mathcal{L}_{\text{gan\_ce\_d}}(\hat{I}, I_s; D) = -\log D(I_s) - \log(1 - D(\hat{I}))$$
$$\mathcal{L}_{\text{gan\_ls\_g}}(\hat{I}; D) = (D(\hat{I}) - 1)^2$$
$$\mathcal{L}_{\text{gan\_ls\_d}}(\hat{I}, I_s; D) = (D(\hat{I}))^2 + (D(I_s) - 1)^2$$

      - Even though the SR models trained with adversarial loss and content loss achieve lower PSNR compared to those trained with pixel loss, they bring significant gains in perceptual quality.

Supervised Super-Resolution

- Learning strategies
  - Loss functions
    - Cycle consistency loss
      - They not only super-resolve the LR image I to the HR image $\hat{I}$ but also downsample $\hat{I}$ back to another LR image I' through another CNN.

$$\mathcal{L}_{\text{cycle}}(I', I) = \frac{1}{hwc}\sqrt{\sum_{i,j,k}(I'_{i,j,k} - I_{i,j,k})^2}$$

→ Unpaired super-resolution

Supervised Super-Resolution

- Learning strategies
  - Loss functions
    - Total variation loss
      - To suppress noise in generated images
      - The sum of the absolute differences between neighboring pixels and measures how much noise is in the images.

$$\mathcal{L}_{\mathrm{TV}}(\hat{I}) = \frac{1}{hwc} \sum_{i,j,k} \sqrt{(\hat{I}_{i,j+1,k} - \hat{I}_{i,j,k})^2 + (\hat{I}_{i+1,j,k} - \hat{I}_{i,j,k})^2}$$

Supervised Super-Resolution

- Learning strategies
  - Batch normalization
    - Lim et al.* (EDSR) argue that the BN loses the scale information of each image and gets rid of range flexibility from networks.
    - So they remove BN and use the saved memory cost (up to 40 percent) to develop a much larger model, and thus increase the performance substantially.

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1\ldots m}\}$;
Parameters to be learned: γ, β
Output: $\{y_i = \mathrm{BN}_{\gamma,\beta}(x_i)\}$

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{// mini-batch mean}$$

$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_{\mathcal{B}})^2 \qquad \text{// mini-batch variance}$$

$$\widehat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \qquad \text{// normalize}$$

$$y_i \leftarrow \gamma \widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \qquad \text{// scale and shift}$$

Algorithm 1: Batch Normalizing Transform, applied to activation x over a mini-batch.

* B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2017, pp. 1132–1140.

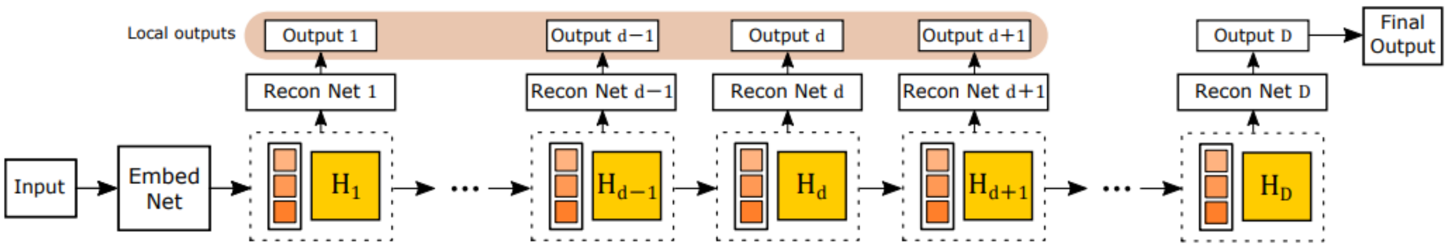
Supervised Super-Resolution

- Learning strategies
  - Curriculum learning
    - $\times 2 \rightarrow \times 4 \rightarrow \ldots$
      - In order to reduce the difficulty of SR with large scaling factors ProSR*, ADRSR**, progressive CARN*** which are progressive not only on architectures but also on training procedure.
    - Easy degradation → more complex degradation
      - SRFBN****

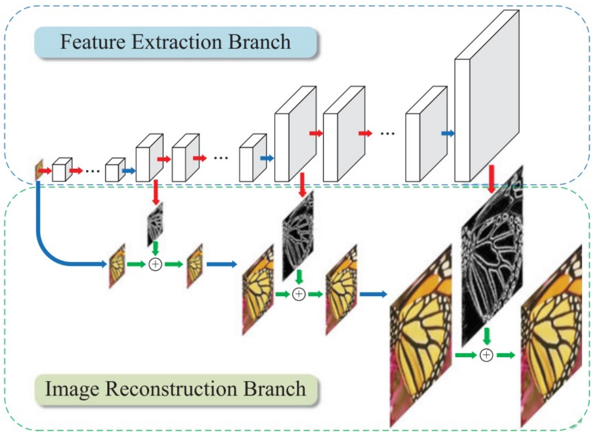
* Y. Wang, F. Perazzi, B. McWilliams, A. Sorkine-Hornung, O. Sorkine-Hornung, and C. Schroers, "A fully progressive approach to single-image super-resolution," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, 2018, pp. 977–97709.
** Y. Bei, A. Damian, S. Hu, S. Menon, N. Ravi, and C. Rudin, "New techniques for preserving global structure and denoising with low information loss in single-image super-resolution," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, 2018, pp. 987–9877.
*** N. Ahn, B. Kang, and K.-A. Sohn, "Image super-resolution via progressive cascading residual network," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, 2018. pp, 904–912.
**** Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2019, pp. 3862–3871.

Supervised Super-Resolution

- Learning strategies
  - Multi-supervision
    - Multi-supervision with recursive units
      - DRCN*, MemNet**, DSRN***



    - Multi-supervision with intermediate results of different scales during propagation
      - LapSRN****



* J. Kim *et al.*, "Deeply-recursive convolutional network for image super-resolution," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 1637–1645.
** Y. Tai, J. Yang, X. Liu, and C. Xu, "Memnet: A persistent memory network for image restoration," in Proc. IEEE Int. Conf. Comput. Vis., 2017, 4539–4547.
*** W. Han *et al.*, "Image super-resolution via dual-state recurrent networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 1654–1663.
**** W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Fast and accurate image super-resolution with deep laplacian pyramid networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 11, pp. 2599–2613, Nov. 2018.

Supervised Super-Resolution - Others and SOTA

Supervised Super-Resolution

- Other improvements
  - Network interpolation
    - ESRGAN* trains a PSNR-based model and train a GAN-based model by fine-tuning, then interpolate all the corresponding parameters of both networks to derive intermediate models.

$$\theta_G^{\mathrm{INTERP}} = (1-\alpha)\ \theta_G^{\mathrm{PSNR}} + \alpha\ \theta_G^{\mathrm{GAN}}$$

* X. Wang et al., “Esrgan: Enhanced super-resolution generative adversarial networks,” in Proc. Eur. Conf. Comput. Vis. Workshop, 2018.

