Super-Resolution (SR): Computational and Deep Learning-Based Approaches

Majid Rabbani – EME Visiting Professor

and

Prasanna Reddy Pulakurthi – EME PhD Student

Presented on April 26, 2023 as part of the Society for Imaging Science and Technology (IS&T) Rochester NY chapter seminar series.

Presentation Notes

1. This talk was presented on Wednesday, April 26, 2023 by Majid Rabbani and Prasanna Reddy Pulakurthi as part of the Society for Imaging Science and Technology (IS&T) Rochester NY chapter seminar series.

2. This presentation file was graciously provided by the authors.

3. The **presentation video file** is available on YouTube at, https://youtu.be/czlEG-QkKRI

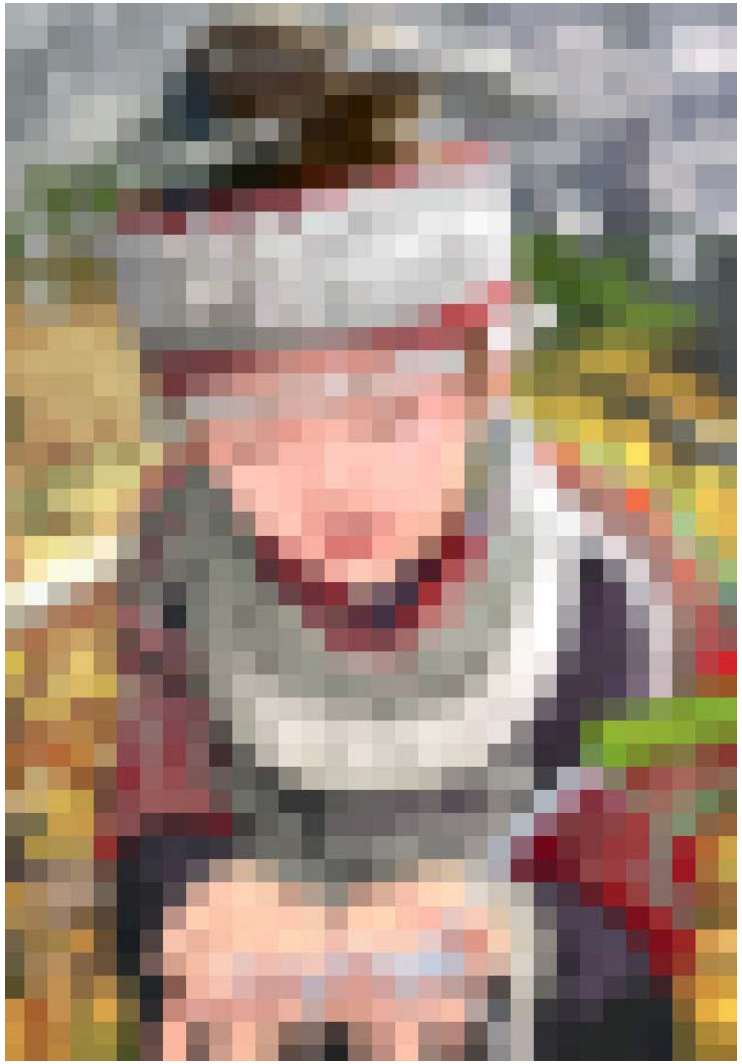
4. **Blade Runner scene**: The presentation included the showing of a clip from the film, *Blade Runner* here, ***https://youtu.be/hHwjceFcF2Q***

- Peter Burns

What Is Image Resolution? *(Presented by M. Rabbani)*

- Resolution can mean different things to different imaging applications:
  - Spatial resolution (pixel density in an image, usually measured as pixels per unit area)
  - Radiometric (tonal) resolution (bit depth)
  - Temporal resolution (frames per second)
  - Spectral resolution (number of color planes, spectral bands, etc.)
- In the context of this presentation, Super-Resolution (SR) refers to obtaining an image at a spatial resolution higher than that of the camera sensor.

Perceptual Appearance of 2X and 4X Increase In Resolution



46×32

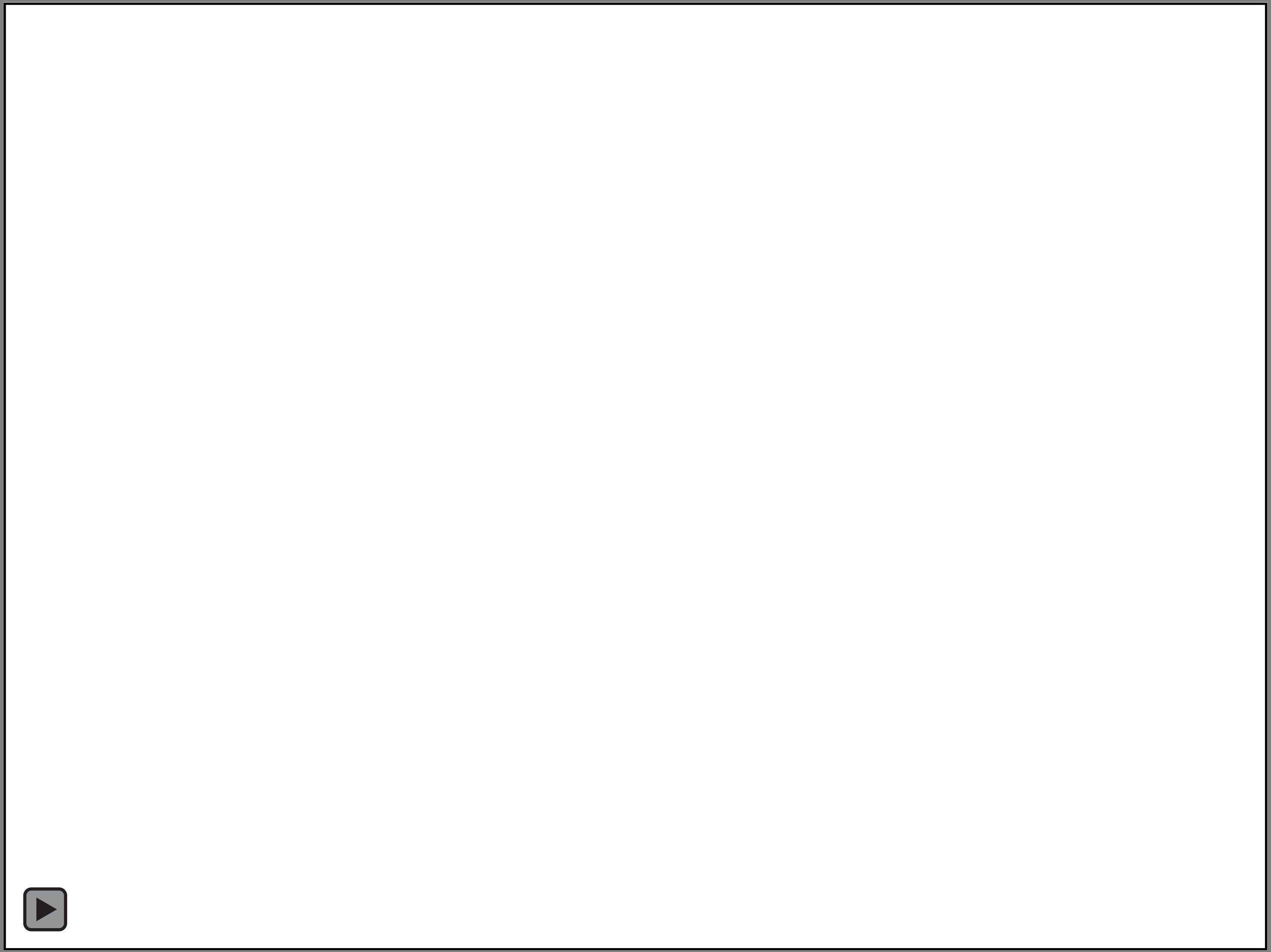
91×63 (2X)

181×125 (4X)

“Blade Runner” Movie: 1982 – Scene Set in November 2019

“Enhance 34 to 36. Pan right and pull back. Stop. Enhance 34 to 46. Pull back. Wait a minute, go right, stop. Enhance 57 to 19. Track 45 left. Stop. Enhance 15 to 23. Give me a hard copy right there.”

https://youtu.be/hHwjceFcF2Q



Seminal Papers on Super-Resolution

- Nearly two years after the release of the “Blade Runner” movie, the paper by Tsai and Huang[1] marked the inception of computational SR (aka reconstruction-based SR).
- Nearly two decades after the release of the “Blade Runner” movie, the paper by Baker and Kanade[2] on the “Face Hallucination” marked the inception of example-based SR.
- Advent of Deep Learning (circa 2015): “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network – CVPR 2017

SRGAN

From the recorded CVPR presentation: https://www.youtube.com/watch?v=BXIR_SVCrsE

1. *R. Tsai, T. Huang, “Multi-frame image restoration and registration,” Advances in Computer Vision and Image Processing,” (1), no. 2, 1984, pp. 317-339*
2. *S. Baker and T. Kanade, “Hallucinating faces,” In IEEE International Conference on Automatic Face and Gesture Recognition, 2000.*

Computational (Traditional Signal Processing) Based Approach

Super-Resolution Framework

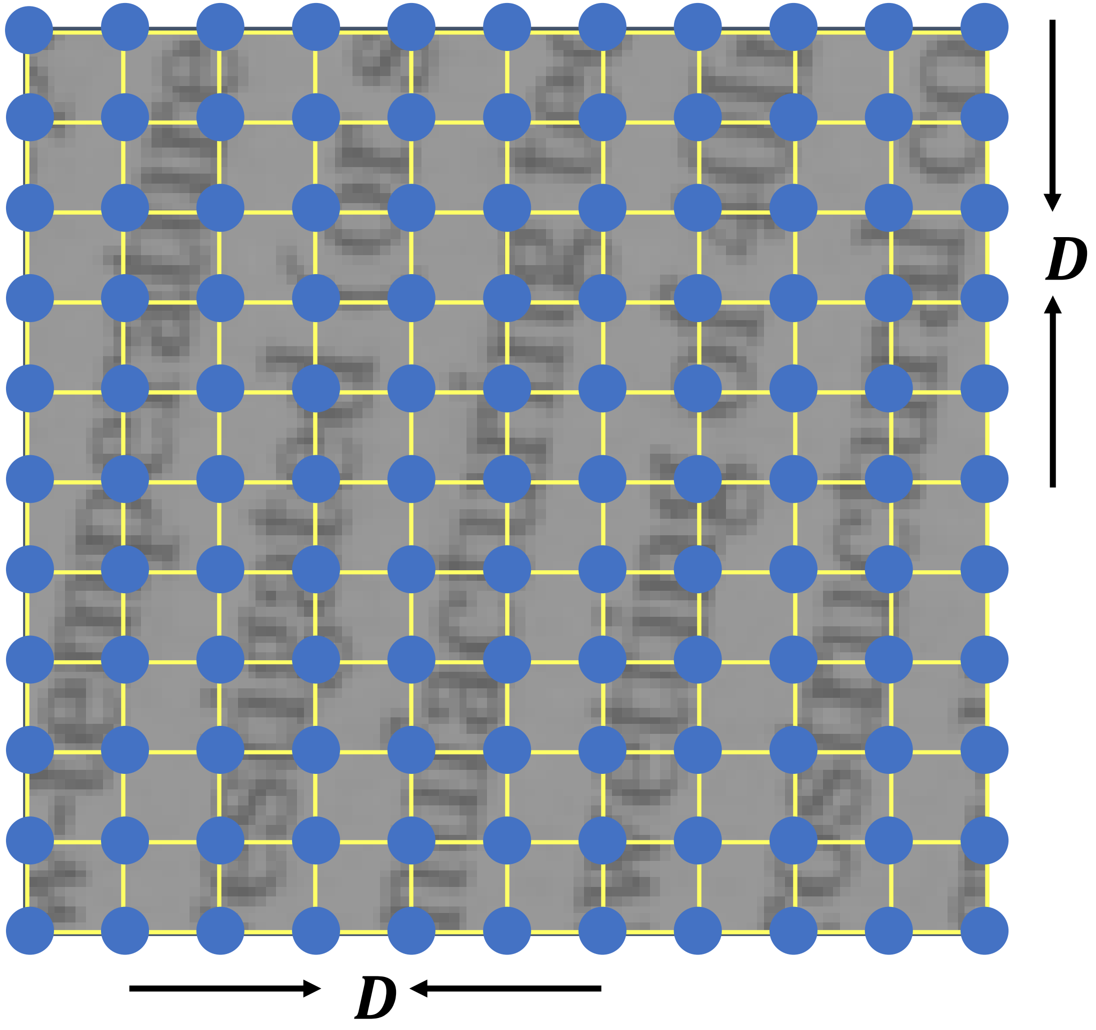
- Super-resolution refers to obtaining an image at a resolution higher than that of the camera sensor.
- Super-resolution from a single low-resolution image is a highly ill-posed problem.
- The problem becomes more manageable when a sequence of low-resolution frames is available, e.g., multiple frames of a scene captured by a video camera.
- Computational SR algorithms construct High-Resolution (HR) images from several observed Low Resolution (LR) images by increasing the high-frequency components and removing any existing degradations.

Super-Resolution Commercial Example (circa 2005)

From http://www.motiondsp.com

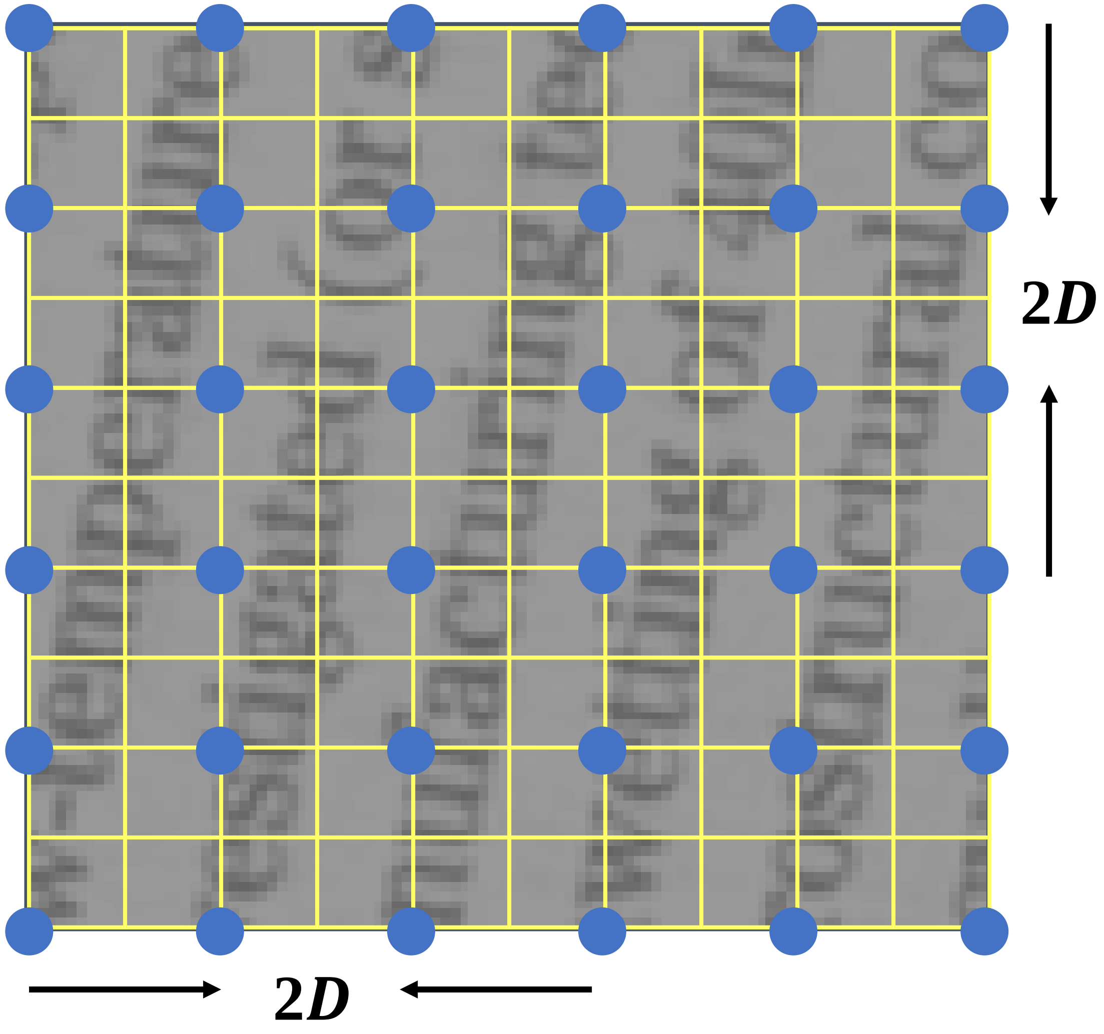
Intuition

- For a given band-limited image, the Nyquist sampling theorem states that if a uniform sampling is fine enough ($\geq \boldsymbol{D}$), perfect reconstruction is possible.



