Deep Learning for Image/Video Restoration and Super-resolution

Other titles in Foundations and Trends® in Computer Graphics and Vision

Discrete Graphical Models - An Optimization Perspective
Bogdan Savchynskyy
ISBN: 978-1-68083-638-7

Line Drawings from 3D Models: A Tutorial
Pierre Bénard and Aaron Hertzmann
ISBN: 978-1-68083-590-8

Publishing and Consuming 3D Content on the Web: A Survey
Marco Potenziani, Marco Callieri, Matteo Dellepiane and Roberto Scopigno
ISBN: 978-1-68083-536-6

Crowdsourcing in Computer Vision
Adriana Kovashka, Olga Russakovsky, Li Fei-Fei and Kristen Grauman
ISBN: 978-1-68083-212-9

The Path to Path-Traced Movies
Per H. Christensen and Wojciech Jarosz
ISBN: 978-1-68083-210-5

(Hyper)-Graphs Inference through Convex Relaxations and Move Making Algorithms
Nikos Komodakis, M. Pawan Kumar and Nikos Paragios
ISBN: 978-1-68083-138-2

Deep Learning for Image/Video Restoration and Super-resolution

A. Murat Tekalp
Department of Electrical and Electronics Engineering
Koç University
Turkey
mtekalp@ku.edu.tr

the essence of knowledge
Boston — Delft

Foundations and Trends® in Computer Graphics and Vision

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

A. M. Tekalp. *Deep Learning for Image/Video Restoration and Super-resolution.* Foundations and Trends® in Computer Graphics and Vision, vol. 13, no. 1, pp. 1–110, 2022.

ISBN: 978-1-68083-973-9

Foundations and Trends® in Computer Graphics and Vision

Volume 13, Issue 1, 2022

Editorial Scope

Topics

Foundations and Trends® in Computer Graphics and Vision publishes survey and tutorial articles in the following topics:

- Rendering
- Shape
- Mesh simplification
- Animation
- Sensors and sensing
- Image restoration and enhancement
- Segmentation and grouping
- Feature detection and selection
- Color processing
- Texture analysis and synthesis
- Illumination and reflectance modeling
- Shape representation
- Tracking
- Calibration
- Structure from motion
- Motion estimation and registration
- Stereo matching and reconstruction
- 3D reconstruction and image-based modeling
- Learning and statistical methods
- Appearance-based matching
- Object and scene recognition
- Face detection and recognition
- Activity and gesture recognition
- Image and video retrieval
- Video analysis and event recognition
- Medical image analysis
- Robot localization and navigation

Information for Librarians

Foundations and Trends® in Computer Graphics and Vision, 2022, Volume 13, 4 issues. ISSN paper version 1572-2740. ISSN online version 1572-2759. Also available as a combined paper and online subscription.

Contents

Deep Learning for Image/Video Restoration and Super-resolution

A. Murat Tekalp

Department of Electrical and Electronics Engineering, Koç University, Turkey; mtekalp@ku.edu.tr

ABSTRACT

Recent advances in neural signal processing led to significant improvements in the performance of learned image/video restoration and super-resolution (SR). An important benefit of data-driven deep learning approaches to image processing is that neural models can be optimized for any differentiable loss function, including perceptual loss functions, leading to perceptual image/video restoration and SR, which cannot be easily handled by traditional model-based methods.

We start with a brief problem statement and a short discussion on traditional vs. data-driven solutions. We next review recent advances in neural architectures, such as residual blocks, dense connections, residual-in-residual dense blocks, residual blocks with generative neurons, self-attention and visual transformers. We then discuss loss functions and evaluation (assessment) criteria for image/video restoration and SR, including fidelity (distortion) and perceptual criteria, and the relation between them, where we briefly review the perception vs. distortion trade-off.

We can consider learned image/video restoration and SR as learning either a nonlinear regressive mapping from degraded to ideal images based on the universal approximation theorem, or a generative model that captures the probability distribution of ideal images. We first review regressive

A. Murat Tekalp (2022), "Deep Learning for Image/Video Restoration and Super-resolution", Foundations and Trends® in Computer Graphics and Vision: Vol. 13, No. 1, pp 1–110. DOI: 10.1561/0600000100.

inference via residual and/or dense convolutional networks (ConvNet). We also show that using a new architecture with residual blocks based on a generative neuron model can outperform classical residual ConvNets in peak-signal-to-noise ratio (PSNR). We next discuss generative inference based on adversarial training, such as SRGAN and ESRGAN, which can reproduce realistic textures, or based on normalizing flow such as SRFlow by optimizing log-likelihood. We then discuss problems in applying supervised training to real-life restoration and SR, including overfitting image priors and overfitting the degradation model seen in the training set. We introduce multiple-model SR and real-world SR (from unpaired training data) formulations to overcome these problems. Integration of traditional model-based methods and deep learning for non-blind restoration/SR is introduced as another solution to model overfitting in supervised learning.