Supervised Super-Resolution

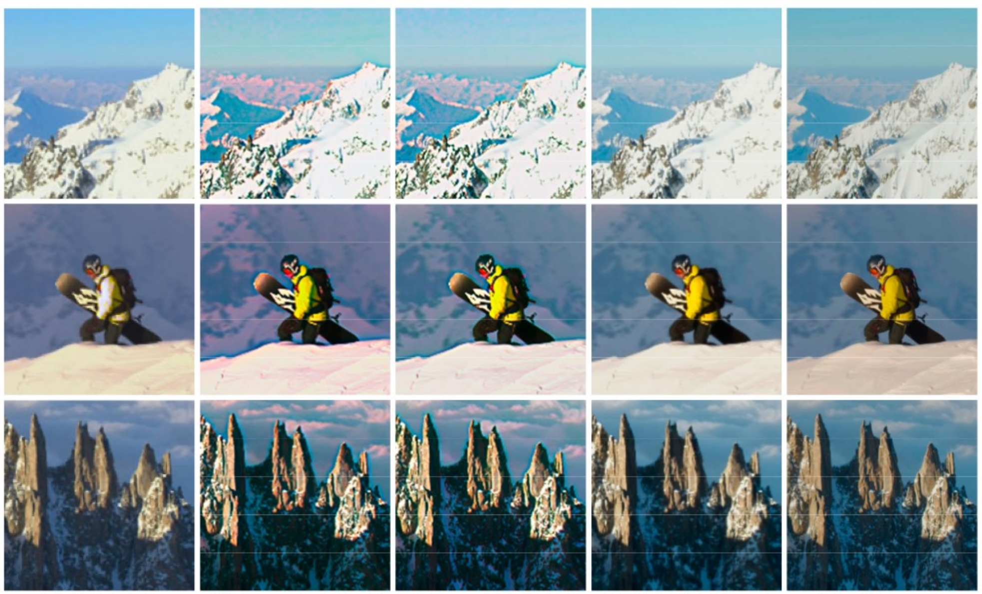
- Other improvements
  - Self-ensemble
    - Self-ensemble is an inference technique commonly used by SR models.
    - Rotations with different angles (0, 90, 180, 270) and horizontal flipping are applied on the LR images to get a set of 8 images.
    - Then these images are fed into the SR model and the corresponding inverse transformation is applied to the reconstructed HR images to get the outputs.

Bicubic	×4	28.42	0.8104	26.00	0.7027	25.96	0.6675	23.14	0.6577	24.89	0.7866
SRCNN [5]	×4	30.48	0.8628	27.50	0.7513	26.90	0.7101	24.52	0.7221	27.58	0.8555
FSRCNN [6]	×4	30.72	0.8660	27.61	0.7550	26.98	0.7150	24.62	0.7280	27.90	0.8610
VDSR [16]	×4	31.35	0.8830	28.02	0.7680	27.29	0.0726	25.18	0.7540	28.83	0.8870
LapSRN [19]	×4	31.54	0.8850	28.19	0.7720	27.32	0.7270	25.21	0.7560	29.09	0.8900
MemNet [35]	×4	31.74	0.8893	28.26	0.7723	27.40	0.7281	25.50	0.7630	29.42	0.8942
EDSR [23]	×4	32.46	0.8968	28.80	0.7876	27.71	0.7420	26.64	0.8033	31.02	0.9148
SRMDNF [43]	×4	31.96	0.8925	28.35	0.7787	27.49	0.7337	25.68	0.7731	30.09	0.9024
D-DBPN [10]	×4	32.47	0.8980	28.82	0.7860	27.72	0.7400	26.38	0.7946	30.91	0.9137
RDN [44]	×4	32.47	0.8990	28.81	0.7871	27.72	0.7419	26.61	0.8028	31.00	0.9151
RCAN (ours)	×4	32.63	0.9002	28.87	0.7889	27.77	0.7436	26.82	0.8087	31.22	0.9173
RCAN+ (ours)	×4	**32.73**	**0.9013**	**28.98**	**0.7910**	**27.85**	**0.7455**	**27.10**	**0.8142**	**31.65**	**0.9208**

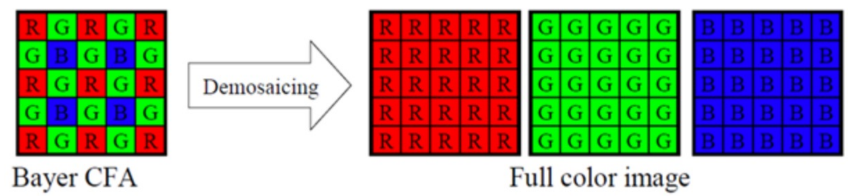
Self-ensemble

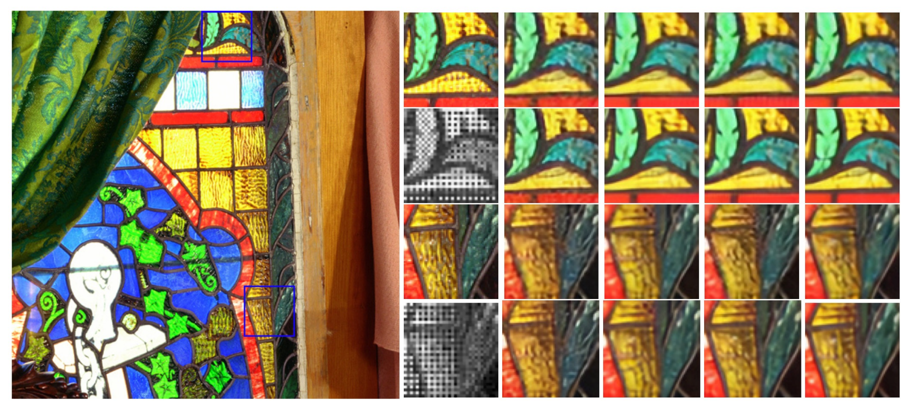
Supervised Super-Resolution

- Other improvements
  - Super-resolution + alpha
    - Inverse Tone-mapping (SDR -> HDR)
    - Denoising
    - Demosaicing



