PTYCHNET : CNN BASED FOURIER PTYCHOGRAPHY

Armin Kappeler[1], *Sushobhan Ghosh*[2], *Jason Holloway*[3], *Oliver Cossairt*[2], *Aggelos Katsaggelos*[2]

[1]Yahoo Inc. [2]Northwestern University [3]Columbia University

ABSTRACT

Fourier ptychography is an imaging technique that overcomes the diffraction limit of conventional cameras with applications in microscopy and long range imaging. Diffraction blur causes resolution loss in both cases. In Fourier ptychography, a coherent light source illuminates an object, which is then imaged from multiple viewpoints. The reconstruction of the object from these set of recordings can be obtained by an iterative phase retrieval algorithm. However, the retrieval process is slow and does not work well under certain conditions. In this paper, we propose a new reconstruction algorithm that is based on convolutional neural networks and demonstrate its advantages in terms of speed and performance.

Index Terms**—** Fourier ptychography, Convolutional Neural Network, CNN

1. INTRODUCTION

Imaging using traditional optical systems is constrained by the space-bandwidth product (SBP) [1], which describes the trade-off between high resolution and large field of view. Fourier ptychography (FP) is a coherent imaging technique which aims to overcome the SBP limitation by capturing a sequence of SBP limited images and computationally combining them to recover a high resolution, large FOV image and thus overcoming the SBP barrier. Fourier ptychography has been applied to wide field, high resolution microscopy [2], quantitative phase imaging [3], adaptive Fourier ptychography imaging [4], long distance, sub-diffraction imaging [5] and other applications. In Fourier ptychography, a high resolution image is recovered from a set of frequency limited low resolution images of an object illuminated with a coherent light source. To achieve this, an iterative phase retrieval algorithm [6] recovers the phase information that is lost in the incoherent imaging process. A detailed overview of different phase reconstruction techniques is provided in [7, 8].

Iterative phase retrieval algorithms perform well if the set of low resolution images have overlapping frequency bands in the Fourier domain, but the reconstruction quality quickly degrades as the overlap between the Fourier patches decreases [9]. The requirement of overlap between neighboring patches requires sequential scanning to obtain all the low resolution images and provides a major barrier to single shot ptychography [10]. Reducing or eliminating the overlap-requirement would lead to a much faster acquisition time. In this paper, we focus on the algorithm for retrieving the high resolution image. In place of a phase retrieval algorithm, we propose a Convolutional Neural Network (CNN) based solution (PtychNet), that directly restores the image in the spatial domain without explicitly recovering the phase information. CNNs have been proven to be very effective for image classification [11–13], and have become increasingly popular with other image processing tasks such as super-resolution [14–16], image segmentation [17], etc.

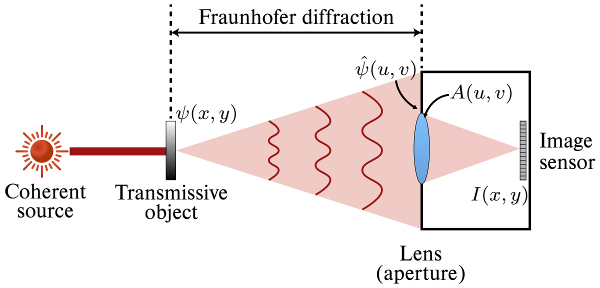


Fig. 1: Example setup for Fourier ptychography (FP). Coherent light diffracts through a translucent medium into the far-field. A lens samples a portion of the Fourier domain which is recorded as intensity images at the sensor. See Section 2.1 for details.

We show that PtychNet obtains better reconstruction results in considerably less time if the low resolution images have no overlapping frequency bands. When the low-resolution images contain overlapping support in the frequency domain, we can use PtychNet to significantly reduce the computation time of an iterative phase retrieval algorithm.

The remainder of the paper is organized as follows. In Section 2 we briefly introduce Fourier ptychography, in Section 3 we explain our proposed framework PtychNet. Section 4 contains our results and experimental evaluation and Section 5 concludes the paper.