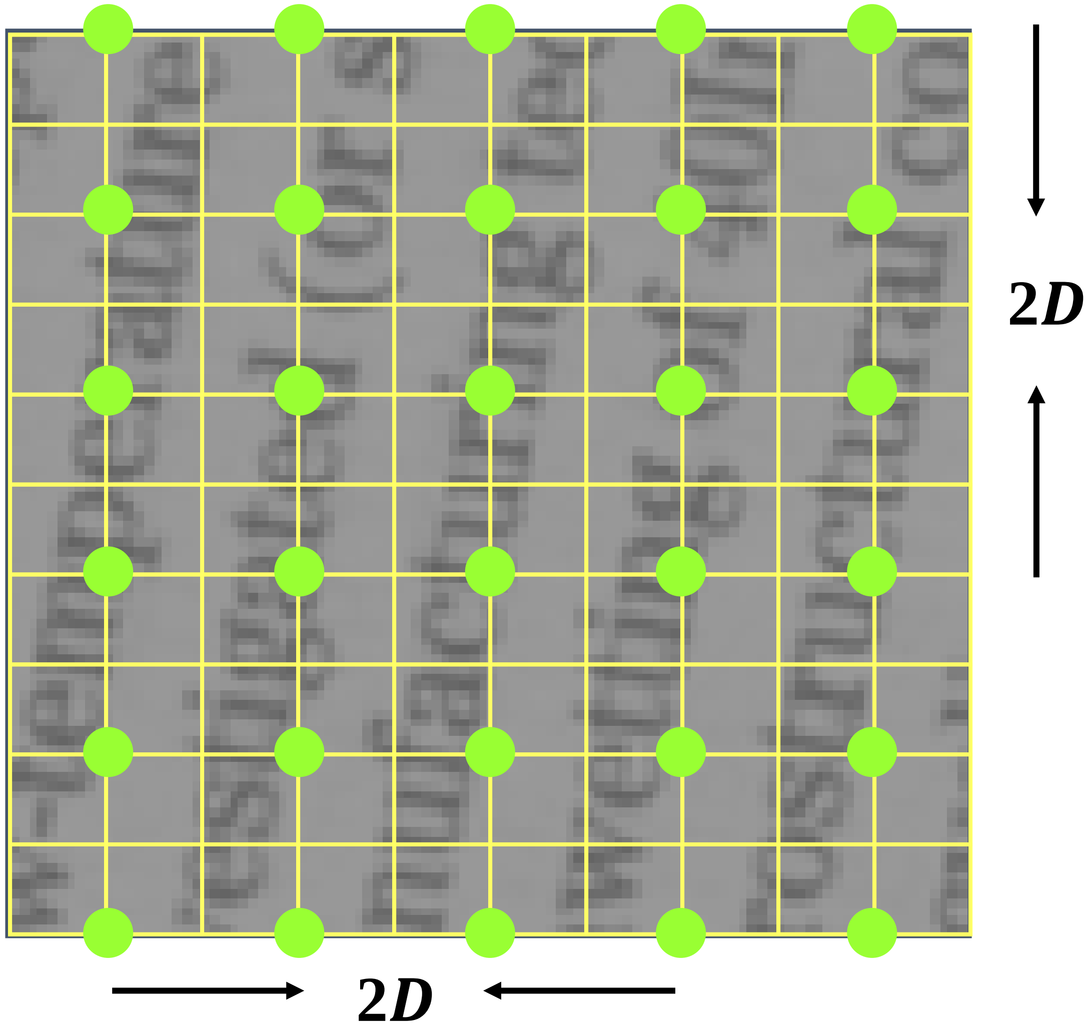
Intuition

- Due to the limited camera resolution, the scene is sampled using an insufficient two-dimensional grid.



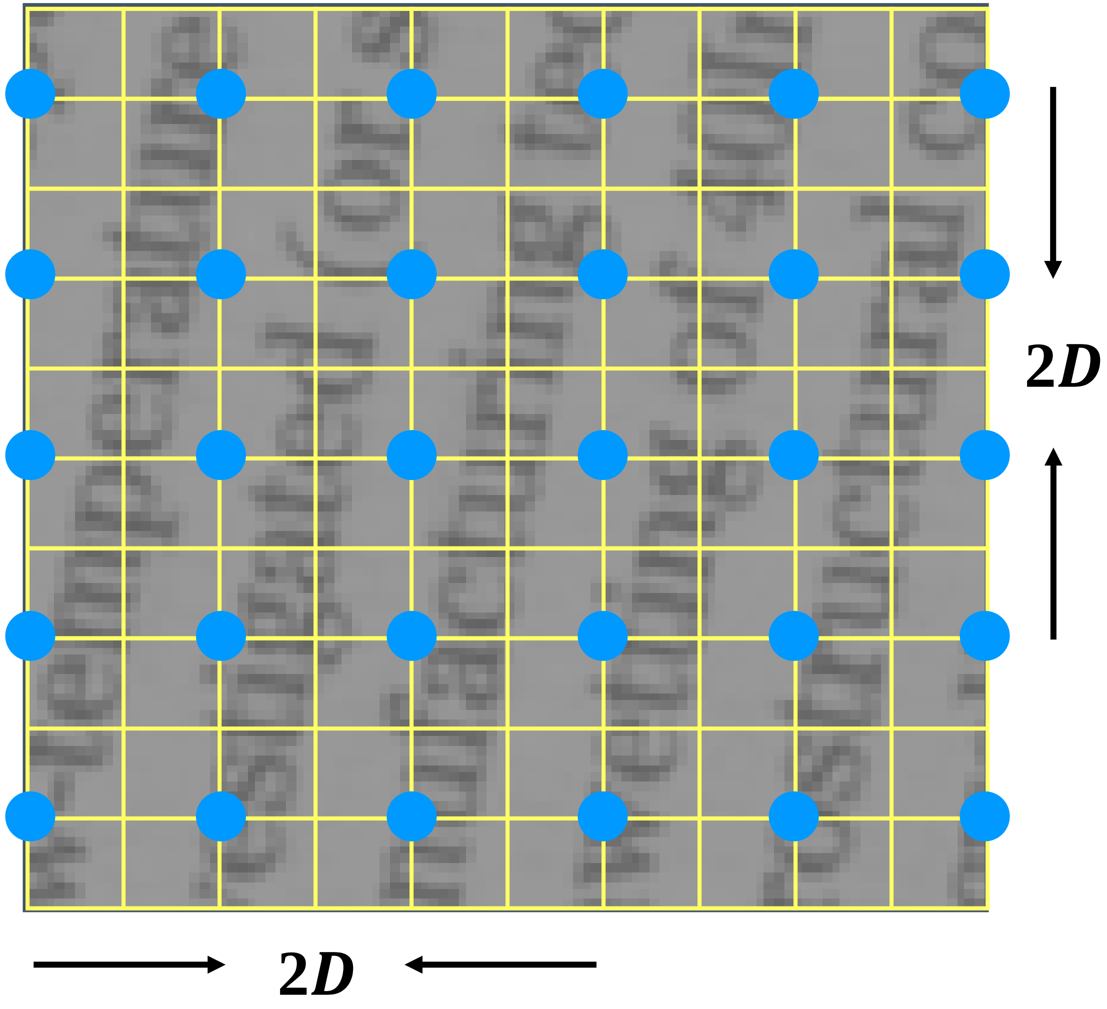
Intuition

- However, if a second picture is taken while shifting the camera ‘slightly to the right,’ (in this case, ½ pixel displacement) a different low-resolution rendition of the scene would be captured.



Intuition

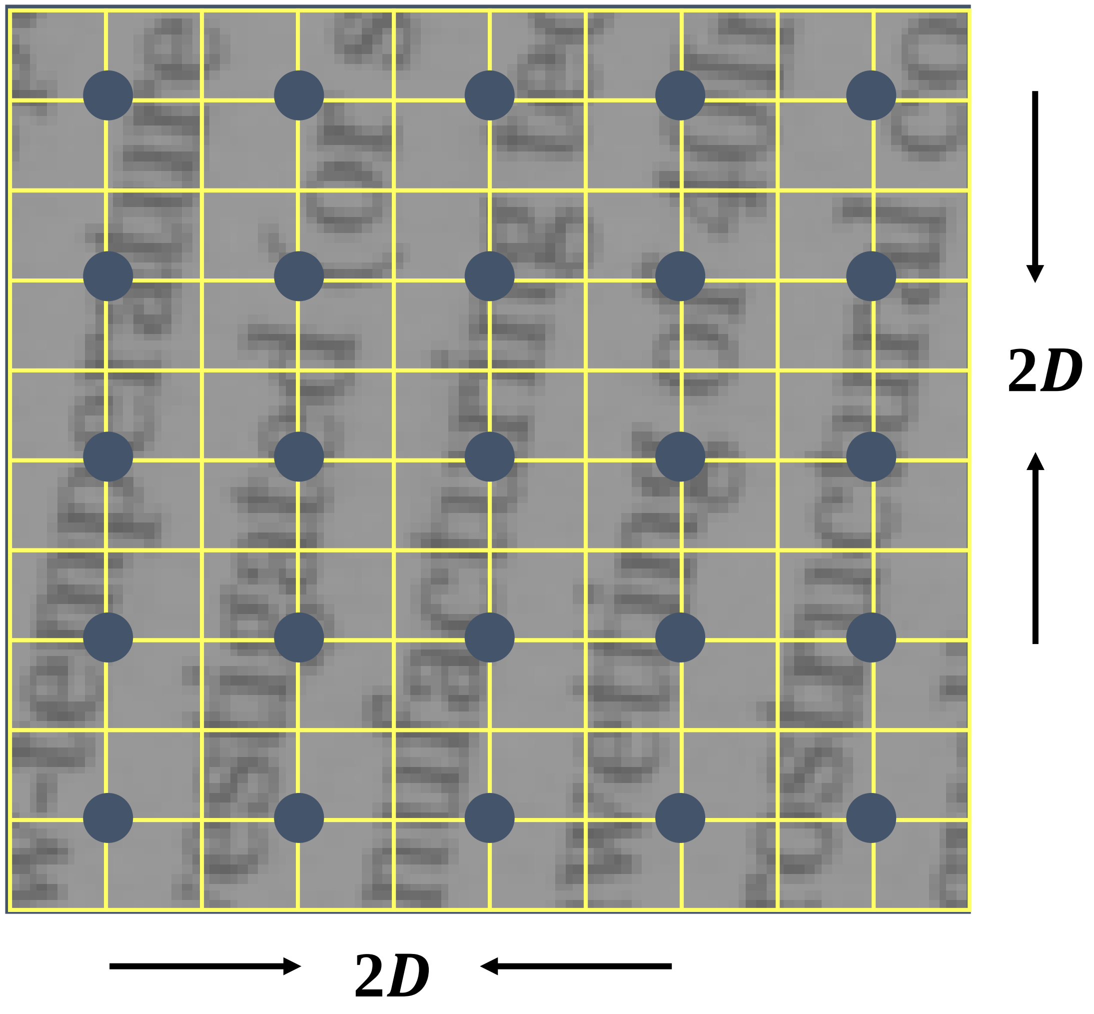
- Similarly, by shifting the camera down by ½ pixel, a third but different low-resolution image is obtained.



From **Rossi Rubner**: www1.idc.ac.il/toky/CompPhoto-09/Lectures/08_SR_09.ppt

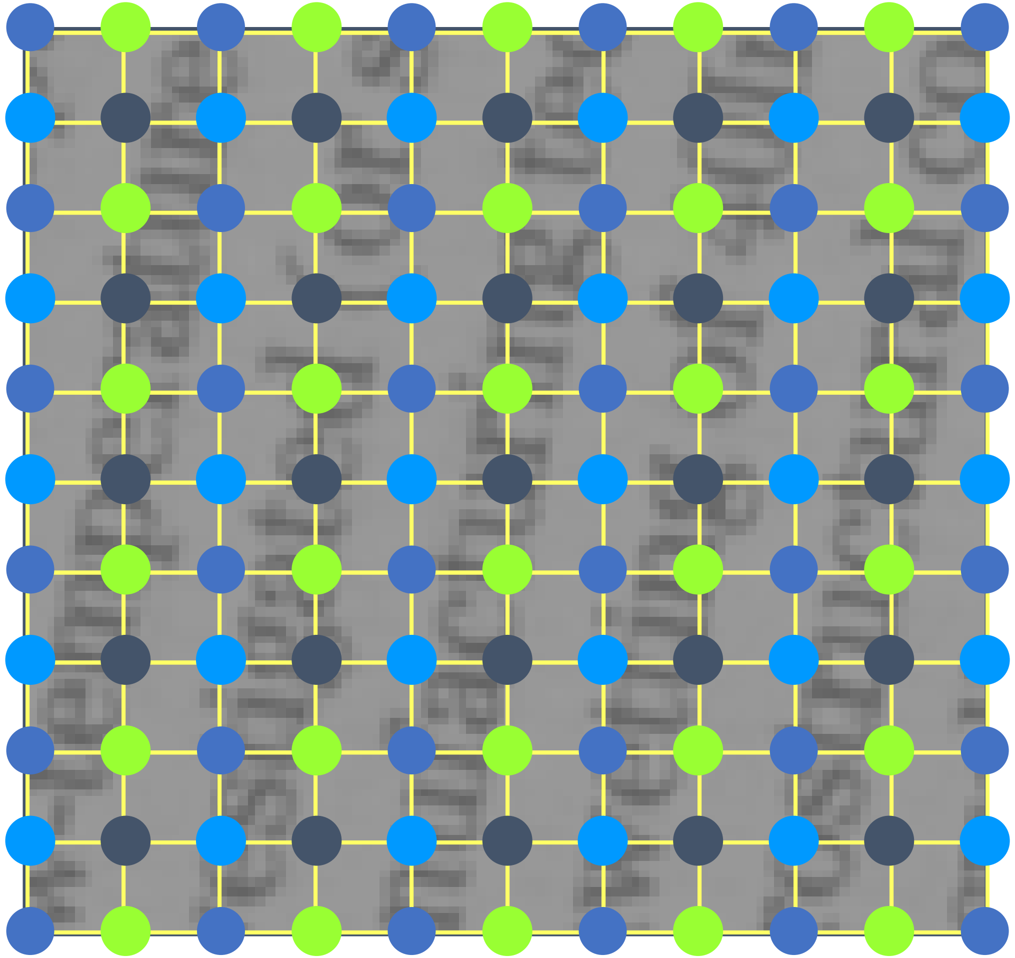
Intuition

- Finally, by shifting down and to the right a fourth and final low-resolution image can be captured.