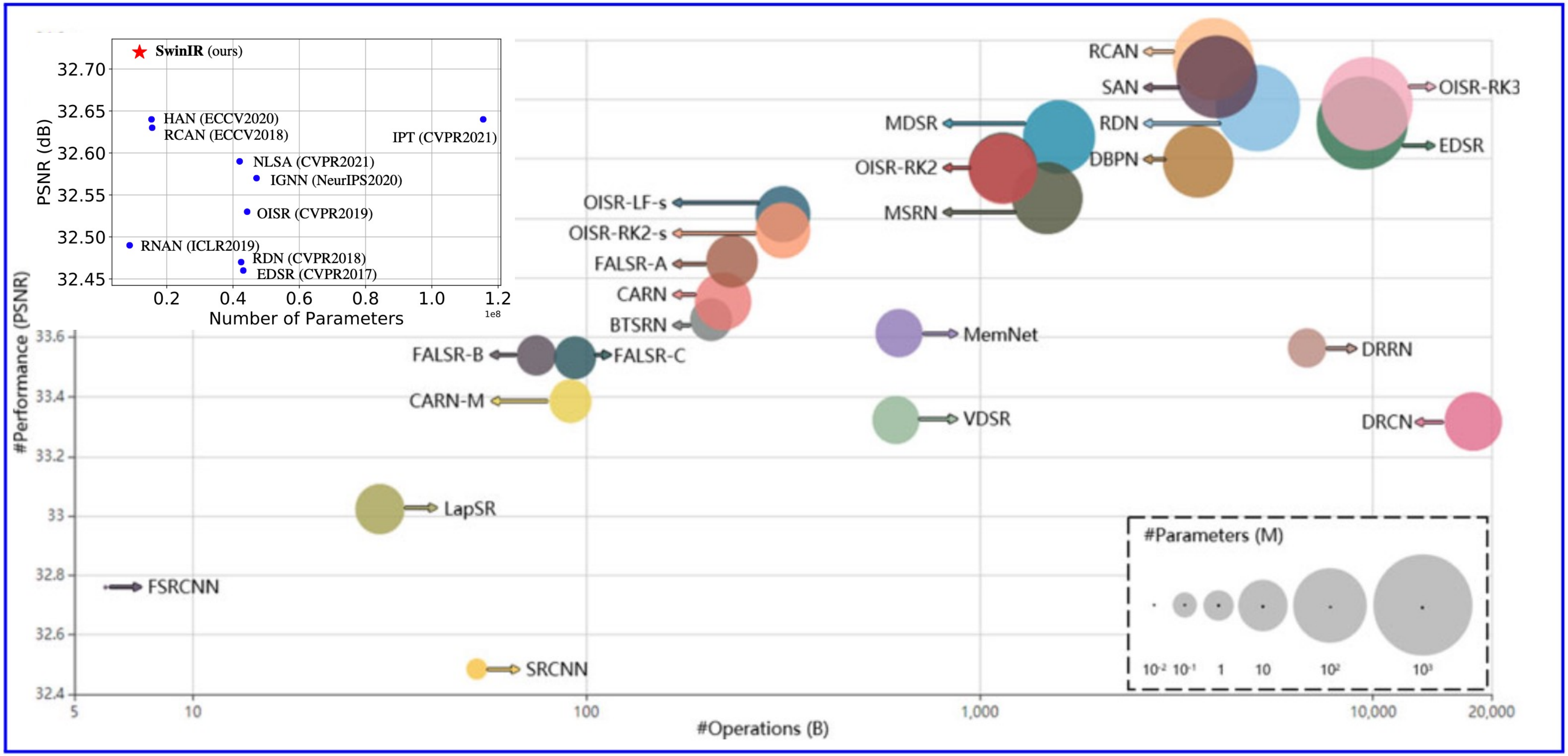
(a) LR-SDR (b) Deep SR-ITM (c) Deep SR-ITM+ (d) **FDAN (Ours)** (e) HR-HDR





Supervised Super-Resolution

- State-of-the-art super-resolution models



Exercise #1
Supervised SR

Exercise #1 – Supervised SR

- EDSR-pytorch 코드 실행 준비
- RCAN github https://github.com/yulunzhang/RCAN
  - EDSR github로 통합. pretrained model 다운로드.
- EDSR github https://github.com/sanghyun-son/EDSR-PyTorch

- conda create –n edsr python=3.8
- conda activate edsr
- pip install torch==1.12.0+cu102 torchvision==0.13.0+cu102 torchaudio==0.12.0 --extra-index-url https://download.pytorch.org/whl/cu102
- pip install matplotlib imageio tqdm scikit-image
- git clone https://github.com/sanghyun-son/EDSR-PyTorch
- cd EDSR-PyTorch

Exercise #1 – Supervised SR

- 데이터셋 다운로드
  - https://github.com/ChaofWang/Awesome-Super-Resolution/blob/master/dataset.md
  - Dataset 폴더에 아래 데이터셋 다운로드
    - Set5, Set14, BSD300, 91-Image
      - 91-Image는 링크 문제로 페이지 내의 SR_testing_datasets에서 다운로드 가능
  - Set5, Set14, Urban100, Manga109
    - 평가용으로 주로 사용
  - DIV2K
    - Supervised SR 학습에 가장 많이 사용
  - Real SR
    - Unsupervised SR 학습 및 평가에 가장 많이 사용

Exercise #1 – Supervised SR

- 1. SR pretrained 모델 활용 (RCAN)
  - Pre-trained 모델 다운로드
    - EDSR-PyTorch/models 아래에 다운로드
    - https://github.com/yulunzhang/RCAN에서 다운로드 링크 제공
      - 원내 클라우드 접속 허가 필요
  - cd EDSR-Pytorch/src
  - python main.py --template RCAN --data_test Demo --scale 4 --pre_train ../models/RCAN_BIX4.pt --test_only --save_results (--self_ensemble)
    - EDSR-PyTorch/test 폴더 내의 모든 영상에 대해서 SR 실행
      - --dir_demo 수정하여 폴더 위치 변경 가능
    - EDSR-PyTorch/experiment/test/results-Demo 내에 결과 저장

Exercise #1 – Supervised SR

- 2. 학습
  - 데이터셋 준비
    - DIV2K를 주로 사용하지만 용량 문제로 SR291 데이터셋으로 대체
    - SR291 (BSDS200 + T91)
      - BSDS200: BSDS300 Training set
      - T91: 91-Image
      - HR, LR_bicubic 폴더 생성
    - benchmark/Set5
      - HR, LR_bicubic 폴더 생성
    - HR 폴더에 원래 image들 넣고 make_lr.m (MATLAB) 실행
      - MATLAB이 없는 관계로 실습 X
      - 논문 성능을 위해서는 MATLAB 필수
        - MATLAB이 없을 경우, 대체 Python 코드가 있지만 MATLAB 권장
        - https://github.com/fatheral/matlab_imresize
      - 실제 사용을 위해서는 Python 사용 무관