2. FOURIER PTYCHOGRAPHY

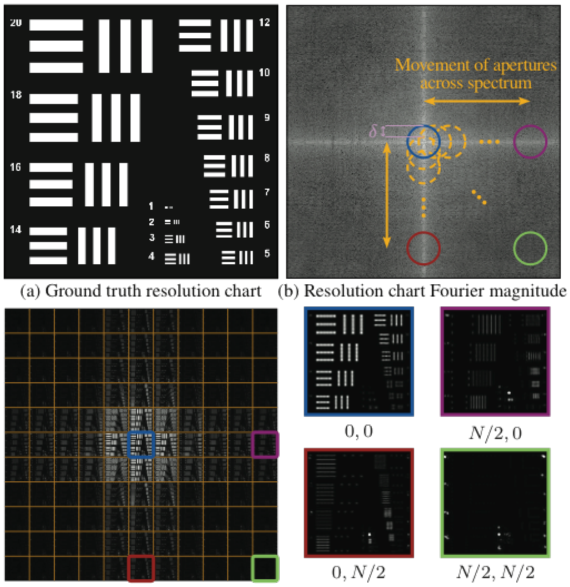
2.1. Image Formation Model

Consider the generalized imaging setup shown in Figure 1. A monochromatic source with wavelength λ illuminates a transparent object. Let the 2D complex field that emanates from the object be denoted as $\psi(x,y)$. If a camera is placed in the far-field and satisfies the Fraunhofer approximation, the field incident on the lens is a scaled Fourier transform of the scene,

$$\hat{\psi}(u,v) = \mathcal{F}_{\frac{1}{\lambda z}}\left\{\psi(x,y)\right\},$$

where λ is the wavelength of illumination, z is the distance between object and lens, $\hat{\psi}(u,v)$ is the field at the lens and (u,v)

978-1-5090-2175-8/17/$31.00 ©2017 IEEE



(a) Ground truth resolution chart (b) Resolution chart Fourier magnitude

$0, 0$ $N/2, 0$ $0, N/2$ $N/2, N/2$

(c) Observed image mosaic (d) Detail of observed images

Fig. 2: Example of image acquisition in Fourier ptychography. $N \times N$ images with limited, overlapping frequency bands are captured to recover one high resolution image. Image used from [5] with permission.

are coordinates in the frequency domain. The frequency spectrum is limited by the finite aperture of the lens, $A(u - c_u, v - c_v)$, where (c_u, c_v) is the center of the lens. The lens focuses the light on the image plane–which also satisfies the Fraunhofer approximation–and the intensity of the resulting field is recorded by the sensor. The measured intensity is thus given by

$$I(x, y, c_u, c_v) \propto \left| \mathcal{F}\left\{ \hat{\psi}(u, v) \odot A(u - c_u, v - c_v) \right\} \right|^2 \quad (1)$$

where $\odot$ signifies an element-wise multiplication. For simplicity, we will drop the scaling factor of the Fraunhofer approximation in this paper, though it may be accounted for after image reconstruction if desired.

To emulate capturing the scene with a larger lens, $N \times N$ images are captured by translating the lens, (c_u, c_v), to cover a larger portion of the Fourier spectrum. An example of the data acquisition process is shown in Figure 2.

2.2. Iterative Error Reduction Algorithm

Recovering the complex field $\hat{\psi}(u, v)$ from the set of measured intensity images I_i, $i = 1, \ldots, N$, is a non-convex optimization problem. That is, recovering $\hat{\psi}(u, v)$ reduces to solving the optimization problem:

$$\hat{\psi}^* = \text{argmin}_{\hat{\psi}} \sum_i \left\| \psi_i - \mathcal{F}\left\{ A_i \odot \hat{\psi} \right\} \right\|_2 \quad \text{s.t. } |\psi_i|^2 = I_i,$$

where the spatial arguments have been omitted for compactness. For an ideal lens with radius r, light within the support is passed uniformly and all other light is rejected, $A = ||(u - c_u, v - c_v)||_2 \leq r$.