Intuition

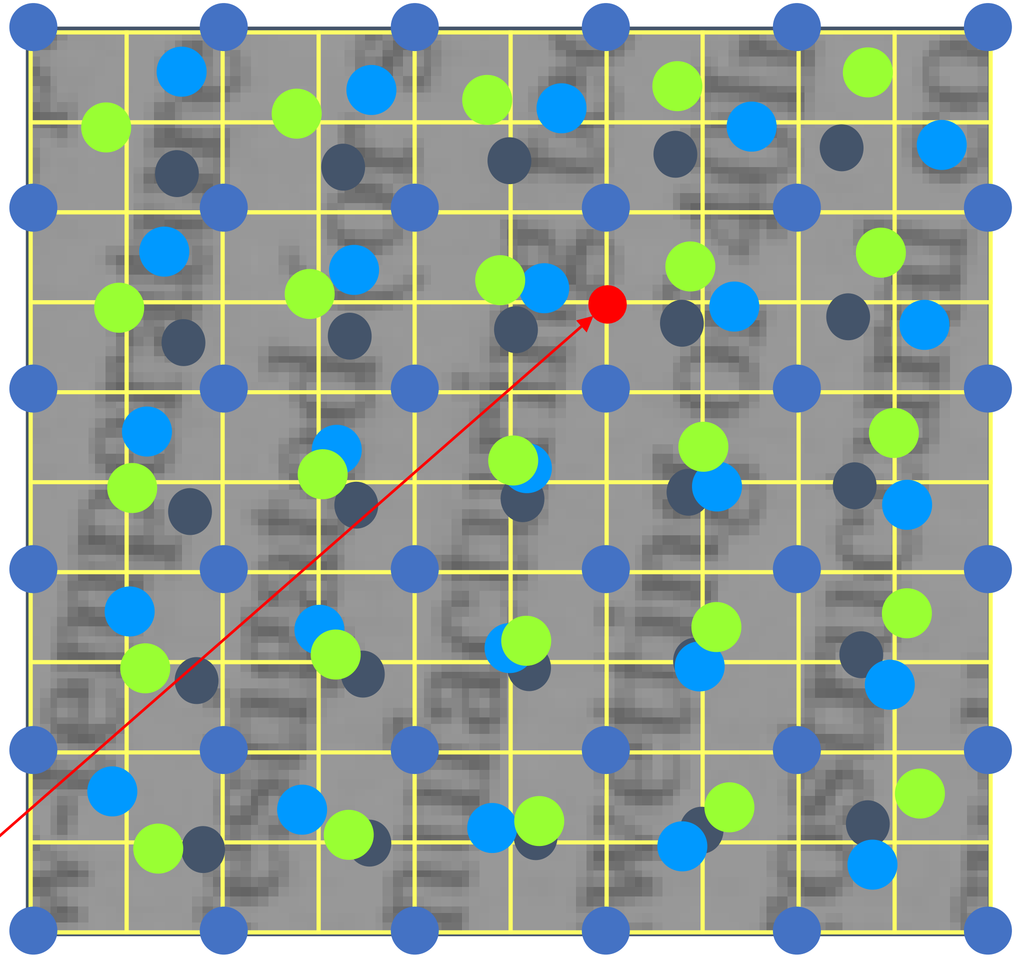
- It is straightforward to see that by interlacing the four images, the desired resolution can be obtained, and thus perfect reconstruction is possible.



From **Rossi Rubner**: www1.idc.ac.il/toky/CompPhoto-09/Lectures/08_SR_09.ppt

Rotation/Scale/Displacement – Capture Issues in Real Life

- What if the camera displacement is arbitrary?
- What if the camera rotates?
- What if the camera gets closer to the object (zoom)?
- Unfortunately, this is typically the case in practice.



Reconstruct at this location

Super-Resolution Observation Model

+

=

Hi-Res Scene	Geometric Transformation	Optical Blur	Down Sampling		Noise		k^{th} Low-Res Observation
$\boldsymbol{x}$	$\boldsymbol{M}_k$	$\boldsymbol{B}_k$	$\boldsymbol{D}$	+	$\boldsymbol{n}_k$	=	$\boldsymbol{y}_k$

$$\boldsymbol{y}_k = \boxed{\boldsymbol{D}\boldsymbol{B}_k\boldsymbol{M}_k}\,\boldsymbol{x} + \boldsymbol{n}_k = \boxed{\boldsymbol{W}_k}\,\boldsymbol{x} + \boldsymbol{n}_k$$

Observation Model

Super-Resolution Observation Model Assumptions

$\boldsymbol{y}_k = \boldsymbol{D}\boldsymbol{B}_k\boldsymbol{M}_k\boldsymbol{x} + \boldsymbol{n}_k = \boldsymbol{W}_k\boldsymbol{x} + \boldsymbol{n}_k$ observation model

$\boldsymbol{y}_k$ observed low-resolution noisy images **(*M* captures - known)**

$\boldsymbol{x}$ original high-resolution image **(Needs to be calculated)**

$\boldsymbol{M}_k$ motion warp matrix: global or local translation, rotation, etc., **(estimated)**

$\boldsymbol{B}_k$ blur model: optical, motion, sensor pixel size, etc. (**known** or **estimated** based on application)

$\boldsymbol{D}$ downsampling matrix dictated by required resolution ratio **(known)**

$\boldsymbol{n}_k$ Noise: (exact value not known but statistics are **known**)

Latest Advancements In Reconstruction-Based SR

- In practice, the noise may not be exactly white and motion estimates may not all be accurate. When using the L_2 norm energy function, a single outlier can potentially ruin the entire estimation process.

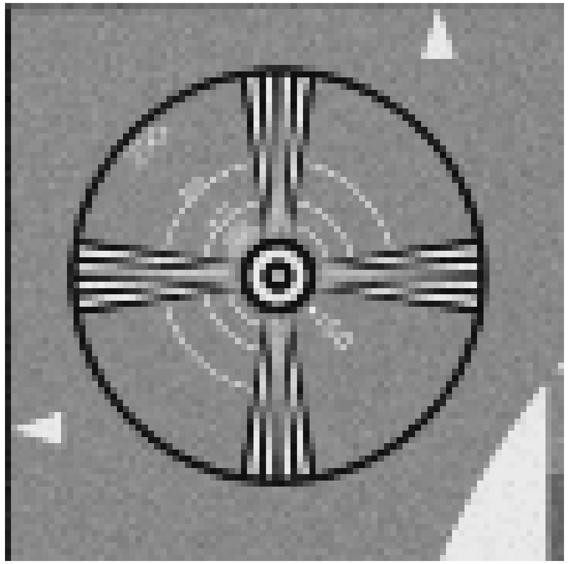
- Farsiu et al[5] develop two methods for robust super-res by (i) minimizing L_1 norm energy function for both data fidelity and regularization, which is much more robust to outliers, and (ii) using the bilateral filter as a regulating term.

  1. First method is general and works for any type of motion.
  2. Second method is extremely fast, but is only appropriate for translational motion.

5. *Farsiu, Robinson, Elad, and Milanfar, "Fast and Robust Multiframe Super Resolution", IEEE TIP, (13), No. 10, pp. 1327-1344, (2004).*

Latest Advancements In Reconstruction-Based SR: Examples

20 images, Resolution enhancement ratio 4X



L_2-norm based – Old

L_1-norm based – New

- Mean minimizes the sum of squared deviations – L_2–norm (mathematically tractable)
- Median minimizes the sum of absolute deviations – L_1–norm (insensitive to outliers)

5. *Farsiu, Robinson, Elad, and Milanfar, "Fast and Robust Multiframe Super Resolution", IEEE TIP, (13), No. 10, pp. 1327-1344, (2004).*

Latest Advancements In Reconstruction-Based SR: Examples

One of 8 LR captured frames

L_1+bilateral TV

You can download a Matlab version of the code written by Oded Hanson and experiment with your own images: https://faculty.idc.ac.il/toky/old_courses/videoProc-07/projects/SuperRes/srproject.html

5. *Farsiu, Robinson, Elad, and Milanfar, "Fast and Robust Multiframe Super Resolution", IEEE TIP, (13), No. 10, pp. 1327-1344, (2004).*

Deep Learning Based Approach

What Made Deep Learning Feasible For Images and Video

- Hardware advances – faster processors, parallelization, cheaper hardware and memory (Nvidia RTX 4090 is ~100 TFLOPS and sells for $1,599)
- Deeper (more) layers – Darknet-53, Resnet-152, GPT-3 uses a model with over 175 billion parameters.
- Dense instead of sparse models (e.g., convolutional networks)
- Advanced neuron learning models (e.g., ReLU activation function, batch normalization, Adam optimizers, transfer learning, etc.)
- Large and diverse training datasets (ImageNet is a large database >14M images, Microsoft Celeb (MS-Celeb-1M) is a dataset of 10M faces.

Super Resolution Using Deep Learning

1. Techniques that improve the resolution of a randomly chosen single-frame LR image.
    a) Residual Dense Network (RDN)
    b) Residual Channel Attention Network (RCAN)
    c) Super-Resolution using a Generative Adversarial Network (SRGAN) – (2017)
2. Techniques using an encoder-decoder paradigm: (i) Starting with a HR image, a LR image is generated using a specific deep network encoder. (ii) The LR image can be upscaled into a HR image using a matched deep network decoder.
    a) Task-Aware Image Downsampling (2018)
    b) Learned Image Downscaling for Upscaling Using Content Adaptive Resampler
    c) Invertible Rescaling Network (IRN) – (2020)

Generative Adversarial Networks

GANs consist of two neural networks that compete against each other (hence the term adversarial) in order to generate new, synthetic instances of data that can pass for real data. The generator tries to generate realistic samples to fool discriminator, while the discriminator tries to distinguish between real and actual samples.

GAN's[6] are the leading technology for producing "Deep Fake" videos on the internet

2x Downscaled

Generator

Discriminator

Real or Fake?

6. *Ian J. Goodfellow, et al, "Generative adversarial nets," In Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'14). MIT Press, Cambridge, MA, USA, 2672–2680, (2014)*

Photo-Realistic Single Image Super-Resolution Using GANs

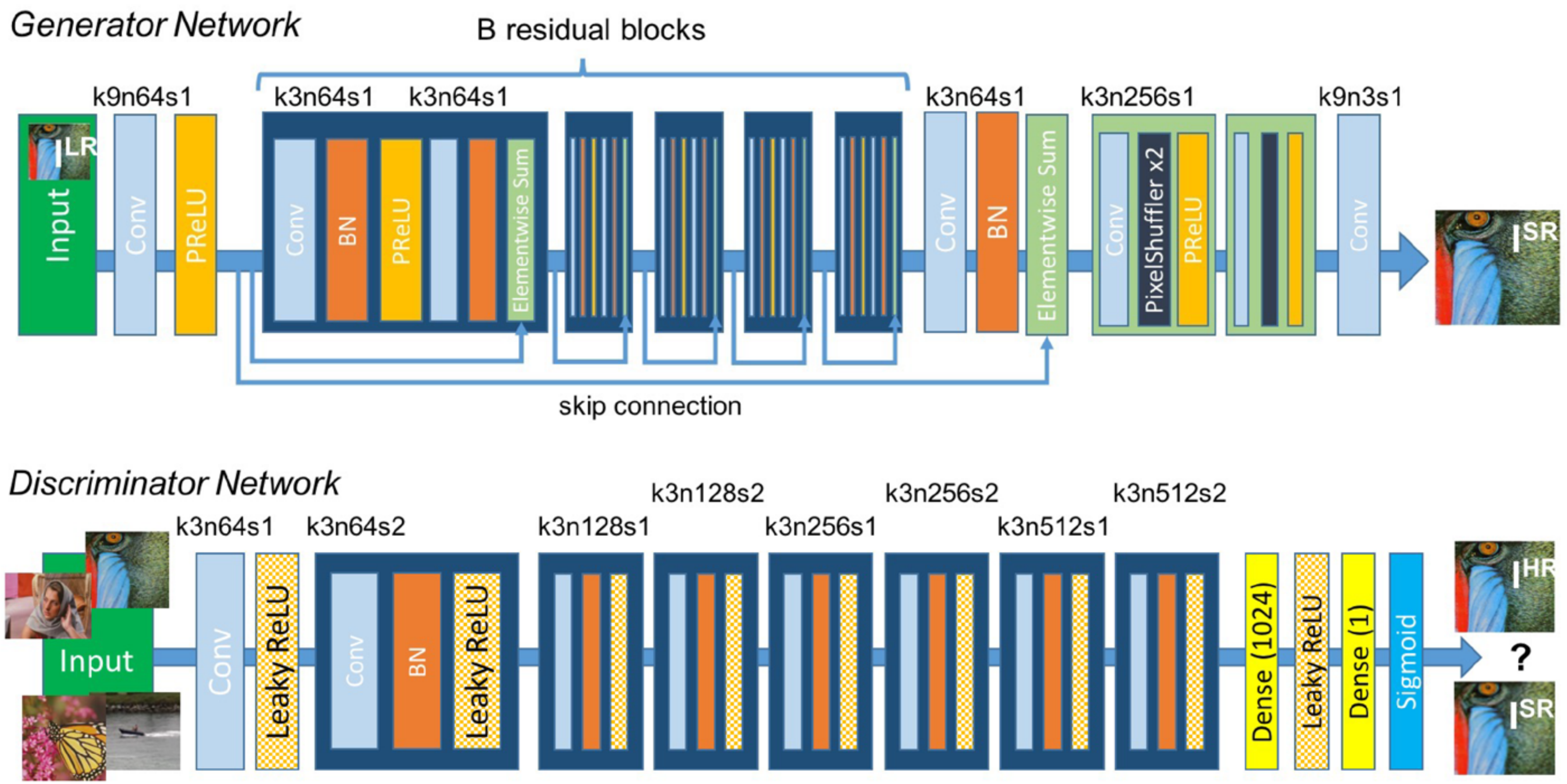


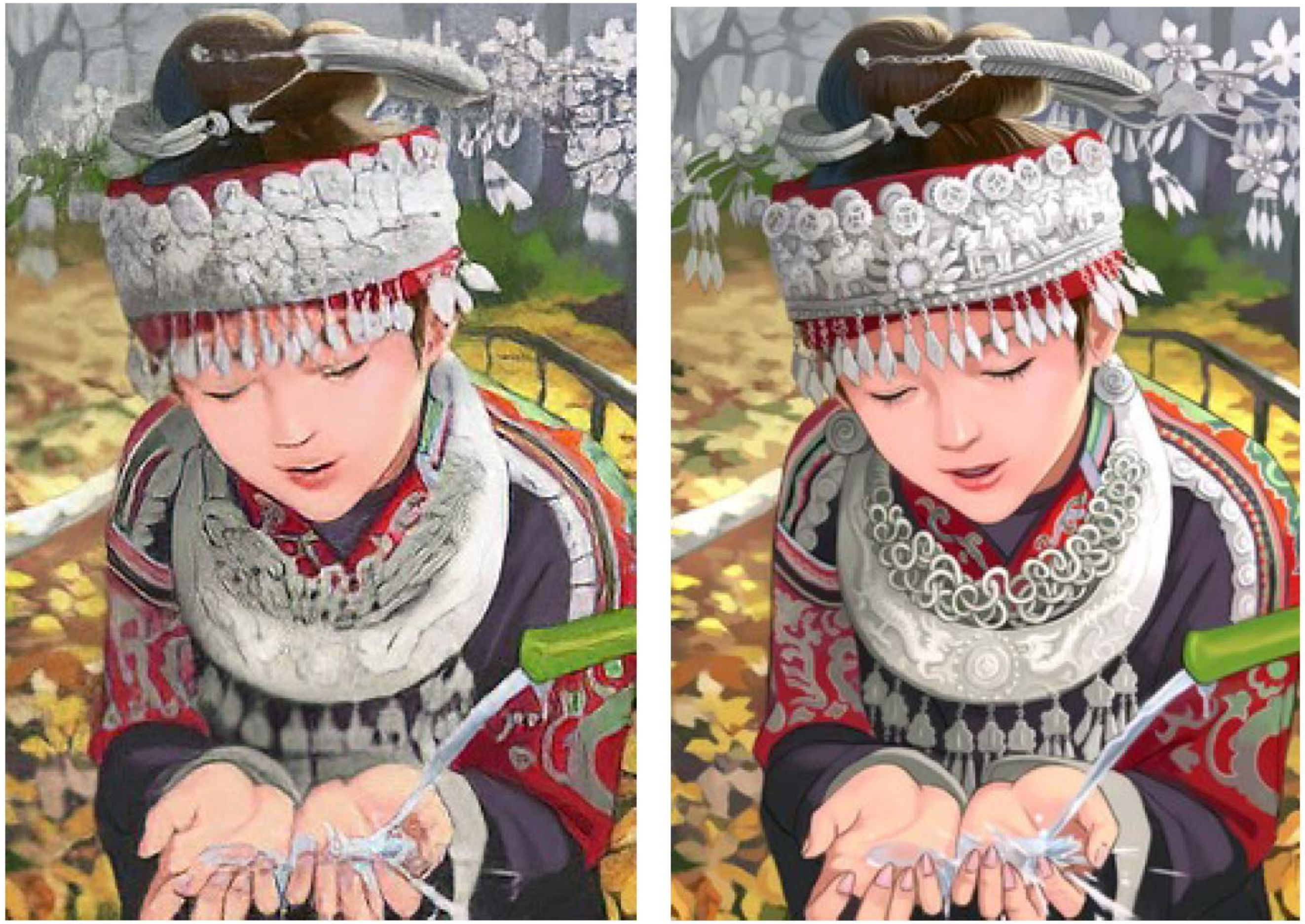
Figure 4: Architecture of Generator and Discriminator Network with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer.