Exercise #1 – Supervised SR

- 2. 학습
  - 데이터셋 폴더 생성
    - Dataset/SR291 폴더 이름을 A2001로 변경
  - src/data 내에 데이터셋 추가
    - sr291.py 복사 후 a2001.py로 이름 수정
      - SR291 대신 원하는 폴더명 사용 (예: A2001)
    - SR291 → A2001로 수정

```
from data import srdata
class SR291(srdata.SRData):
    def __init__(self, args, name='SR291', train=True, benchmark=False):
        super(SR291, self).__init__(args, name=name)
```

Exercise #1 – Supervised SR

- 2. 학습
  - cd EDSR-Pytorch/src
  - conda activate edsr

  - Training from scratch
    - python main.py --template RCAN --save RCAN_BIX2_G10R20P48 ₩
      --scale 2 --reset --save_results --patch_size 64 ₩
      --dir_data ../../Dataset --data_train A2001 --data_test Set5

  - Finetuning
    - python main.py --template RCAN --save RCAN_BIX2_G10R20P48 ₩
      --scale 2 --reset --save_results --patch_size 64 ₩
      --dir_data ../../Dataset --data_train A2001 --data_test Set5 ₩
      --pre_train ../models/RCAN_BIX2.pt

Exercise #1 – Supervised SR

- 3. model 추가
  - src/model 내에 모델 추가
  - 모델에 따라 필요시 options.py에 argument 추가
  - (실습) Post upsampling 모델 + Global residual learning구현
    - 5×5conv32 – ReLU – 3×3conv64 – ReLU – 3×3conv256 – PixelShuffle – 3×3conv3
    - torch.nn.functional.interpolate(x, scale_factor=2, mode='bicubic')
    - edsr.py 복사하여 a2001sr.py 생성
      - EDSR → A2001SR로 수정 후 모델 내용 작성

```
def make_model(args, parent=False):
return EDSR(args)

class EDSR(nn.Module):
def __init__(self, args, conv=common.default_conv):
super(EDSR, self).__init__()
```

    - python main.py --model A2001SR --save A2001_SR --scale 2 --reset ₩
      --save_results --patch_size 64 --dir_data ../../Dataset ₩
      --data_train A2001 --data_test Set5

Exercise #1 – Supervised SR

- 4. loss 추가
  - src/loss 내에 추가
  - (실습) Total Variation loss 구현

$$\mathcal{L}_{\mathrm{TV}}(\hat{I}) = \frac{1}{hwc}\sum_{i,j,k}\sqrt{(\hat{I}_{i,j+1,k}-\hat{I}_{i,j,k})^2+(\hat{I}_{i+1,j,k}-\hat{I}_{i,j,k})^2}$$

- src/loss/__init__.py 28번째 줄에 아래 내용 추가

```
elif loss_type.find('TV') >= 0:
module = import_module('loss.tv')
loss_function = getattr(module, 'TV’)()
```

- vgg.py 복사하여 tv.py 생성
  - VGG → TV로 수정 후 모델 내용 작성

```
class VGG(nn.Module):
def __init__(self~~, conv_index, rgb_range=1~~):
super(VGG, self).__init__()
```

Exercise #1 – Supervised SR

- 4. loss 추가
  - src/loss 내에 추가
  - (실습) Total Variation loss 구현

$$\mathcal{L}_{\mathrm{TV}}(\hat{I}) = \frac{1}{hwc}\sum_{i,j,k}\sqrt{(\hat{I}_{i,j+1,k} - \hat{I}_{i,j,k})^2 + (\hat{I}_{i+1,j,k} - \hat{I}_{i,j,k})^2}$$

  - dx = sr[:, :, :-1, :-1] – sr[:, :, :-1, 1:]
    (dy와 개수를 맞추기 위해 y index도 :-1로 설정)

  - python main.py --model A2001SR --save A2001_SR_TV --scale 2 ₩
    --reset --save_results --patch_size 64 --dir_data ../../Dataset ₩
    --data_train A2001 --data_test Set5 --loss 1*L1+0.001*TV

Real-World Super-Resolution

Real-World Super-Resolution

- Unsupervised super-resolution
  - Since it is difficult to collect images of the same scene but with different resolutions, the LR images in SR datasets are often obtained by performing predefined degradation on HR images.
  - Thus the trained SR models actually learn a reverse process of the predefined degradation.
  - In order to learn the real-world LR-HR mapping without introducing manual degradation priors, researchers pay more and more attention to unsupervised SR, in which case only unpaired LR-HR images are provided for training, so that the resulting models are more likely to cope with the SR problems in real-world scenarios.

Real-World Super-Resolution

- Unsupervised super-resolution
  - It is easy to obtain a large amount of data by downsampling HR images.

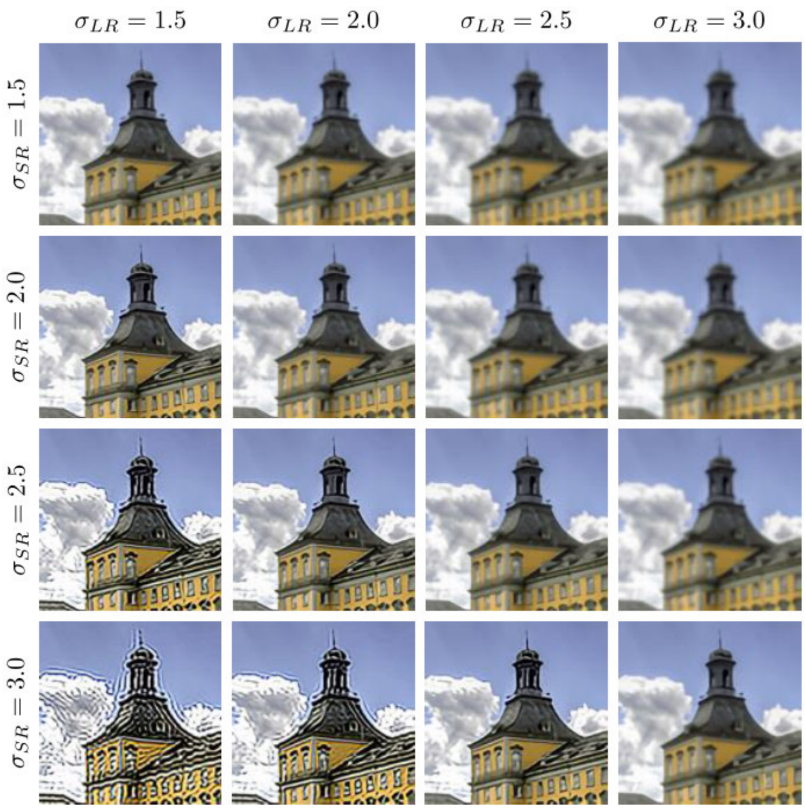
$$\mathbf{I}_{LR} = (\mathbf{I}_{HR} * \mathbf{k}) \downarrow_s$$

$\mathbf{I}_{LR}$: a low-resolution image
$\mathbf{I}_{HR}$: a high-resolution image
$\mathbf{k}$: downscaling kernel
s: a scale factor

  - In general,
    - $\mathbf{k}$ is assumed to be the bicubic kernel.
    - Noise and compression artifact are ignored.