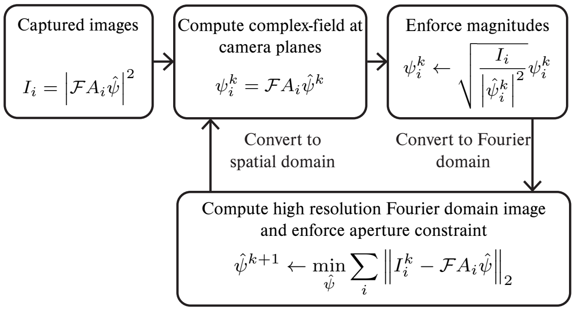

Captured images
$I_i = \left| \mathcal{F} A_i \hat{\psi} \right|^2$
Compute complex-field at camera planes
$\psi_i^k = \mathcal{F} A_i \hat{\psi}^k$
Enforce magnitudes
$\psi_i^k \leftarrow \sqrt{\frac{I_i}{\left|\hat{\psi}_i^k\right|^2}} \psi_i^k$
Convert to spatial domain
Convert to Fourier domain
Compute high resolution Fourier domain image and enforce aperture constraint
$\hat{\psi}^{k+1} \leftarrow \min_{\hat{\psi}} \sum_i \left\| I_i^k - \mathcal{F} A_i \hat{\psi} \right\|_2$

Fig. 3: Block diagram for IERA used in [5], modified with permission.

Conventional methods estimate $\hat{\psi}(u, v)$ using variations on iterative error reduction algorithms (IERAs) that enforce magnitude constraints in the spatial domain and support constraints in the Fourier domain [6, 7]. Figure 3 shows the block diagram of the IERA used in [5].

3. PTYCHNET

We propose a learning-based algorithm of recovering the high resolution image based on Convolutional Neural Networks. A high-level representation of the network structure is shown in Figure 4. Our network learns a non-linear mapping from the intensity images I_i to the original input light field ψ. Both, input I_i and output ψ are in the spatial domain. The inverse filters of the band-passes applied to the original light field can be approximated with convolutional filters, and the reconstruction process is locally independent which makes this a well-suited problem for a CNN. The input data of the CNN consists of the concatenation of all the intensity images I_i to a 3D-cube with dimensions $w \times h \times N^2$ where w and h are the width and height of the image and N^2 are the number of sampled images. The output of the CNN is the amplitude of the desired high resolution field ψ.

3.1. Architecture

The proposed CNN is based on the architecture used in [14]. It consists of three convolutional layers. The two hidden layers H_1 and H_2 are each followed by a ReLU activation function. The first layer consists of 64 kernels with a kernel size of 9×9. The second layer has 32 kernels with a size of 5×5 and the output layer has a kernel size of 5×5. The output layer has only one kernel that will directly produce the reconstructed image in the spatial domain. The weights are initialized with random Gaussian distributed values with a standard deviation of 0.001. We use the Euclidean distance as our loss function[1].

3.2. Training Procedure

Our algorithm is implemented with the Caffe framework [18] and trained using 91 publicly available images from Set91 [19].

[1] Experiments with TV-minimization as loss function did not lead to any improvements in PSNR

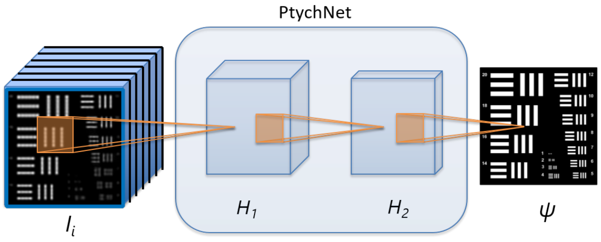


Fig. 4: PtychNet overview, three layer CNN with two hidden layers. ReLU activation functions follow the hidden layers.

The images are converted to grayscale and resized to $w \times h$ pixels, where $w = h = 512$ pixels. These images represent our ground truth data ψ. For each training image, equation (1) is used to generate $N^2, N = 7$ images with low spatial resolution I_i. The observed images are concatenated into a 3D-cube of size $w \times h \times N^2$. Using all 91 datacubes, approximately 15000 $48 \times 48 \times 49$ patches were extracted to train the CNN. Note that since both the input and the output image of the CNN are in the spatial domain, our reconstruction algorithm is spatially invariant and therefore we can divide the input and output data into patches for parallel processing. To avoid border effects introduced by zero-padding for the convolutional layers, only the 32×32 center pixels of a training patch are used to calculate the Euclidean loss. We created two separate training datasets of input images with overlapping and non-overlapping frequency bands. We achieve better performance for the non-overlapping dataset by subtracting the center input image (which contains the DC term) from the reconstructed output image. This approach is similar to the idea of residual networks [13]. Our networks were trained for 200,000 iterations with a batch size of 256.