7. *C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," 2017 IEEE Conference on CVPR, Honolulu, HI, USA, 2017, pp. 105-114, doi: 10.1109/CVPR.2017.19*

Trained With Random Samples of 350,000 Images from ImageNet Database

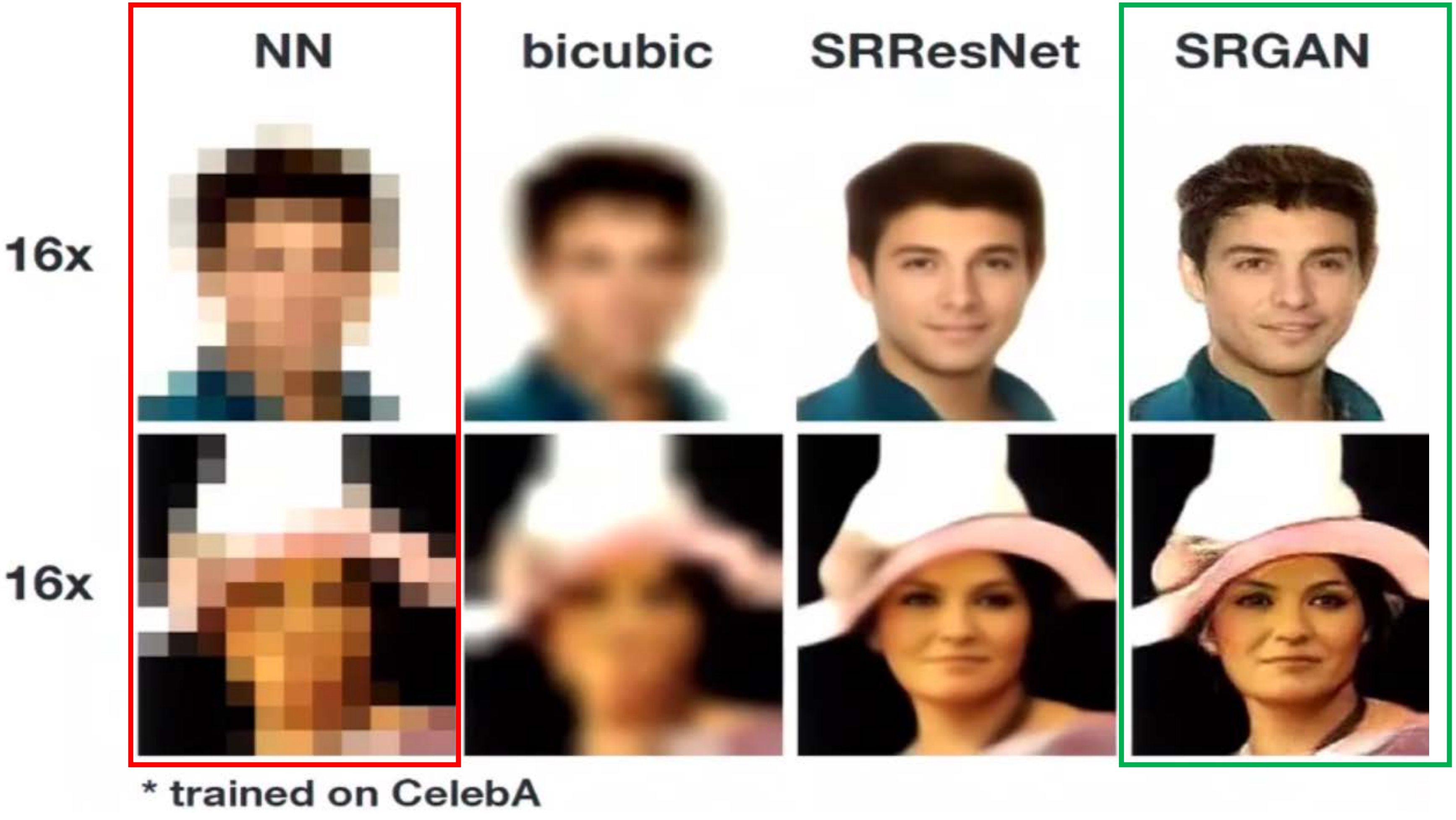
SRGAN
(21.15dB/0.6868)

original



Demonstrating 4X Resolution Upsampling (https://www.youtube.com/watch?v=BXIR_SVCrsE)

Face Hallucination - SRGAN Trained on CelebA Database

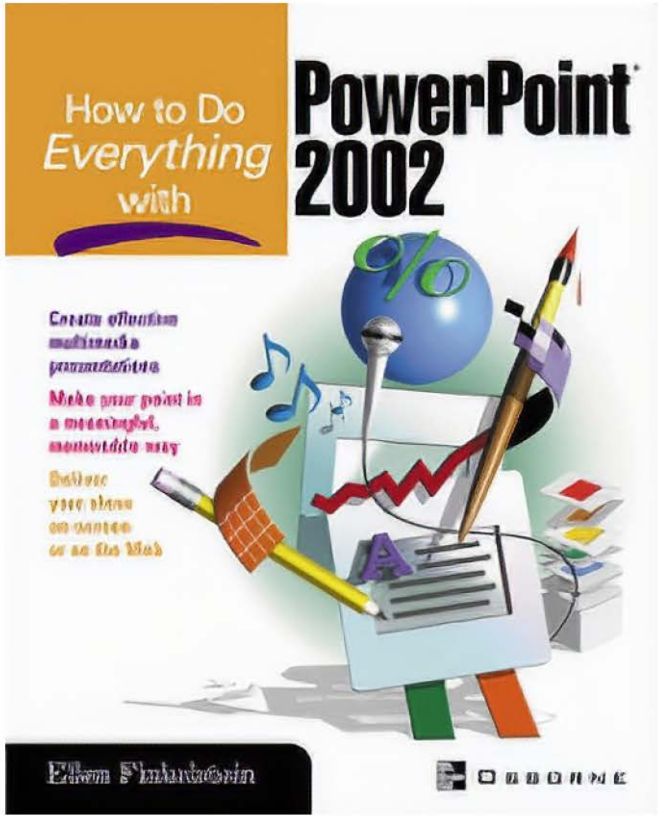


From the recorded CVPR presentation: https://www.youtube.com/watch?v=BXIR_SVCrsE

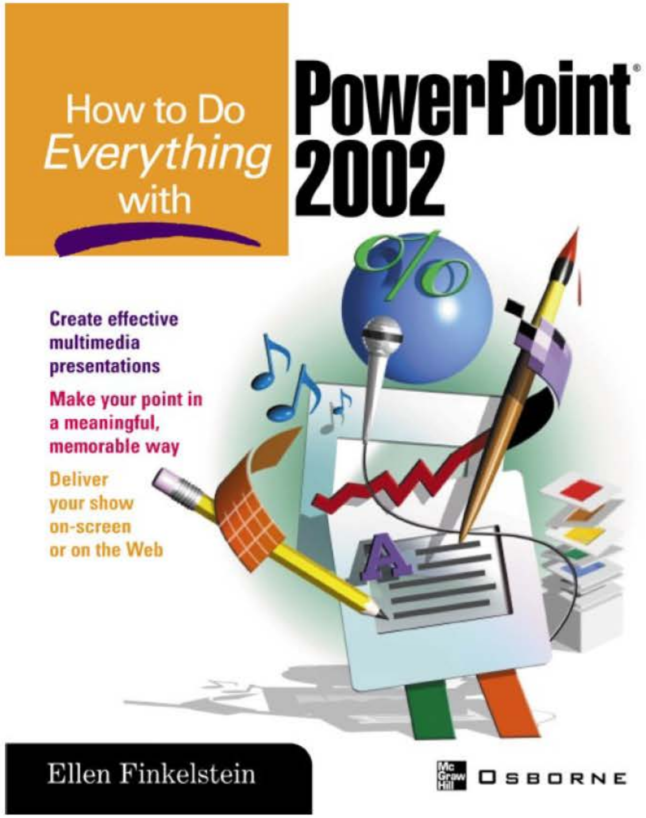
Disadvantages – Training Data Limitation

Bicubic 4X upscaling

SRGAN 4X Upscaling



Original



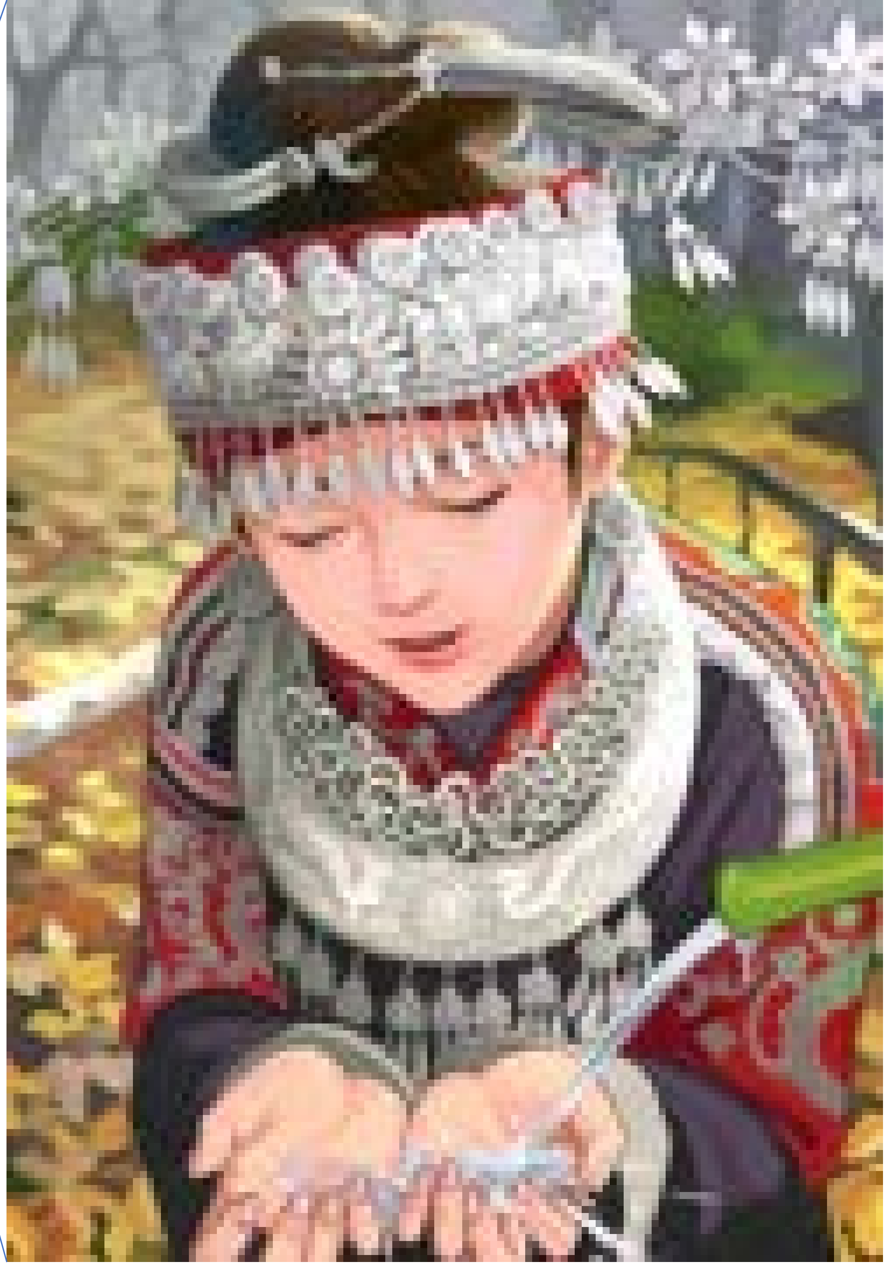
- Since the network was trained on ImageNet which doesn't contain text or numbers it fails to accurately reconstruct text and numbers.

Encoder-Decoder Techniques

- GAN and RCAN-based methods work on *any* single LR image. The only requirement is that the network have been trained on similar images.

- Encoder-Decoder paradigm generates the LR image using a specially designed deep network encoder and upscales the LR image using the matched deep network decoder.
    - If the decoder network is used to upscale a LR image that has not been downscaled by the corresponding encoder, the results could be quite disappointing.
    - Encoder-Decoder techniques allow transmission and storage of LR images that can be upscaled efficiently with the matched generic decoder without the need for additional information.

LR 2X downscaled using conventional resolution reduction techniques



Original Image

LR Image Upscaled with Bi-Cubic Interpolator

Original Image

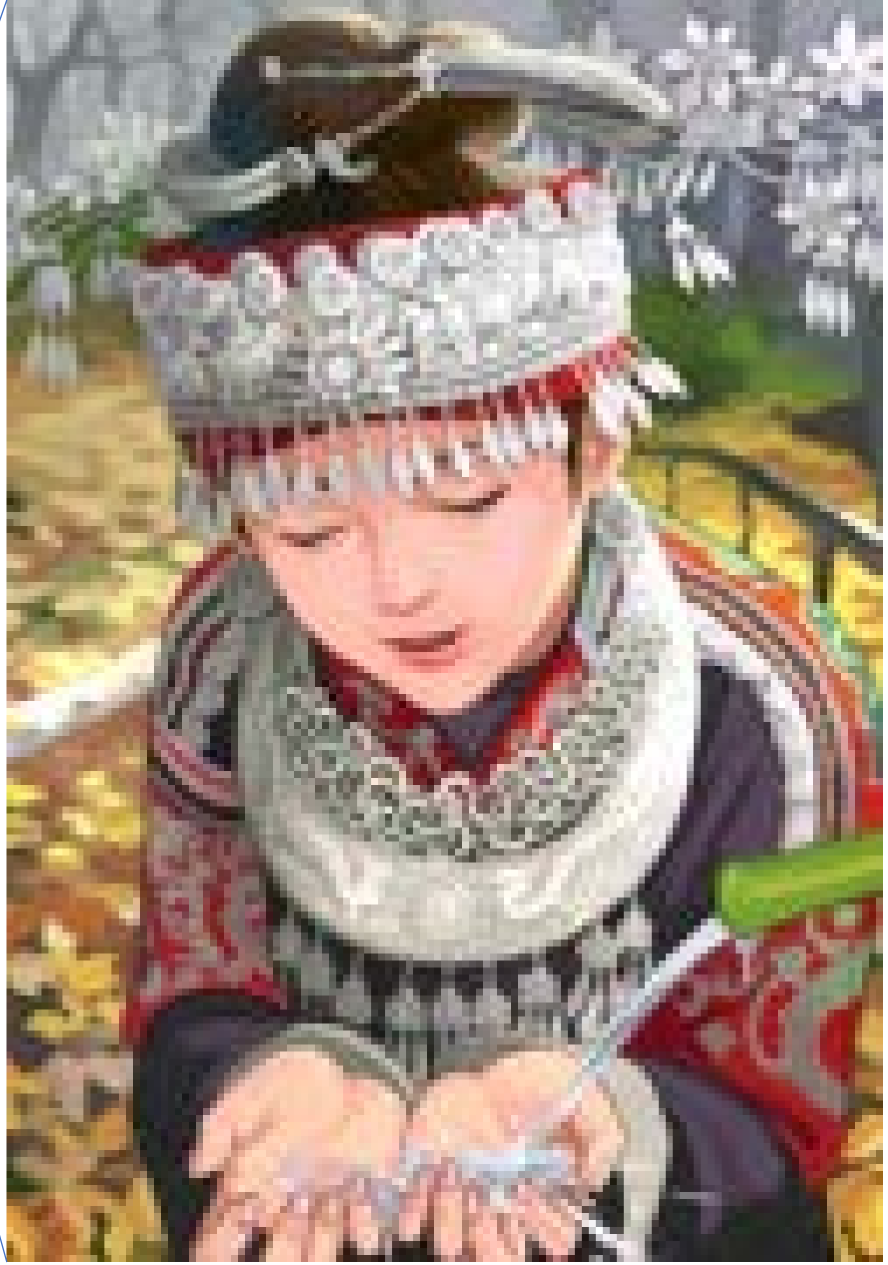
LR 2X downscaled using conventional resolution reduction techniques