Real-World Super-Resolution

- Unsupervised super-resolution
  - However, $\mathbf{k}$ in the real-world and downsampling are different.
  - The discrepancy between ideal and real $\mathbf{k}$ causes performance degradation for the SR methods.



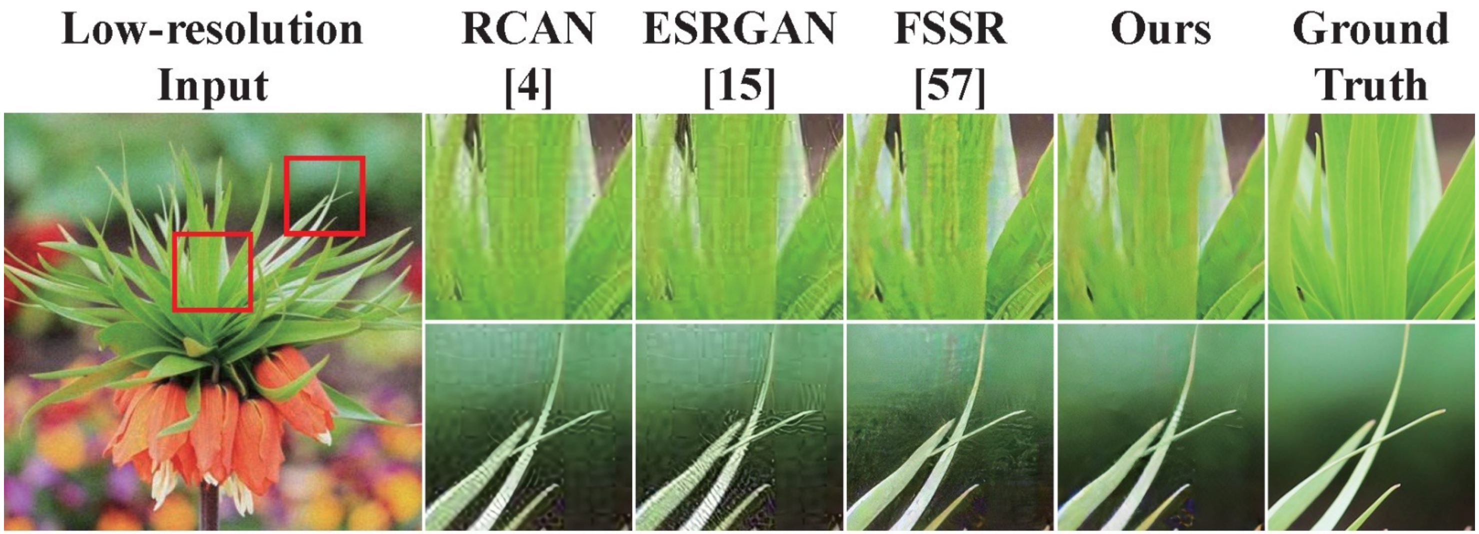
σ_{LR}: σ of Gaussian kernel to simulate $\mathbf{k}$ in the real-world.

σ_{SR}: σ of Gaussian kernel $\mathbf{k}$ employed to construct the training dataset.

→ Blind SR (SR with unknown $\mathbf{k}$.)

Real-World Super-Resolution

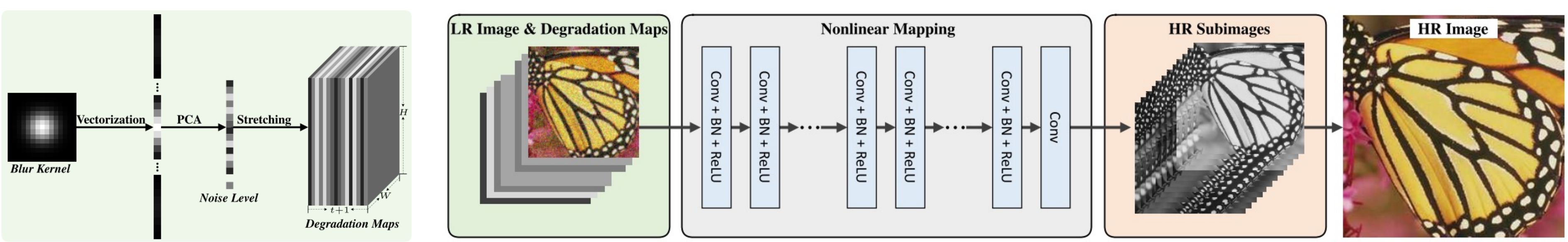
- Unsupervised super-resolution
  - However, noise and compression artifact could be occurred in the real-world.
  - Super-resolution networks trained without the noise and compression artifact cannot deal with the real-world images.



→ Real-World SR (SR with noise and artifact.)

Real-World Super-Resolution

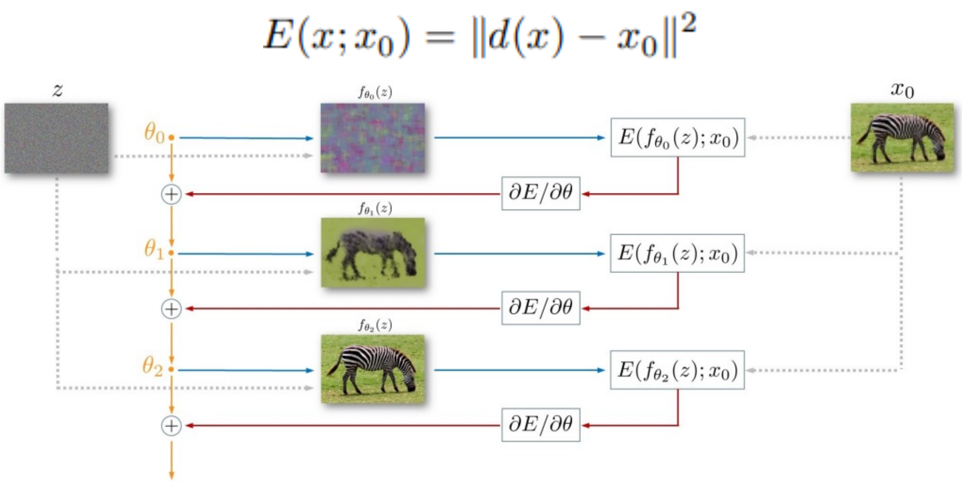
- Unsupervised super-resolution
  - Super-Resolution Network for Multiple Degradations (SRMD*)
    - Training on $\mathbf{I}_{LR} - \mathbf{I}_{HR}$ pairs generated with various $\mathbf{k}$.
    - Generates degradation maps from blur kernel and noise level.
    - $\mathbf{I}_{LR}$ and the degradation maps are employed to recover $\mathbf{I}_{HR}$.
    - It is difficult to deal with $\mathbf{k}$ which is not presented in the dataset.