4. EXPERIMENTAL RESULTS

In this section, we tested the effectiveness of our CNN by comparing it against the IERA algorithm proposed in [5]. We test our algorithm on using a resolution chart (resChart) and Lena image in addition to the Set5 images from [19]. We use peak-signal-to-noise-ratio (PSNR) and structural similarity (SSIM) as our performance metrics. The IERA algorithm was evaluated after 100 iterations which is well after the results have plateaued.

Two testing configurations are considered: the standard 61% overlap used in FP, and 0% overlap which corresponds to a densely packed lens array. Overlap defined as the percentage of the area shared between adjacent input images in the frequency domain (see Figure 2b). The baseline for image performance is the center image (image at position 0, 0 in Figure 2d), which corresponds to the lowpass filtered original image.

4.1. Without overlap

Table 1 shows the PSNR (dB) and SSIM results for the non-overlapping case. Results are shown for the center image, IERA and PtychNet. Figure 5 shows the original image, captured center image, and the reconstructed images using IERA and PtychNet for Lena and the resolution chart. In both cases PtychNet outperforms IERA and improves on the baseline; improvements over IERA are between 0.6 and 2.1 dB.

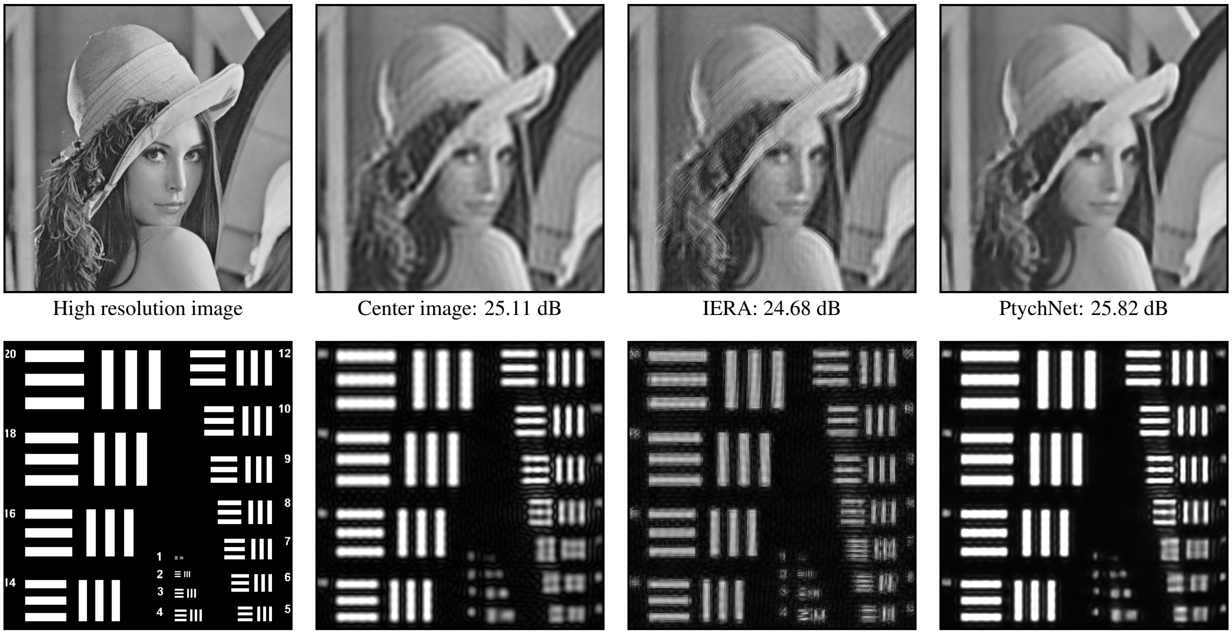
Table 1: PSNR and SSIM without overlap

Image	Metric	Center	IERA	PtychNet
lena	PSNR	25.11	24.68	**25.82**
	SSIM	0.6828	0.6488	**0.7146**
resChart	PSNR	14.41	12.92	**15.05**
	SSIM	0.1981	0.1357	**0.2536**
baby	PSNR	25.97	25.46	**26.50**
	SSIM	0.6836	0.6488	**0.7030**
bird	PSNR	28.49	28.50	**29.70**
	SSIM	0.8175	0.8068	**0.8537**
butterfly	PSNR	21.47	21.57	**23.24**
	SSIM	0.6201	0.5866	**0.7258**
head	PSNR	30.48	30.01	**30.67**
	SSIM	0.7295	0.6964	**0.7410**
woman	PSNR	24.93	24.82	**25.83**
	SSIM	0.7611	0.7431	**0.8106**

4.2. With overlap