LR 2X downscaled
using IRN network
with 1.66M weights
trained w. 800 images

Original Image

Original Image

LR 2X downscaled using IRN network with 1.66M weights trained w. 800 images



IRN LR Image Upscaled with Bi-Cubic Interpolator

Original Image

LR 2X downscaled
using IRN network
with 1.66M weights
trained w. 800 images

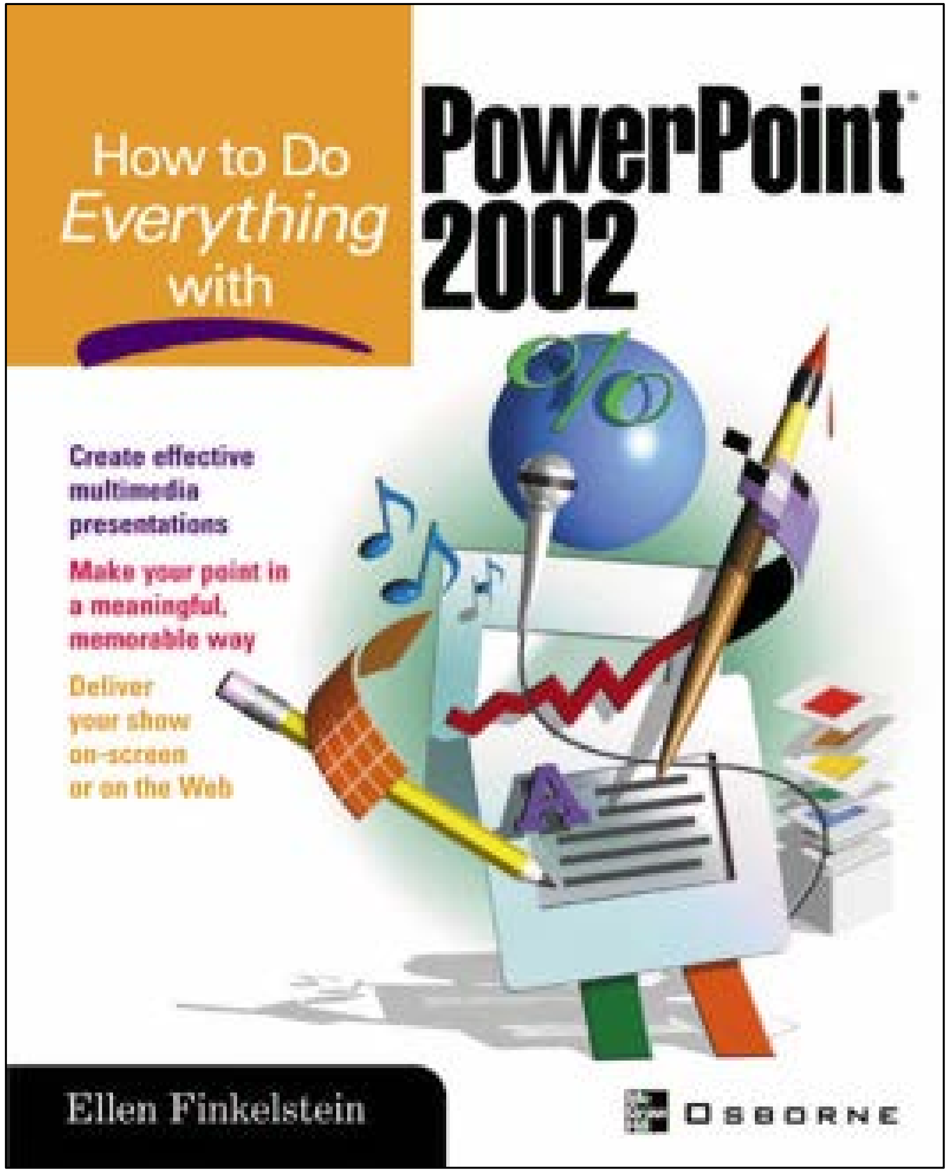
IRN LR image upscaled by IRN inverse network
(PSNR=34.057723 dB/SSIM=0.979321)

Original Image

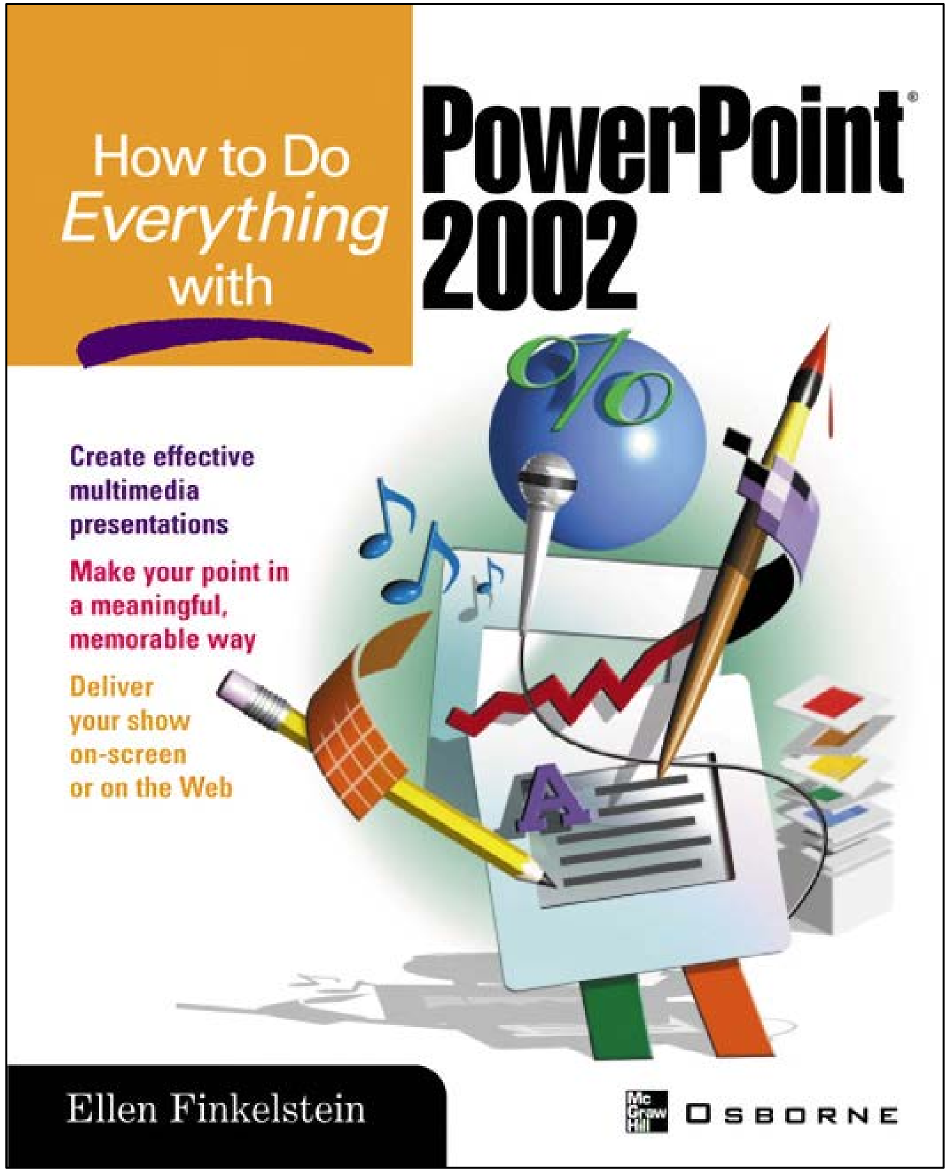
LR 2X downscaled using conventional resolution reduction techniques

Ordinary LR image upscaled by IRN inverse network (PSNR=26.609975 dB/SSIM=0.882251)

IRN Example 2X Upscaling Example

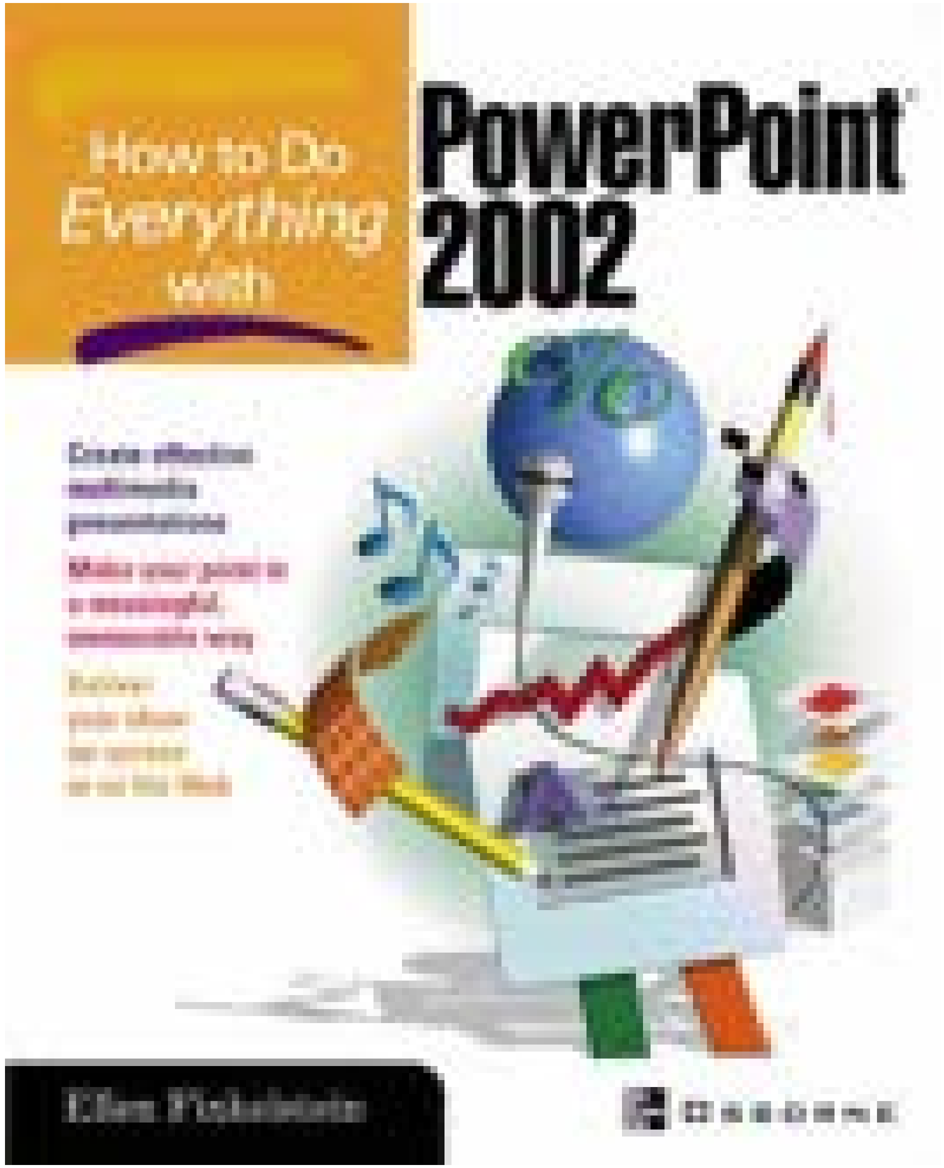


IRN Generated Image 2X Upscaled by Traditional Bi-Cubic Interpolation

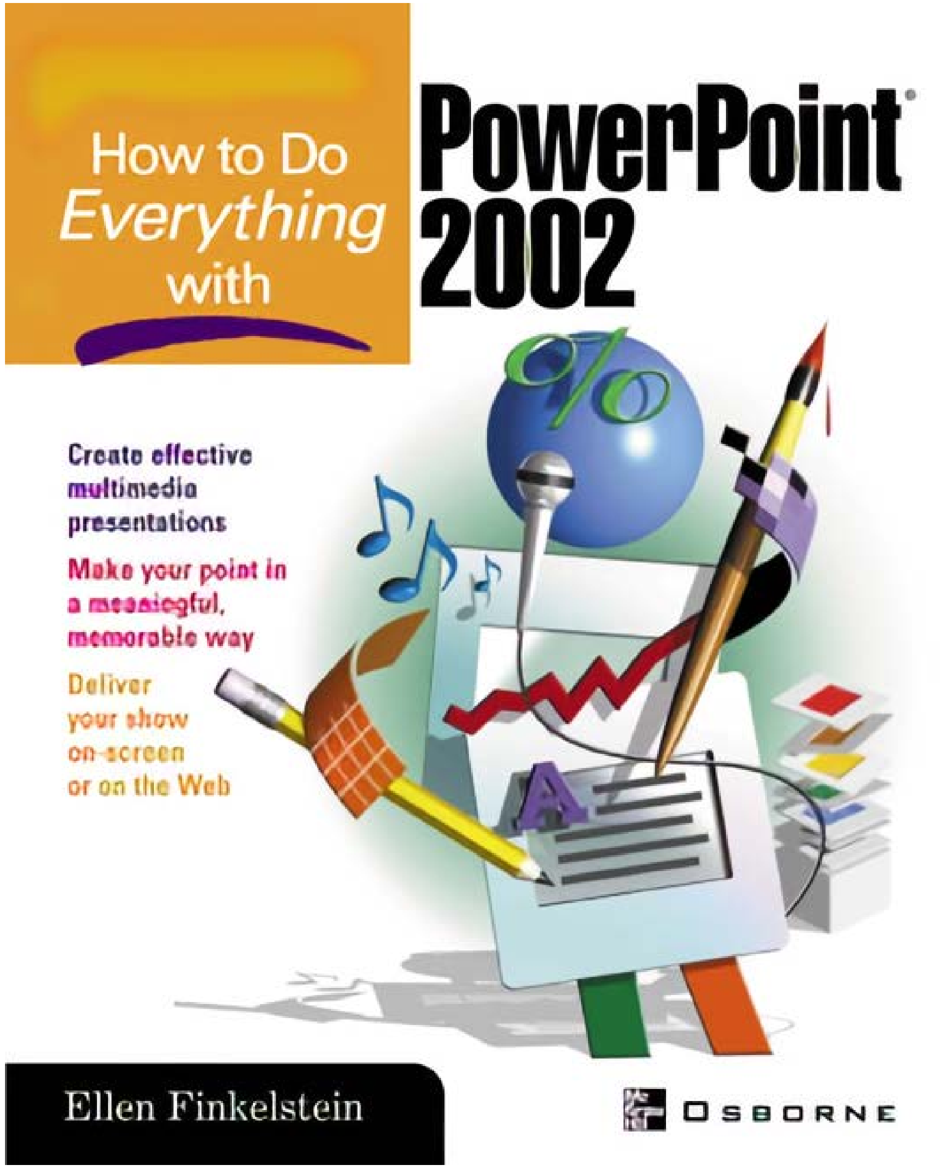


IRN Generated Image 2X Upscaled by the Matched Decoder

IRN Example 4X Upscaling Example

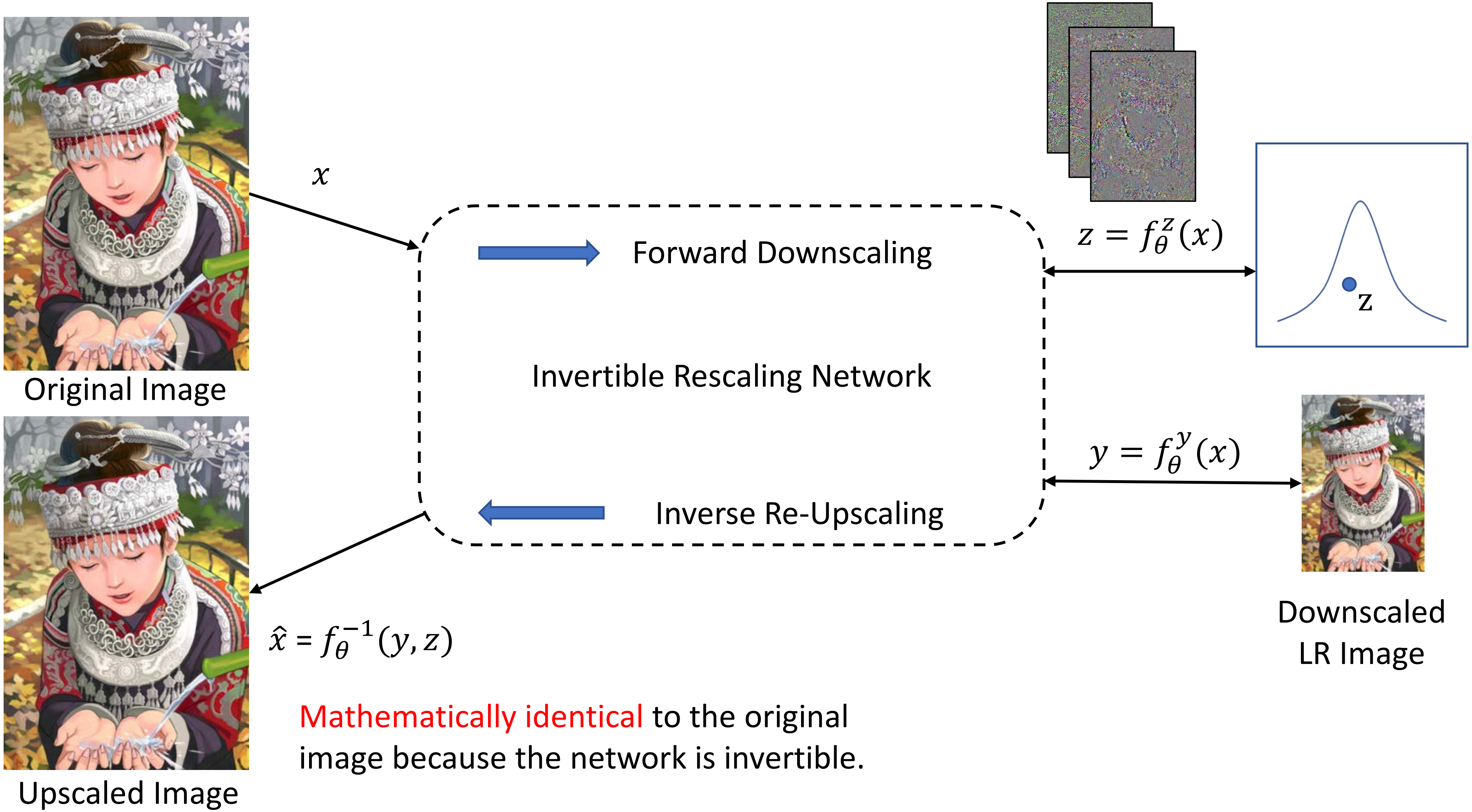


IRN Generated Image 4X Upscaled by Traditional Bi-Cubic Interpolation



IRN Generated Image 4X Upscaled by the Matched Decoder

Invertible Image Rescaling (IRN)



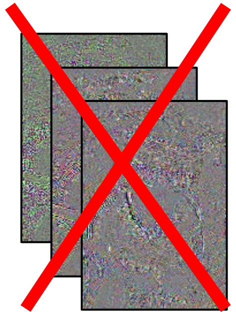
Mathematically identical to the original image because the network is invertible.