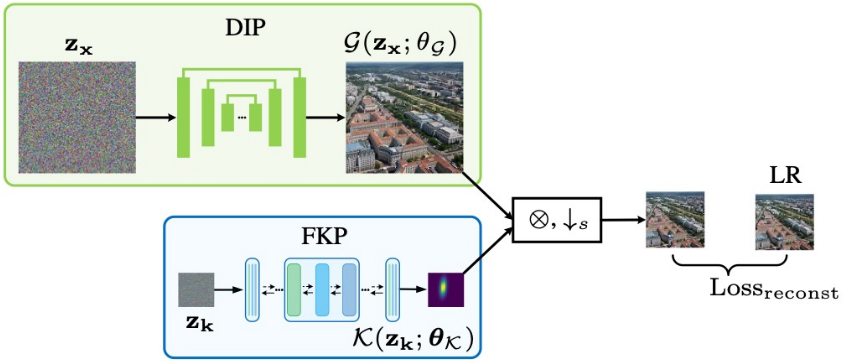
* K. Zhang *et al.*, "Learning a single convolutional super-resolution network for multiple degradations," CVPR, 2018.

Real-World Super-Resolution

- Deep image prior (DIP)
  - The CNN structure is sufficient to capture a great deal of low-level image statistics prior for inverse problems.
  - DIP* employ a randomly-initialized CNN as handcrafted prior to perform SR.



DIP

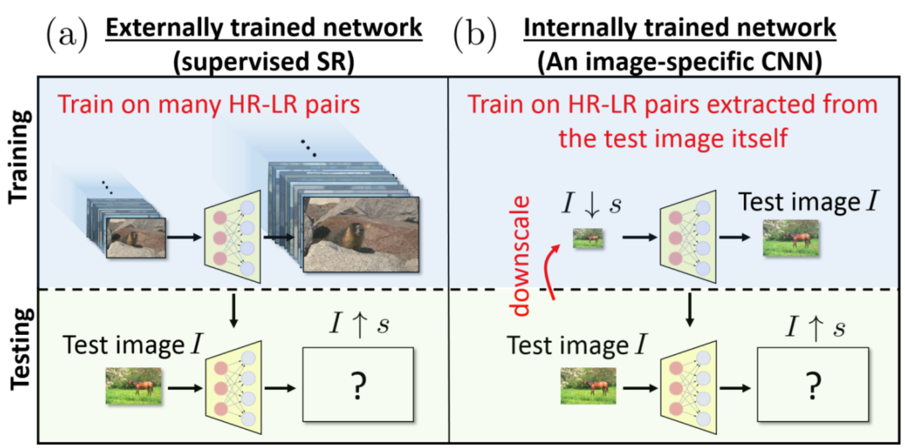


DIP+FKP** (DIP + Flow-based Kernel Prior)

* D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 9446–9454.
** J. Liang, K. Zhang, S. Gu, L. V. Gool, R. Timofte, "Flow-based Kernel Prior with Application to Blind Super-Resolution," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021.

Real-World Super-Resolution

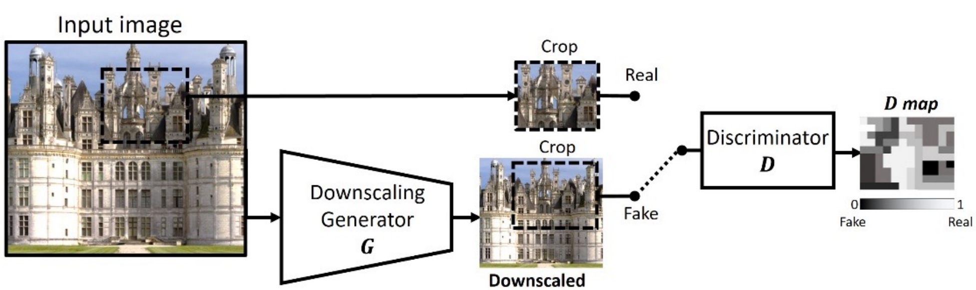
- Zero-shot super-resolution
  - ZSSR* cope with unsupervised SR by training image-specific SR networks at test time rather than training a generic model on large external datasets.
    - Since it needs to train different networks for different images during testing, the inference time is much longer than others.
    - The ZSSR leverages on the cross-scale internal recurrence inside every image, and thus outperforms previous approaches on images under nonideal conditions.
      - i.e., images obtained by non-bicubic degradation and suffered effects like blurring, noise, compression artifacts, which is closer to the real-world scenes.

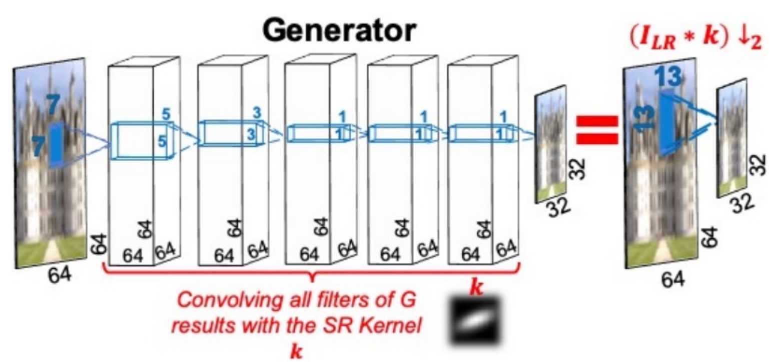


* A. Shocher *et al.*, "Zero-shot" super-resolution using deep internal learning," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 3118–3126

Real-World Super-Resolution

- Zero-shot super-resolution
  - KenelGAN*
    - Zero-shot kernel estimation for blind SR
    - The patch GAN trains on patches of a single input image (real).
    - D tries to distinguish real patches from those generated by G (fake).
    - G learns to downscale the image while fooling D i.e. maintaining the same distribution of patches.
    - Since G is a linear network, convolving all filters results in the downscaling kernel.