The IERA and PtychNet images for 61% overlap are shown in Figure 6. Note that the input and center images are the same as Figure 5. When operating with sufficient overlap the IERA reconstructions are superior to PtychNet, which is particularly evident in the resolution chart. Interestingly, the gap in performance for the resolution chart (IERA: 18.28 dB, PtychNet: 18.04 dB) is much smaller than for Lena (IERA: 31.52 dB, PtychNet: 29.53 dB). In images from Set5, IERA outperforms PtychNet by an average of 2.4 dB (IERA: 35.02 dB, PtychNet: 32.61 dB).

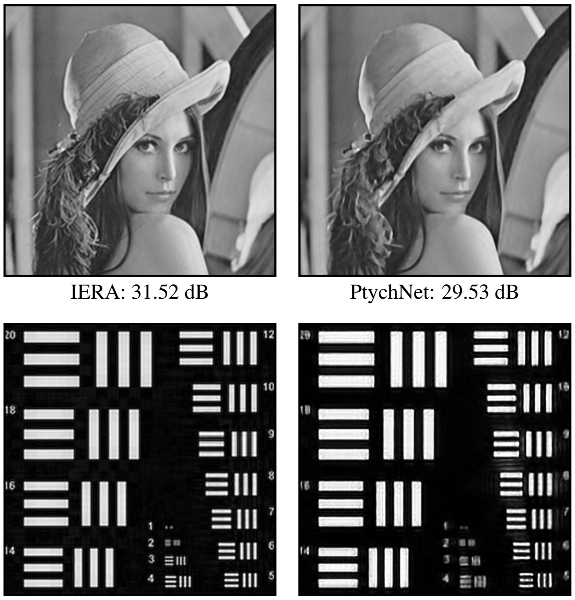
Despite the lower performance, PtychNet has significantly lower runtimes than IERA. For example, the runtime for a 512×512 pixel image for IERA with 100 iterations is about 1 minute, while PtychNet completes in 0.5 seconds. The IERA algorithm requires an initial guess, which is taken as the mean image (averaged over the 49 input images). Alternatively, the output PtychNet can be used to initialize IERA which leads to rapid convergence of the algorithm. The output of PtychNet is itself a good reconstruction of the original image so few additional iterations are needed. In Figure 7 we show the average PSNR versus iteration graph for Set5 and Lena and the resolution chart. We test three initialization schemes: the output of PtychNet, the center image, and the mean image. In our tests, IERA with mean initialization requires 30 iterations to converge, while using the output of PtychNet for initialization only requires 6 iterations to converge to the same PSNR. For all seven images, the final PSNRs are within 0.02 dB regardless of initialization but PtychNet improves convergence time by a factor of five.



High resolution image — Center image: 25.11 dB — IERA: 24.68 dB — PtychNet: 25.82 dB

High resolution image — Center image: 14.41 dB — IERA: 12.92 dB — PtychNet: 15.05 dB