Invertible Image Rescaling (IRN)

Original Image

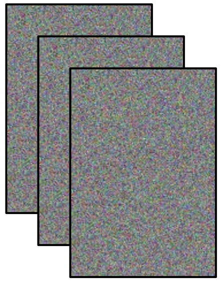
x

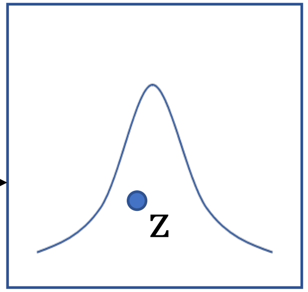
The case-specific high-frequency image z is replaced with a case-agnostic latent variable z that can be generated via a Gaussian distribution with no storage overhead.



Case-agnostic

$z \sim \mathcal{N}(0, I_k)$





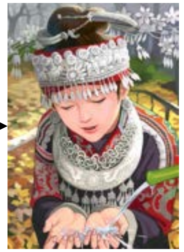
$z = f_\theta^z(x)$

Forward Downscaling

Invertible Rescaling Network

Inverse Re-Upscaling

$y = f_\theta^y(x)$



Downscaled LR Image

Upscaled Image

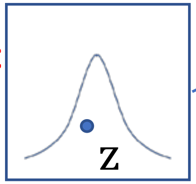
$\hat{x} = f_\theta^{-1}(y, z)$

Although not mathematically identical to the original image, perceptually it contains all the high-frequency information.

Original Image

LR 2X downscaled
using IRN network
with 1.66M weights
trained w. 800 images

Case−agnostic
$z \sim \mathcal{N}(0, I_k)$



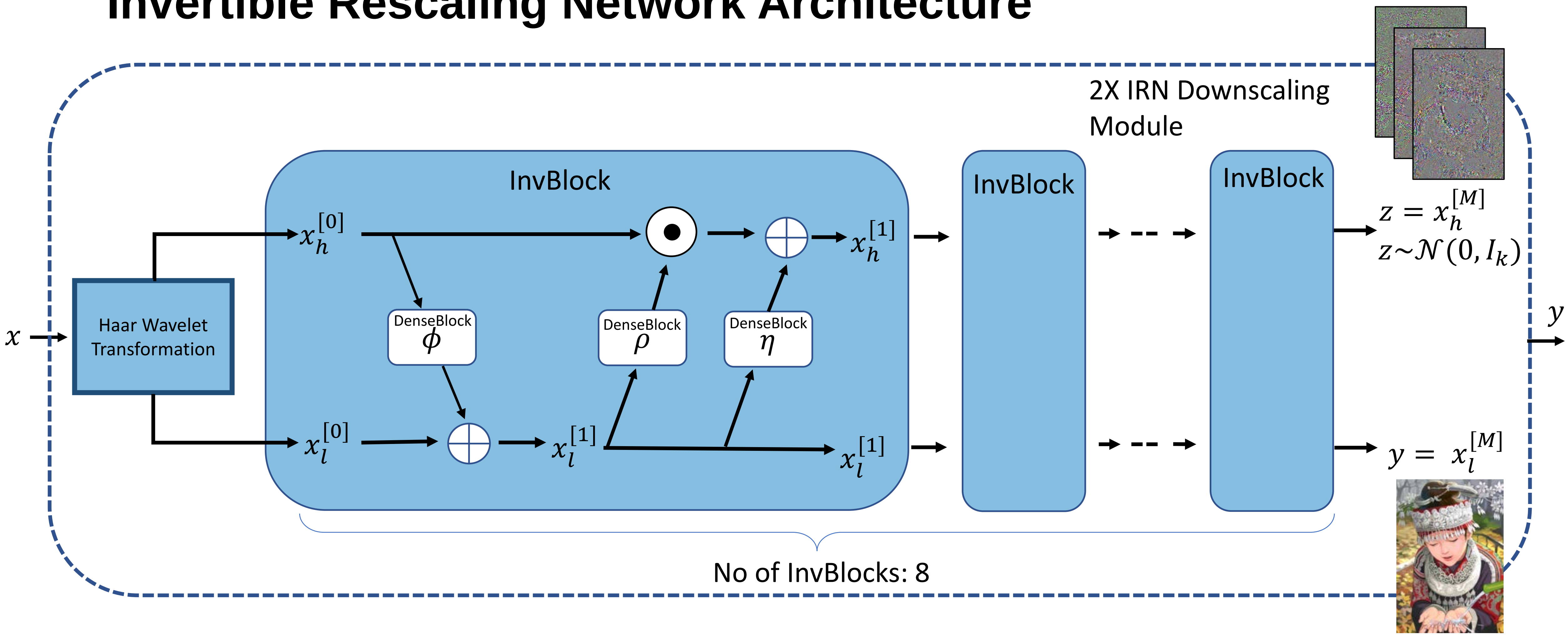
IRN LR image upscaled by IRN inverse network
(PSNR=34.057723 dB/SSIM=0.979321)

LR 2X downscaled
using IRN network
with 1.66M weights
trained w. 800 images

Original Image

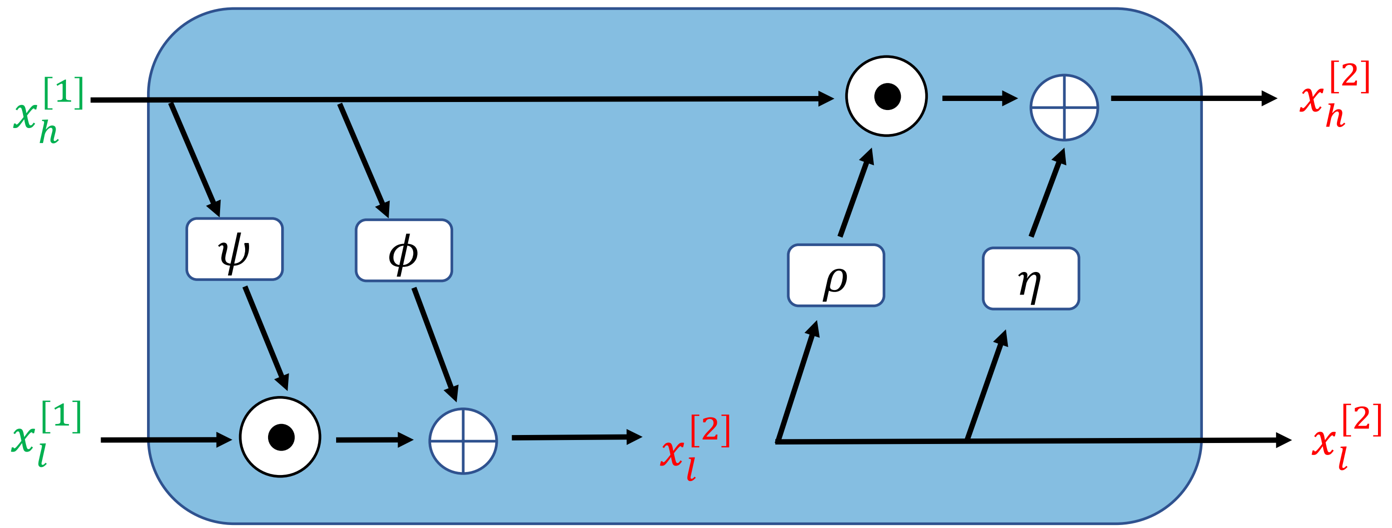
IRN LR Image Upscaled with Bi-Cubic Interpolator

Invertible Rescaling Network Architecture



x: HR image, y: LR image, z: case-agnostic variable, I_k: identity *covariance* matrix, k: dimensions of z, M=8, Grayscale HR image x of size 1024×1024, produces $x_l^{[0]}$, … $x_l^{[M]}$ of dim 512×512 and $x_h^{[0]}$, … $x_h^{[M]}$ of dim size $512 \times 512 \times 3$, ϕ, ρ and η are Neural Networks.

Invertible Neural Network[11]



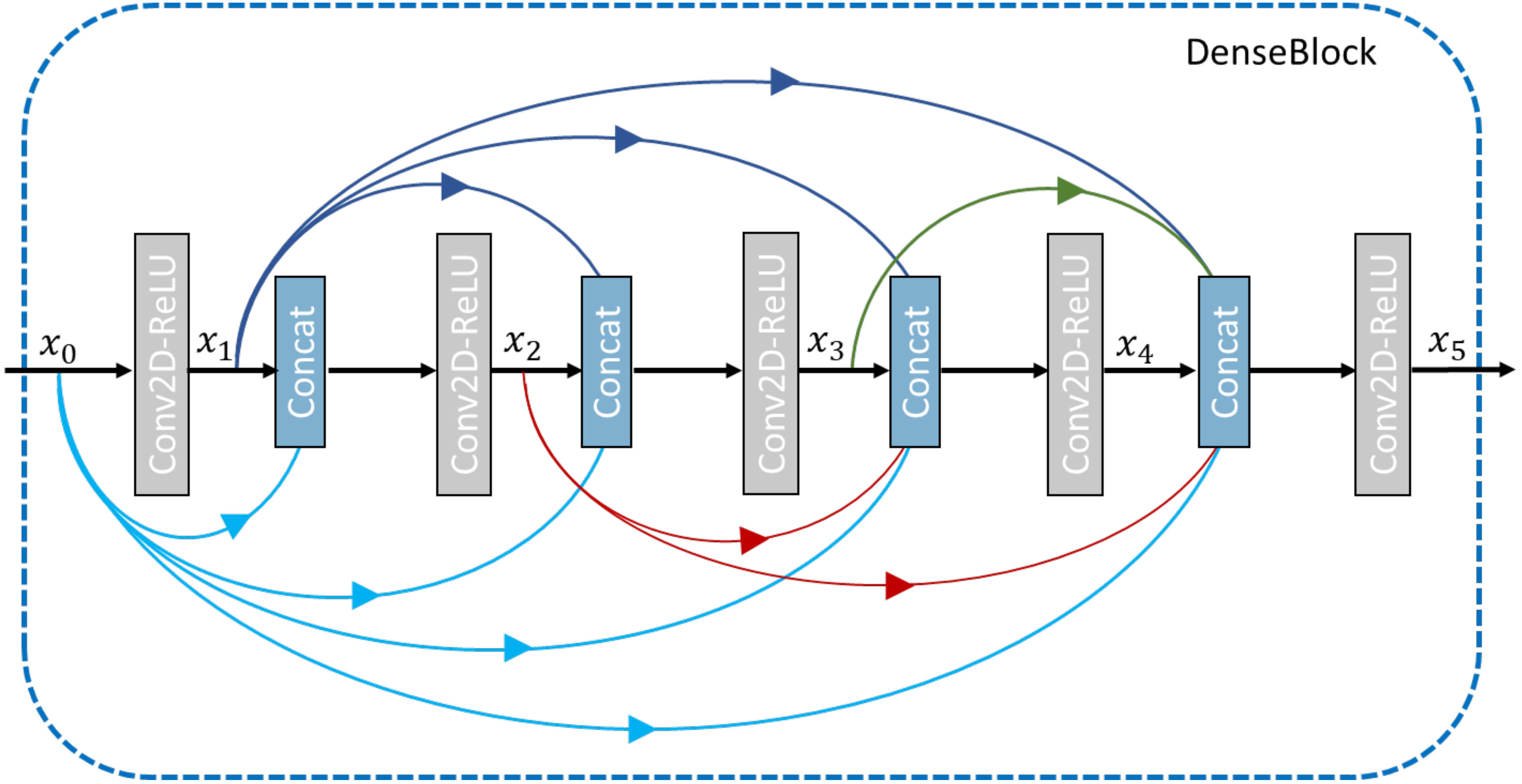
- Element-wise product operator: $\odot$
- The augmentation of $\odot$ $\exp(.)$ helps to enhance the transformation ability of the network.

$$\left\{\begin{aligned} x_l^{[2]} &= x_l^{[1]} \odot \exp(\psi(x_h^{[1]})) + \phi(x_h^{[1]}) \\ x_h^{[2]} &= x_h^{[1]} \odot \exp(\rho(x_l^{[2]})) + \eta(x_l^{[2]}) \end{aligned}\right. \tag{3}$$

$$\left.\begin{aligned} x_h^{[1]} &= (x_h^{[2]} - \eta(x_l^{[2]})) \odot \exp(-\rho(x_l^{[2]})) \\ x_l^{[1]} &= (x_l^{[2]} - \phi(x_h^{[1]})) \odot \exp(-\psi(x_h^{[1]})) \end{aligned}\right\} \tag{4}$$

11. *Dinh, L., Sohl-Dickstein, J., Bengio, S., "Density estimation using real NVP," Proceedings of the International Conference on Learning Representations (2017)*

Dense Block Architecture[12]: $\phi(.), \rho(.), \eta(.)$




```
        Phi Dense Block
(InputDim, NumFilters, [Filtersize])
conv1.weight 2592   (9 × 32 × [3×3])
conv1.bias 32
conv2.weight 11808  ((9+32)×32×[3×3])
conv2.bias 32
conv3.weight 21024  ((9+2*32)×32×[3×3])
conv3.bias 32
conv4.weight 30240  ((9+3*32)×32×[3×3])
conv4.bias 32
conv5.weight 3699   ((9+4*32)×3×[3×3])
conv5.bias 3
Total Parameters = 69494
```

```
        Eta & Rho Dense Block
(InputDim, NumFilters, [Filtersize])
conv1.weight 864    (3 × 32 × [3×3])
conv1.bias 32
conv2.weight 10080  ((3+32)×32×[3×3])
conv2.bias 32
conv3.weight 19296  ((3+2*32)×32×[3×3])
conv3.bias 32
conv4.weight 28512  ((3+3*32)×32×[3×3])
conv4.bias 32
conv5.weight 10611  ((3+4*32)×9×[3×3])
conv5.bias 9
Total Parameters=  69500
```

$\phi(.)$ = 69,494; $\rho(.)$=69,500; $\eta(.)$ = 69,500; InvBlock = 208,494; Total = 1.66M

12. *Huang, G., Liu, Z., Weinberger, K.Q., van der Maaten, L., "Densely connected convolutional networks," In: CVPR. (2017)*