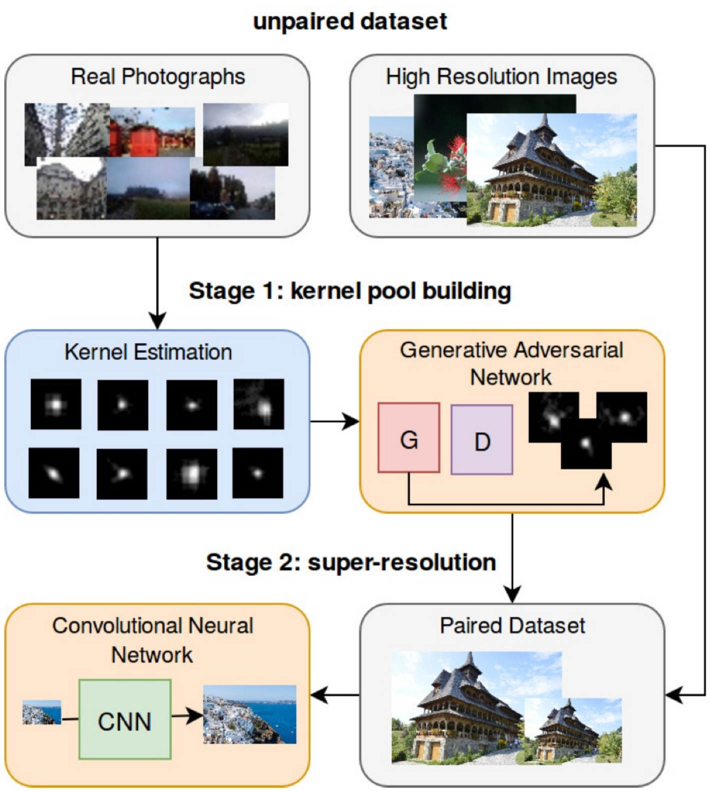


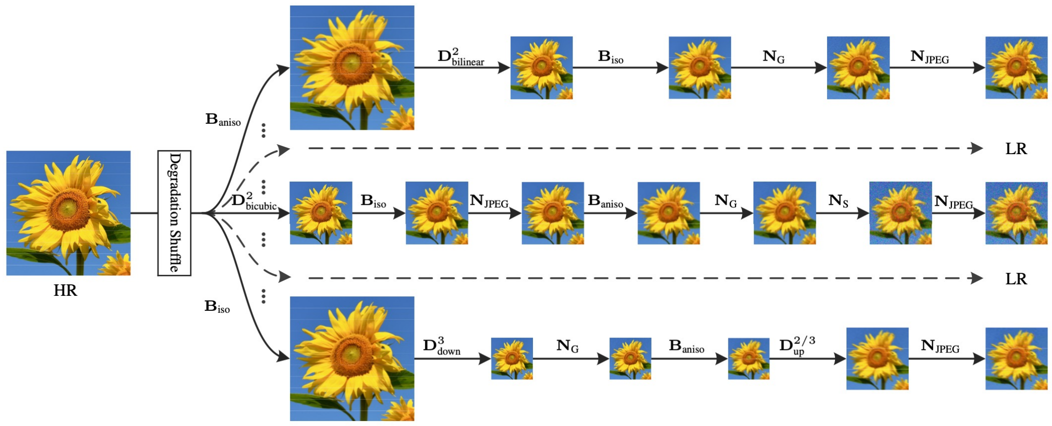


* S. Bell-Kligler *et al.*, "Blind Super-Resolution Kernel Estimation using an Internal-GAN," in Proc. 33rd Int. Conf. Neural Inf. Process. Syst., 2019.

Real-World Super-Resolution

- Degradation modeling
  - Unsupervised SR → Supervised SR
  - KMSR*: Blur kernel generation → Paired data generation
  - BSRGAN**: HR → LR by randomly shuffled degradation process

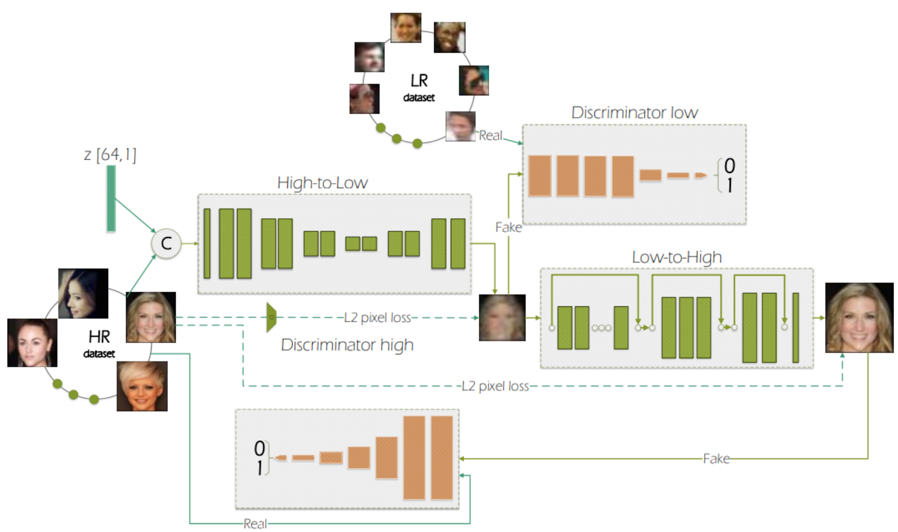




* R. Zhou and S. Süsstrunk, "Kernel Modeling Super-Resolution on Real Low-Resolution Images," in Proc. Int. Conf. Comput. Vis., 2019.
** K. Zhang, J. Liang, L. V. Gool, and R. Timofte, "Designing a Practical Degradation Model for Deep Blind Image Super-Resolution," in Proc. Int. Conf. Comput. Vis., 2021.