Fig. 5: Results for Lena and resolution chart with 0% overlap



IERA: 31.52 dB — PtychNet: 29.53 dB

IERA: 18.28 dB — PtychNet: 18.04 dB

Fig. 6: Results for Lena and resolution chart with 61% overlap

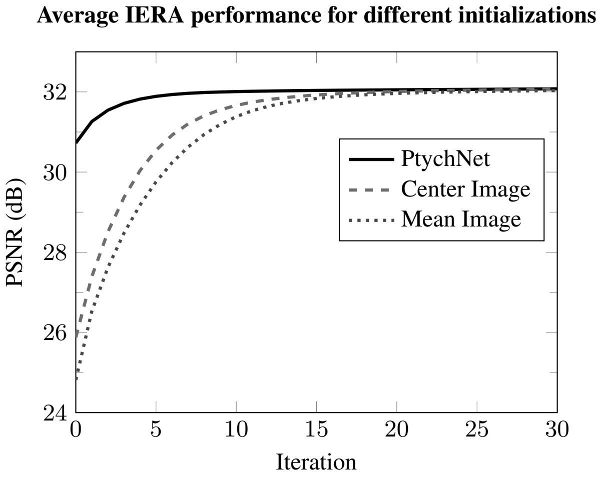


Fig. 7: IERA with different initializations. Using PtychNet for initialization yields faster convergence compared to mean and center image initializations.

5. CONCLUSION

We introduced a recovery algorithm for Fourier ptychography based on deep learning. To the best of our knowledge, there is no pre-existing work on CNN based Fourier ptychography algorithms. We show that for non-overlapped Fourier sampling, PtychNet performed significantly better than an existing FP algorithm, in both speed and quality. Furthermore, PtychNet can be used as the initialization in conventional oversampled FP to improve convergence times five-fold.

Acknowledgments: This work was supported in part by NSF CAREER grant IIS-1453192; ONR grant N00014-15-1-2735; DARPA REVEAL grant HR0011-16-C-0028.

6. REFERENCES

[1] Adolf W. Lohmann, Rainer G. Dorsch, David Mendlovic, Carlos Ferreira, and Zeev Zalevsky, "Space–bandwidth product of optical signals and systems," *J. Opt. Soc. Am. A*, vol. 13, no. 3, pp. 470–473, Mar 1996.

[2] Guoan Zheng, Roarke Horstmeyer, and Changhuei Yang, "Wide-field, high-resolution fourier ptychographic microscopy," *Nat Photon*, vol. 7, no. 9, pp. 739–745, Sep 2013, Article.

[3] Xiaoze Ou, Roarke Horstmeyer, Changhuei Yang, and Guoan Zheng, "Quantitative phase imaging via fourier ptychographic microscopy," *Opt. Lett.*, vol. 38, no. 22, pp. 4845–4848, Nov 2013.

[4] Zichao Bian, Siyuan Dong, and Guoan Zheng, "Adaptive system correction for robust fourier ptychographic imaging," *Opt. Express*, vol. 21, no. 26, pp. 32400–32410, Dec 2013.

[5] Jason Holloway, M. Salman Asif, Manoj Kumar Sharma, Nathan Matsuda, Roarke Horstmeyer, Oliver Cossairt, and Ashok Veeraraghavan, "Toward Long Distance, Sub-diffraction Imaging Using Coherent Camera Arrays," *Computational Imaging, IEEE Transactions on*, Submitted for review.

[6] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of the phase from image and diffraction plane pictures," *Optik (Jena)*, vol. 35, pp. 337, 1972.

[7] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, Aug 1982.

[8] C. Yang, J. Qian, A. Schirotzek, F. Maia, and S. Marchesini, "Iterative Algorithms for Ptychographic Phase Retrieval," *ArXiv e-prints*, May 2011.

[9] Jianliang Qian, Chao Yang, A Schirotzek, F Maia, and S Marchesini, "Efficient algorithms for ptychographic phase retrieval," *Inverse Problems and Applications, Contemp. Math*, vol. 615, pp. 261–280, 2014.

[10] Pavel Sidorenko and Oren Cohen, "Single-shot ptychography," *Optica*, vol. 3, no. 1, pp. 9–14, Jan 2016.

[11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., pp. 1097–1105. Curran Associates, Inc., 2012.

[12] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[14] Armin Kappeler, Seunghwan Yoo, Qiqin Dai, and Aggelos K Katsaggelos, "Video super-resolution with convolutional neural networks," *IEEE Transactions on Computational Imaging*, vol. 2, no. 2, pp. 109–122, 2016.

[15] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, Feb 2016.

[16] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al., "Photo-realistic single image super-resolution using a generative adversarial network," *arXiv preprint arXiv:1609.04802*, 2016.

[17] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[18] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.

[19] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, Nov 2010.