Quantitative Results (Table 1)

Downscaling & Upscaling	Scale	Param	Set5	Set14	BSD100	Urban100	DIV2K
Bicubic & Bicubic	2×	/	33.66 / 0.9299	30.24 / 0.8688	29.56 / 0.8431	26.88 / 0.8403	31.01 / 0.9393
Bicubic & SRCNN	2×	57.3K	36.66 / 0.9542	32.45 / 0.9067	31.36 / 0.8879	29.50 / 0.8946	–
Bicubic & EDSR	2×	40.7M	38.20 / 0.9606	34.02 / 0.9204	32.37 / 0.9018	33.10 / 0.9363	35.12 / 0.9699
Bicubic & RDN	2×	22.1M	38.24 / 0.9614	34.01 / 0.9212	32.34 / 0.9017	32.89 / 0.9353	–
Bicubic & RCAN	2×	15.4M	38.27 / 0.9614	34.12 / 0.9216	32.41 / 0.9027	33.34 / 0.9384	–
Bicubic & SAN	2×	15.7M	38.31 / 0.9620	34.07 / 0.9213	32.42 / 0.9028	33.10 / 0.9370	–
TAD & TAU	2×	–	38.46 / –	35.52 / –	36.68 / –	35.03 / –	39.01 / –
CNN-CR & CNN-SR	2×	–	38.88 / –	35.40 / –	33.92 / –	33.68 / –	–
CAR & EDSR	2×	51.1M	38.94 / 0.9658	35.61 / 0.9404	33.83 / 0.9262	35.24 / 0.9572	38.26 / 0.9599
IRN (Paper)	2×	1.66M	43.99 / 0.9871	40.79 / 0.9778	41.32 / 0.9876	39.92 / 0.9865	44.32 / 0.9908
IRN (My Results)	2×	1.66M	43.997/0.9871	40.788/0.9777	41.322/0.9875	39.918/0.9865	44.324/0.9908
Bicubic & Bicubic	4×	/	28.42 / 0.8104	26.00 / 0.7027	25.96 / 0.6675	23.14 / 0.6577	26.66 / 0.8521
Bicubic & SRCNN	4×	57.3K	30.48 / 0.8628	27.50 / 0.7513	26.90 / 0.7101	24.52 / 0.7221	–
Bicubic & EDSR	4×	43.1M	32.62 / 0.8984	28.94 / 0.7901	27.79 / 0.7437	26.86 / 0.8080	29.38 / 0.9032
Bicubic & RDN	4×	22.3M	32.47 / 0.8990	28.81 / 0.7871	27.72 / 0.7419	26.61 / 0.8028	–
Bicubic & RCAN	4×	15.6M	32.63 / 0.9002	28.87 / 0.7889	27.77 / 0.7436	26.82 / 0.8087	30.77 / 0.8460
Bicubic & ESRGAN	4×	16.3M	32.74 / 0.9012	29.00 / 0.7915	27.84 / 0.7455	27.03 / 0.8152	30.92 / 0.8486
Bicubic & SAN	4×	15.7M	32.64 / 0.9003	28.92 / 0.7888	27.78 / 0.7436	26.79 / 0.8068	–
TAD & TAU	4×	–	31.81 / –	28.63 / –	28.51 / –	26.63 / –	31.16 / –
CAR & EDSR	4×	52.8M	33.88 / 0.9174	30.31 / 0.8382	29.15 / 0.8001	29.28 / 0.8711	32.82 / 0.8837
IRN (Paper)	4×	4.35M	36.19 / 0.9451	32.67 / 0.9015	31.64 / 0.8826	31.41 / 0.9157	35.07 / 0.9318
IRN (My Results)	4×	4.35M	36.191/0.9451	32.667/0.9015	31.638/0.8825	31.406/0.9156	35.071/0.9318

SRCNN [13], EDSR [14], RDN [15], RCAN [9], SAN [16], TAD & TAU [17], CNN-CR & CNN-SR [18], CAR & EDSR [19]

Results On Our Own Images Not In The Training or Testing Set

- Prasanna has implemented this algorithm as part of his qualifier exam and confirmed the reported results. He has also generated the above results on datasets not reported in the paper.
- Link to the code and Prasanna's results: https://drive.google.com/drive/folders/1OtIcPNhjchZX683MCdjNv-RdYJQGPG40?usp=sharing

Downscaling & Upscaling	Scale	Param	T91	manga109	BSDS200	General100	Open_Images
IRN (My Results)	2×	1.66M	41.911/0.9852	43.684/0.9926	42.276/0.9894	44.808/0.9920	45.999/0.9922
IRN (My Results)	4 ×	4.35M	34.727/0.9261	35.938/0.9615	32.498/0.9022	36.403/0.9457	37.113/0.9444

Open Image Dataset - Original Image

2X IRN Downscaling/Upscaling

4X IRN Downscaling/Upscaling

Results On Our Own Images Not In The Training or Testing Set

- Prasanna has implemented this algorithm as part of his qualifier exam and confirmed the reported results. He has also generated the above results on datasets not reported in the paper.
- Link to the code and Prasanna's results: https://drive.google.com/drive/folders/1OtIcPNhjchZX683MCdjNv-RdYJQGPG40?usp=sharing

2X Bicubic Method

2X IRN Downscaling/Upscaling

Cropped Original Image

References

1. *R. Tsai, T. Huang, “Multi-frame image restoration and registration,” Advances in Computer Vision and Image Processing,” (1), no. 2, 1984, pp. 317-339*
2. *S. Baker and T. Kanade, “Hallucinating faces,” In IEEE International Conference on Automatic Face and Gesture Recognition, March 2000.S. C. Park, et al, “Super-resolution image reconstruction: A technical overview,” IEEE Signal Processing Magazine, pp. 21-36, **20**(3), (2003)..*
3. *S. C. Park, M. K. Park, and M. G. Kang, “Super-resolution image reconstruction: A technical overview,” IEEE Signal Processing Magazine, pp. 21-36, **20**(3), May 2003.*
4. *Lin and Shum, “Fundamental limits of reconstruction-based super-resolution algorithms under local translation”, IEEE Trans. PAMI, **26**(1), pp. 83–97, (2004)*
5. *Farsiu, Robinson, Elad, and Milanfar, “Fast and Robust Multiframe Super Resolution”, IEEE TIP, (13), No. 10, pp. 1327-1344, (2004).*
6. *Ian J. Goodfellow, et al, “Generative adversarial nets,” In Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'14). MIT Press, Cambridge, MA, USA, 2672–2680, (2014)*
7. *C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," 2017 IEEE Conference on CVPR, Honolulu, HI, USA, 2017, pp. 105-114, doi: 10.1109/CVPR.2017.19.*
8. *O. Russakovsky, et al. “Imagenet large scale visual recognition challenge,” International Journal of Computer Vision, pages 1–42, (2014).*
9. *Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: “Image super-resolution using very deep residual channel attention networks,” In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 286–301 (2018)*
10. *M. Xiao, et al, “Invertible image rescaling,” ArXiv, vol. abs/2005.05650, (2020).*
11. *Dinh, L., Sohl-Dickstein, J., Bengio, S., “Density estimation using real NVP,” Proceedings of the International Conference on Learning Representations (2017)*
12. *Huang, G., Liu, Z., Weinberger, K.Q., van der Maaten, L., “Densely connected convolutional networks,” In: CVPR. (2017)*
13. *Dong, C., Loy, C.C., He, K., Tang, X., “Image super-resolution using deep convolutional networks,” IEEE Trans. PAMI, 38(2), 295–307 (2015)*
14. *Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K., “Enhanced deep residual networks for single image super-resolution,” In: Proceedings of the IEEE conference on CVPR workshops. pp. 136–144 (2017)*
15. *Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y., “Residual dense network for image super resolution,” In: Proceedings of the IEEE Conference on CVPR, pp. 2472–2481 (2018)*
16. *Dai, T., Cai, J., Zhang, Y., Xia, S.T., Zhang, L., “Second-order attention network for single image super-resolution,” In: Proceedings of the IEEE Conference on CVPR, pp. 11065–11074 (2019)*
17. *Kim, H., Choi, M., Lim, B., Mu Lee, K, “Task-aware image downscaling,” In: Proceedings of (ECCV). pp. 399–414 (2018)*
18. *Li, Y., Liu, D., Li, H., Li, L., Li, Z., Wu, F., “Learning a convolutional neural network for image compact resolution,” IEEE TIP, 28(3), 1092–1107 (2018)*
19. *Sun, W., Chen, Z., “Learned image downscaling for upscaling using content adaptive resampler,” IEEE TIP 29, 4027–4040 (2020)*

Thank You!

Face Hallucination

Inferring a high-resolution face image from a low-resolution input.

(a) Input 24×32 (b) Hallucinated result (c) Original 96×128

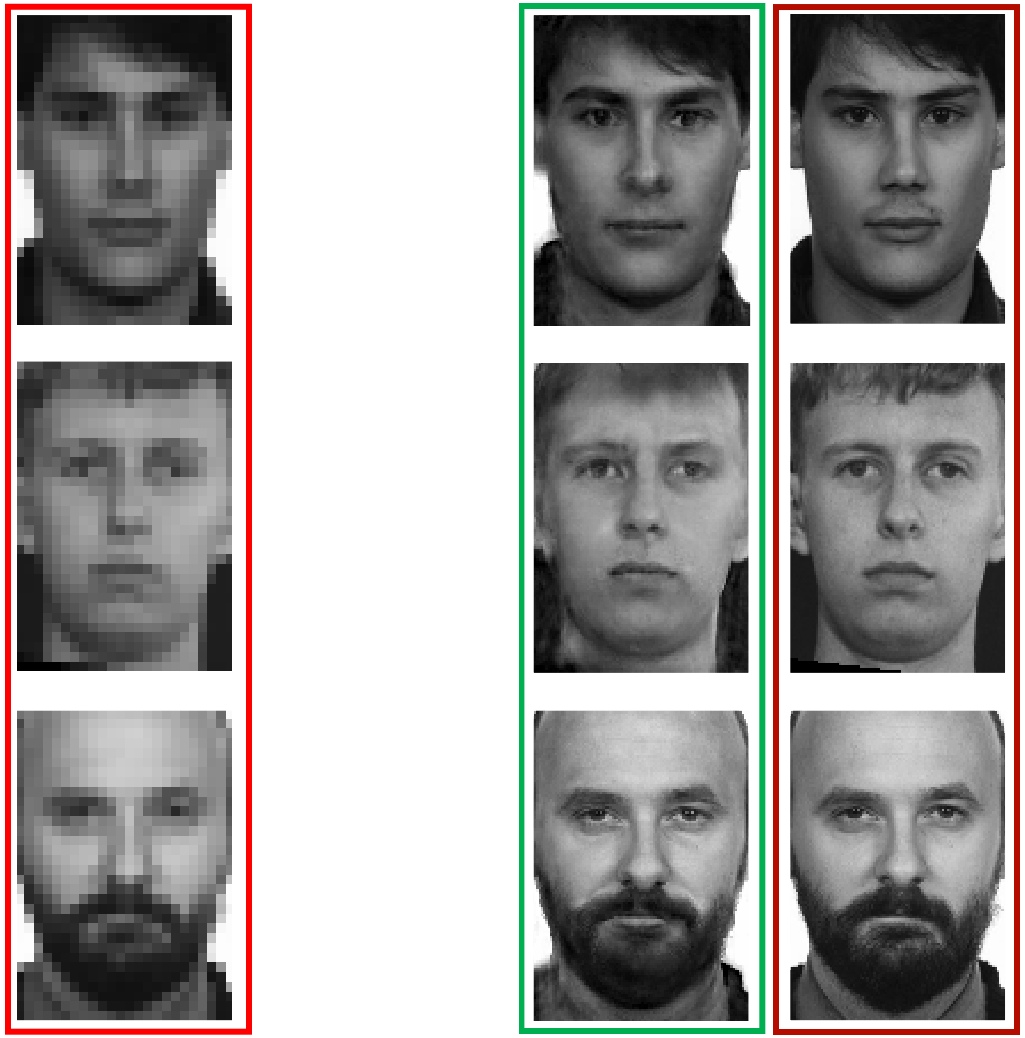
http://people.csail.mit.edu/celiu/FaceHallucination/fh.html

Face Hallucination - How does it work?

- Needs a large collection of other high-resolution face images.
- The theoretical contribution is a two-step statistical modeling that integrates both a global parametric model (generalizes well with common faces), with a local nonparametric model (learns local textures from example faces). The hallucinated face is the maximum *a posteriori* (MAP) solution.
- The practical contribution is a robust warping algorithm to align the LR face images, which is an extremely difficult problem. The LR faces are aligned by finding an affine transform, determined from an eigenspace representation, to warp the input image to a template that maximizes the probability of the captured LR face image.

http://people.csail.mit.edu/celiu/FaceHallucination/fh.html

Face Hallucination



a) Input LR face 24×32

b) Inferred global face

c) Hallucinated result

d) Original HR face 96×128

http://people.csail.mit.edu/celiu/FaceHallucination/fh.html

Invertible Networks for Super Resolution

Presented by Prasanna Reddy Pulakurthi

- Xiao[10] et al. propose an invertible architecture that generates a visually pleasing LR image and embeds the high-frequency content into a high-frequency image z, where all the values are independent Gaussian distributed with mean zero and variance of 1.

- The main contribution of the paper is to replace the case-specific high-frequency image z with a case-agnostic latent variable z that can be generated via a Gaussian distribution with no storage overhead. This is accomplished by using a novel set of loss functions in training the neural network.

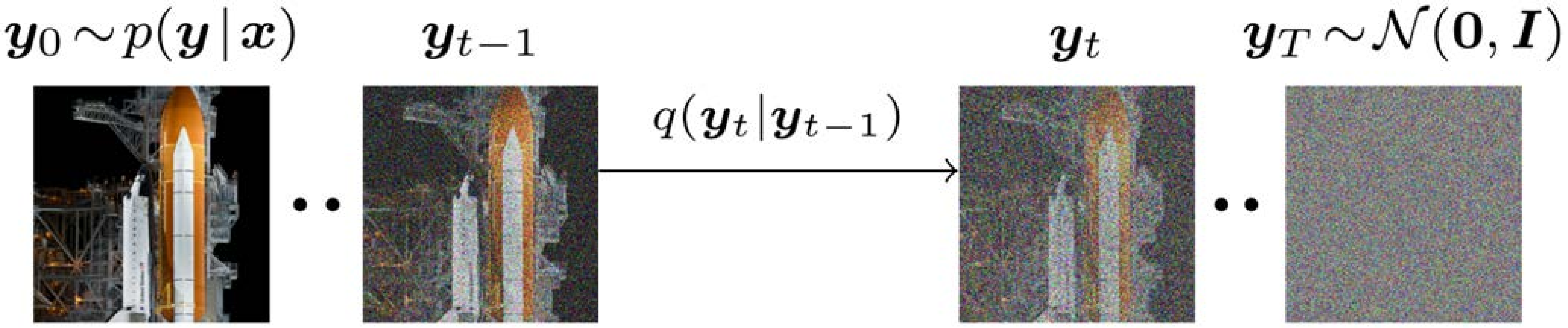
- HR reconstruction is lossy but achieves state-of-the-art visual performance.