Real-World Super-Resolution

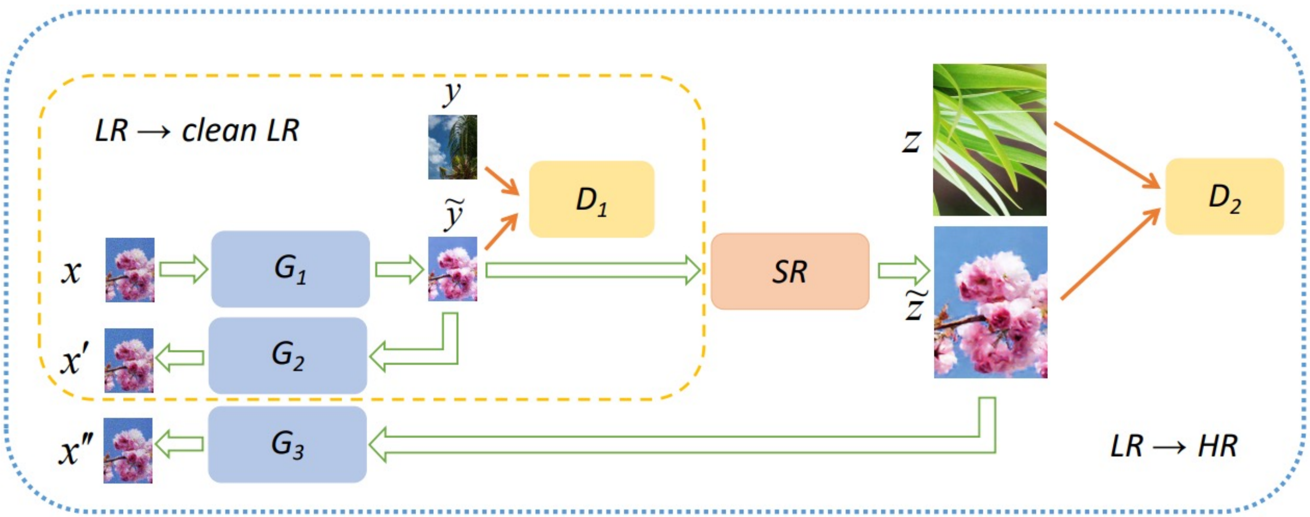
- Weakly-supervised super-resolution
  - Degradation learning
    - Since the predefined degradation is suboptimal, learning the degradation from unpaired LR-HR datasets is a feasible direction.
    - A two-stage process which first trains an HR-to-LR GAN to learn degradation using unpaired LR-HR images and then trains an LR-to-HR GAN for SR using paired LR-HR images conducted base on the first GAN.
    - Bulat et al.*, FSSR**



* A. Bulat *et al.*, "To learn image super-resolution, use a gan to learn how to do image degradation first," in Proc. Eur. Conf. Comput. Vis., 2018, pp. 185–200.
** M. Fritsche, S. Gu and R. Timofte, "Frequency Separation for Real-World Super-Resolution," 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019, pp. 3599-3608.

Real-World Super-Resolution

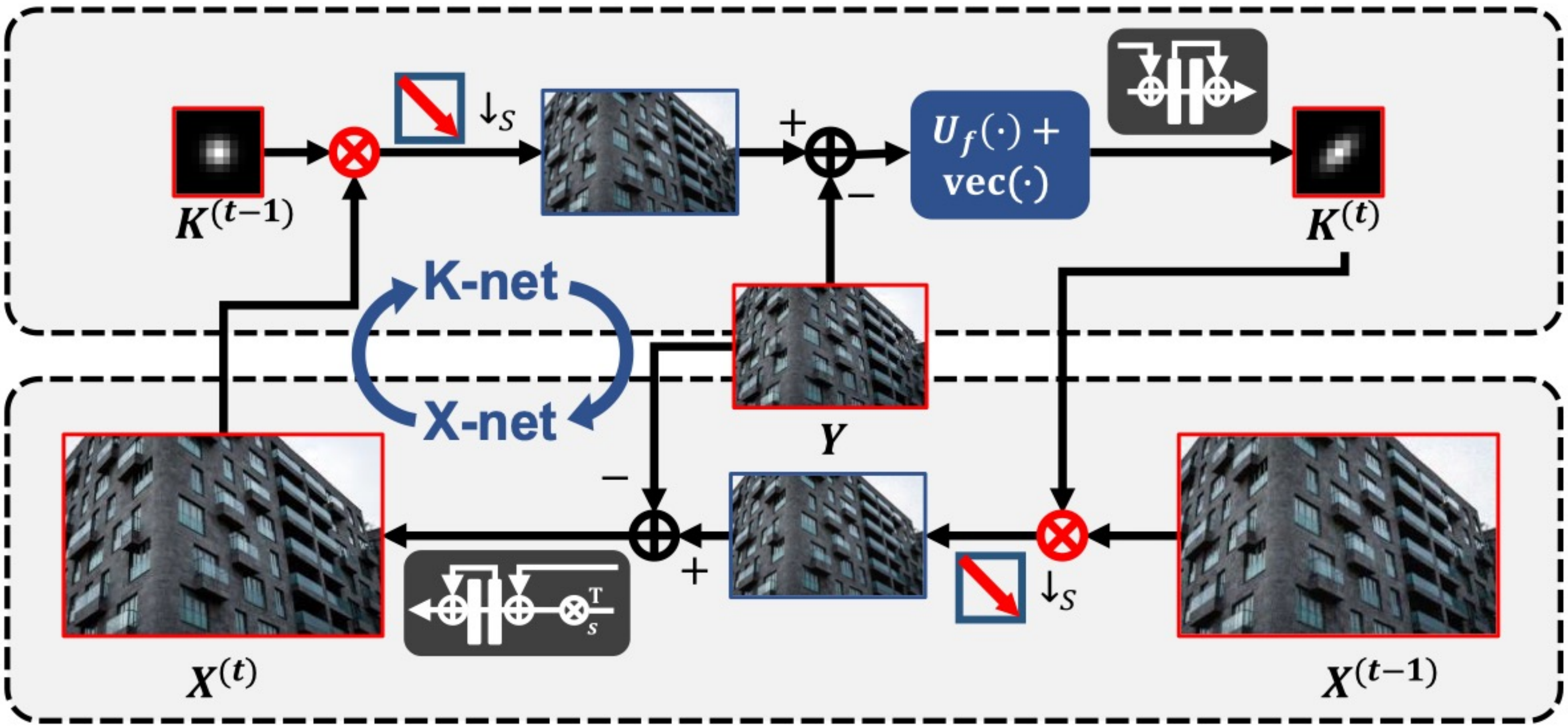
- Weakly-supervised super-resolution
  - Cycle-in-Cycle GAN (CinCGAN*)
    - Treat the LR space and the HR space as two domains, and use a cycle-in-cycle structure to learn the mappings between each other.
    - The training objectives include pushing the mapped results to match the target domain distribution and making the images recoverable through round-trip mappings.
    - Composed of 4 generators and 2 discriminators, making up two CycleGANs for noisy LR ↔ clean LR and clean LR ↔ clean HR mappings.



* Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, and L. Lin, "Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, 2018, pp. 814–81409.

Real-World Super-Resolution

- Weakly-supervised super-resolution
  - KXNet*: Simultaneous blur kernel estimation and super-resolution



* J. Fu, H. Wang, Q. Xie, Q. Zhao, D. Meng, and Z. Xu., "KXNet: A Model-Driven Deep Neural Network for Blind Super-Resolution," in Proc. Eur. Conf. Comput. Vis., 2022.

Exercise #2
Unsupervised SR

Exercise #2 – Unsupervised SR

- Zero-shot Super-resolution
  - pip install gputil
  - pip install opencv-python
  - git clone https://github.com/HarukiYqM/pytorch-ZSSR
    - Official github: https://github.com/assafshocher/ZSSR
    - Tensorflow 기반으로 되어 있어 pytorch 기반 github 사용.
  - cd pytorch-ZSSR
  - cp -r set14 test_data
  - python run_ZSSR.py