10. M. Xiao, et al. "Invertible image rescaling," ArXiv, vol. abs/2005.05650, 2020.

SR3: Image Super-Resolution via Iterative Refinement

https://iterative-refinement.github.io/

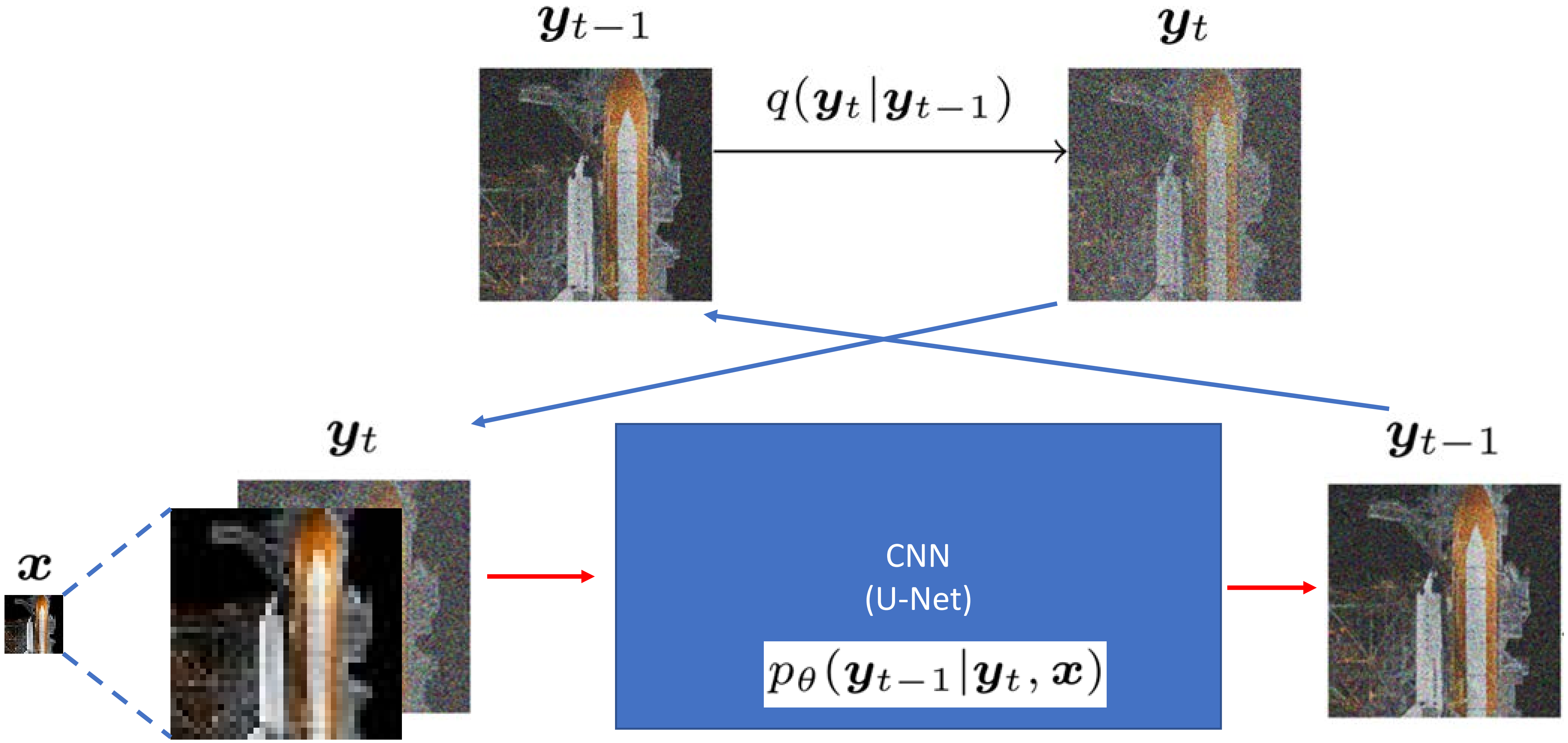
Forward Diffusion Process



$$q(\boldsymbol{y}_{1:T} \mid \boldsymbol{y}_0) = \prod_{t=1}^{T} q(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1})$$

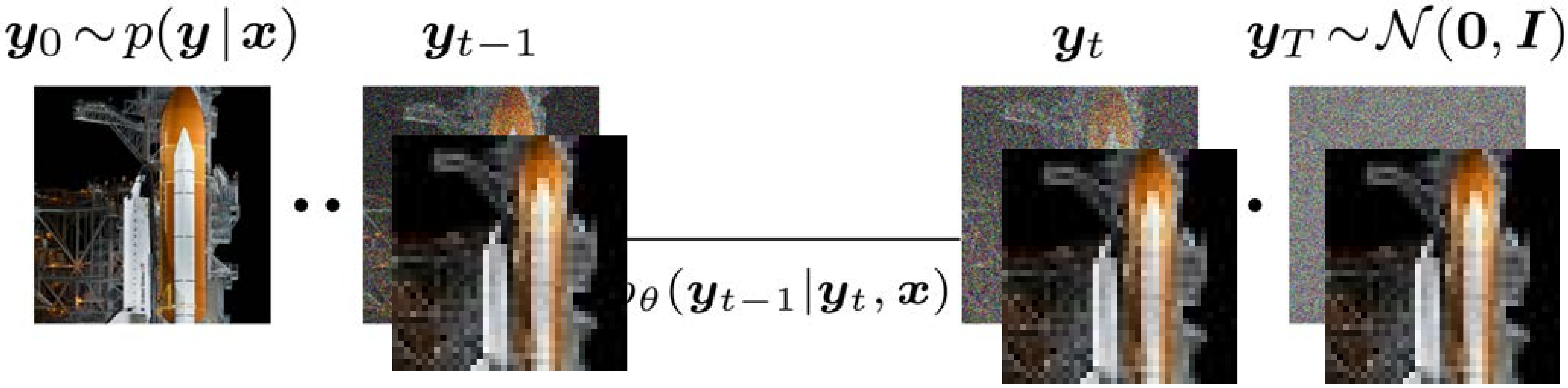
Saharia, Chitwan, et al. "Image super-resolution via iterative refinement." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).

Training the Model



Saharia, Chitwan, et al. "Image super-resolution via iterative refinement." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).

Reverse Diffusion Process



Saharia, Chitwan, et al. "Image super-resolution via iterative refinement." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).

Architecture

Task	Channel Dim	Depth Multipliers	# ResNet Blocks	# Parameters
$16 \times 16 \rightarrow 128 \times 128$	128	$\{1, 2, 4, 8, 8\}$	3	550M
$64 \times 64 \rightarrow 256 \times 256$	128	$\{1, 2, 4, 4, 8, 8\}$	3	625M
$64 \times 64 \rightarrow 512 \times 512$	64	$\{1, 2, 4, 8, 8, 16, 16\}$	3	625M
$256 \times 256 \rightarrow 1024 \times 1024$	16	$\{1, 2, 4, 8, 16, 32, 32, 32\}$	2	150M

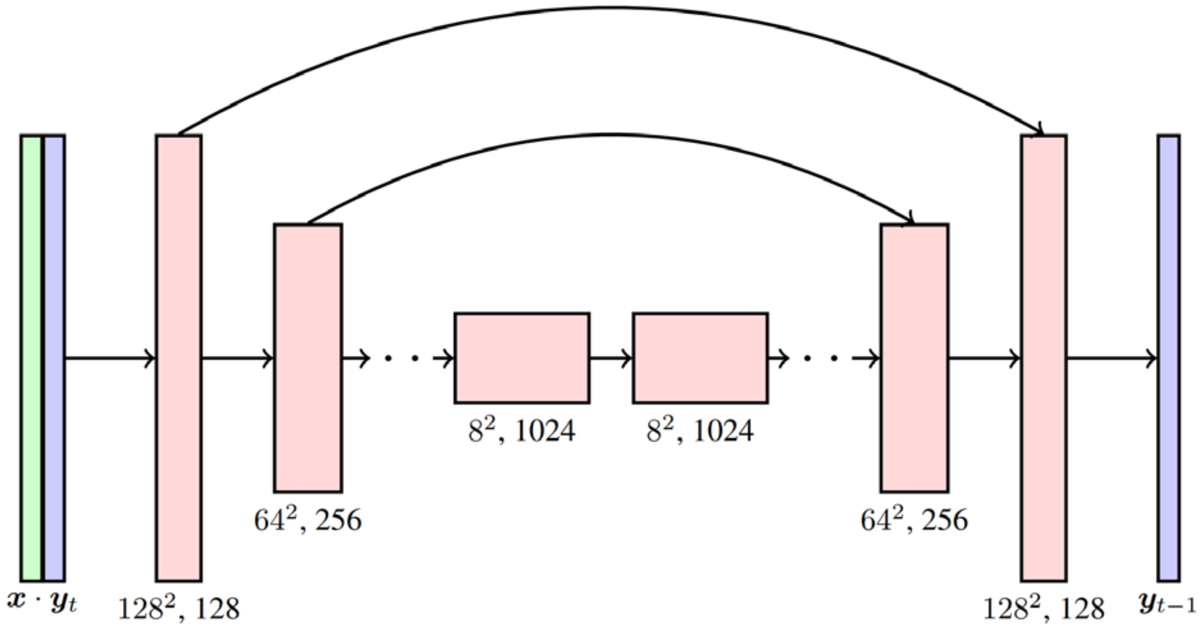
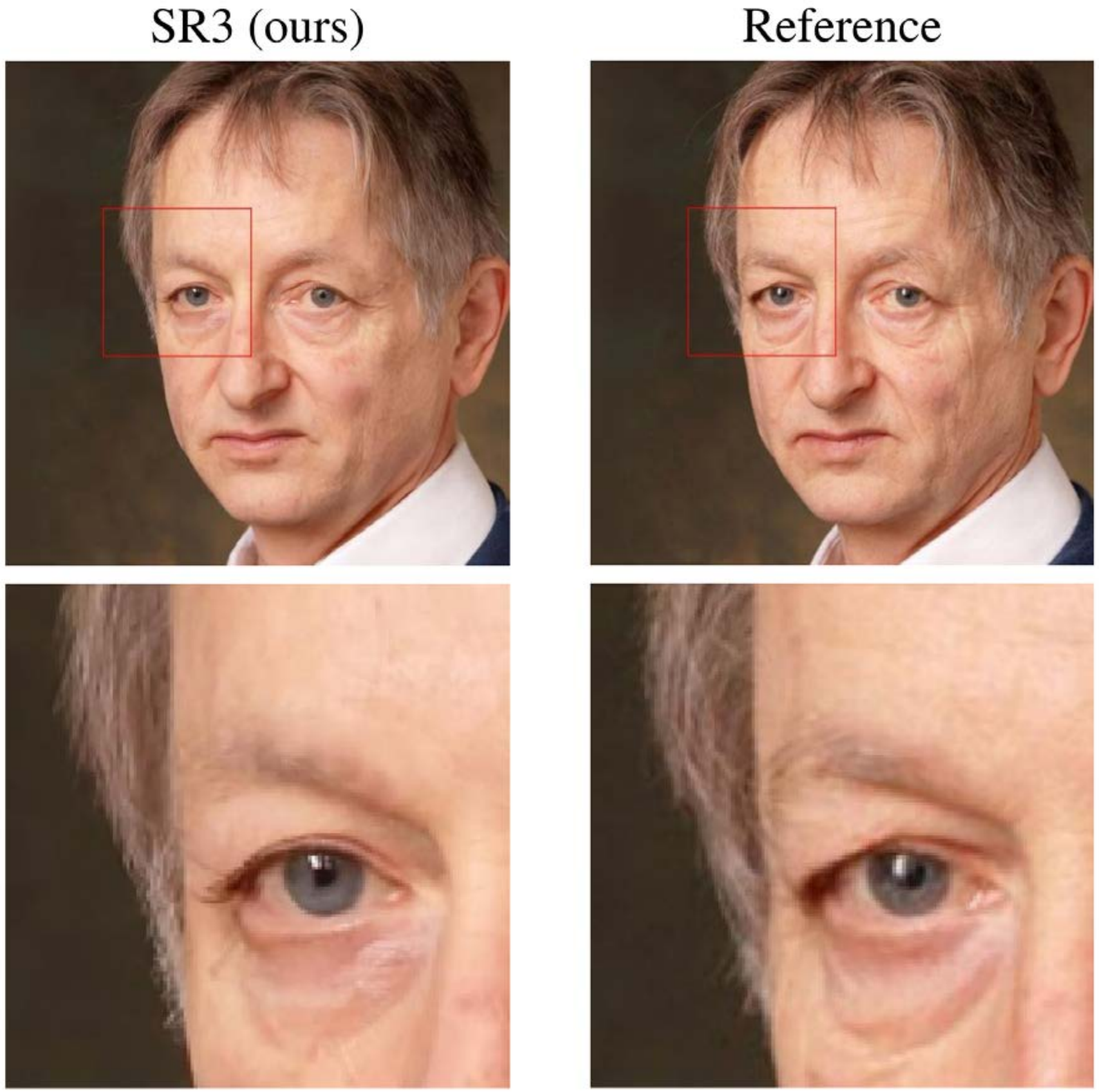


Figure A.1: Description of the U-Net architecture with skip connections. The low resolution input image $\boldsymbol{x}$ is interpolated to the target high resolution, and concatenated with the noisy high resolution image $\boldsymbol{y}_t$. We show the activation dimensions for the example task of $16{\times}16 \rightarrow 128{\times}128$ super resolution.

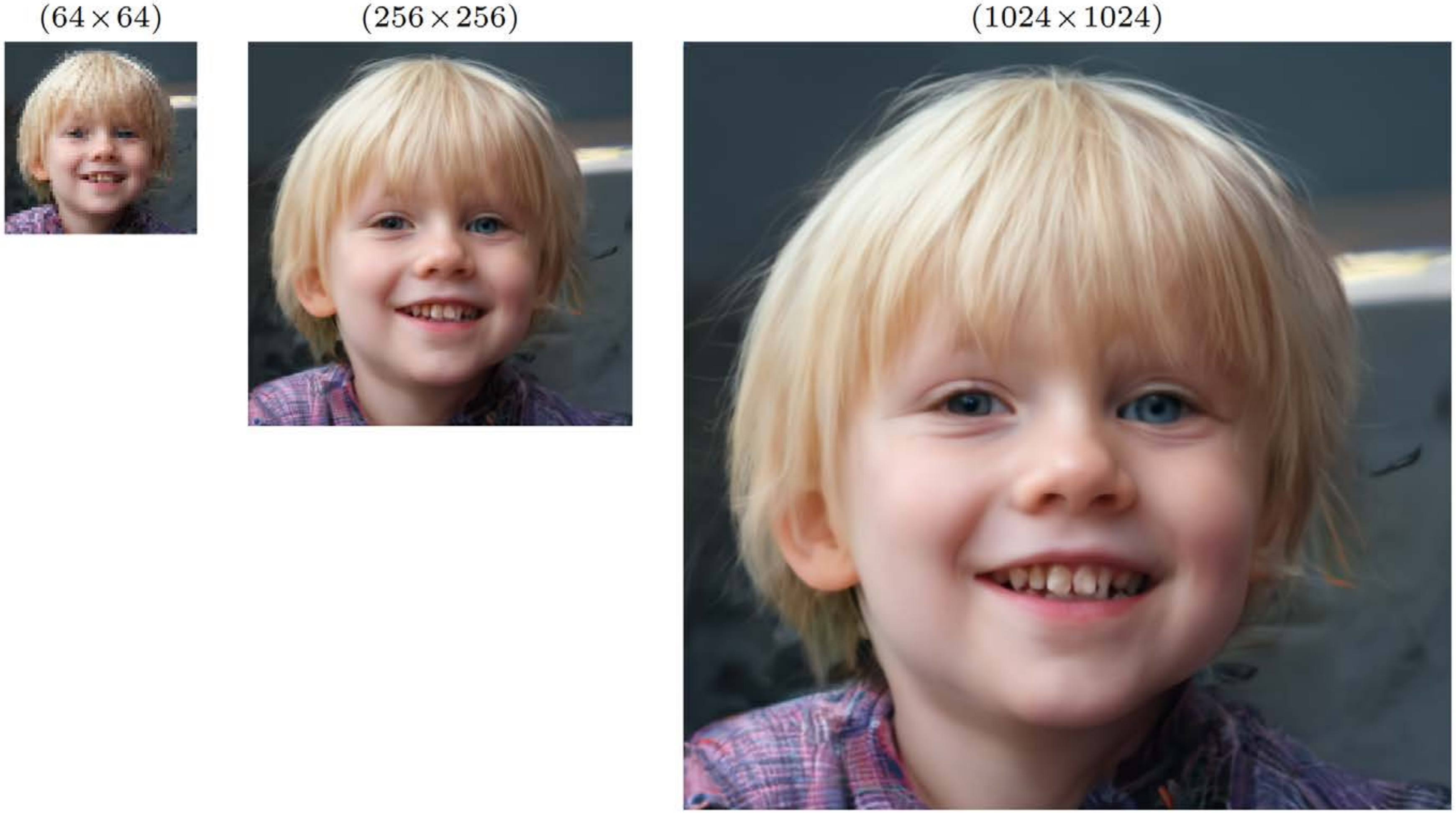
Saharia, Chitwan, et al. "Image super-resolution via iterative refinement." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).

Results of SR3 model (64×64 → 512×512)



Saharia, Chitwan, et al. "Image super-resolution via iterative refinement." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).

Cascaded Face Generation 1024 × 1024



Saharia, Chitwan, et al. "Image super-resolution via iterative refinement." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).