In learned video restoration and SR (VSR), we first discuss how to best exploit temporal correlations in video, including sliding temporal window vs. recurrent architectures for propagation, and aligning frames in the pixel domain using optical flow vs. in the feature space using deformable convolutions. We next introduce early fusion with feature-space alignment, employed by the EDVR network, which obtains excellent PSNR performance. However, it is well-known that videos with the highest PSNR may not be the most appealing to humans, since minimizing the mean-square error may result in blurring of details. We then address perceptual optimization of VSR models to obtain natural texture and motion. Although perception-distortion tradeoff has been well studied for images, few works address perceptual VSR. In addition to using perceptual losses, such as MS-SSIM, LPIPS, and/or adversarial training, we also discuss explicit loss functions/criteria to enforce and evaluate temporal consistency. We conclude with a discussion of open problems.

1

Introduction

Deep learning has made a significant impact not only on computer vision and natural language processing but also on classical signal processing problems such as image/video restoration/super-resolution (SR) and compression. This monograph reviews recent advances and the state of the art in image/video restoration and SR using deep learning. It is worth noting that the nonlinear neural signal processing techniques discussed in this monograph also apply to other inverse problems in imaging.

This section provides an introduction to image restoration and SR problems, including a general overview of classical model-based vs. modern data-driven solutions. We start with the problem statement in Section 1.1, where we pose image restoration/SR as an ill-posed inverse problem. Linear model-based regularization of ill-posed inverse problems is reviewed in Section 1.2. Limitations of linear, shift-invariant (LSI) regularization are discussed in Section 1.3. Next, Section 1.4 provides an overview of classical nonlinear model-based regularized inversion vs. modern data-driven learned approaches. We introduce the three pillars of learned image/video restoration and SR solutions: the architecture, the optimization and evaluation criteria, and training in Section 1.5. Finally, we briefly discuss other related survey articles in Section 1.6.

1.1 Problem Statement

Inverse problems in imaging are those problems, where we want to solve for the ideal image vector $\mathbf{x}$ given a nonlinear observation model

$$\mathbf{y} = \mathcal{D}(\mathbf{x}) + \mathbf{v} \tag{1.1}$$

where $\mathbf{y}$ denotes the observation vector, $\mathcal{D}$ is a nonlinear degradation operator, and $\mathbf{v}$ is the observation noise vector. In the traditional formulation of inverse problems, the degradation (forward) model is assumed to be linear, which can be expressed as

$$\mathbf{y} = \mathbf{DHx} + \mathbf{v} \tag{1.2}$$

where $\mathbf{H}$ denotes a linear degradation operator, and $\mathbf{D}$ is an observation matrix. This linear observation model includes the following image restoration problems as special cases:

- The denoising problem, where $\mathbf{D}=\mathbf{H}=\mathbf{I}$ (identity matrix).
- The deblurring problem, where $\mathbf{D}=\mathbf{I}$ and the matrix $\mathbf{H}$ is determined by the blur point spread function (PSF).
- The super-resolution (SR) problem, where $\mathbf{D}$ and $\mathbf{H}$ represent the sub-sampling operation and the anti-alias filter, respectively.
- The image inpainting problem, where the elements of matrix $\mathbf{D}$ that correspond to missing pixels are set to zero.

1.1.1 Ill-Posed Problems

According to Hadamard, a problem is well-posed if it satisfies the following conditions (Tikhonov and Arsenin, 1977): i) a solution exists, ii) the solution is unique, and iii) small perturbations (noise) in the observations (input) results in small changes in the solution. Problems that are not well-posed in the sense of Hadamard are called ill-posed.

Inverse problems in imaging are often ill-posed because the matrices $\mathbf{D}$ and/or $\mathbf{H}$ may be non-square with more unknowns than the number of equations; hence, the solution either does not exist and/or is not unique, and/or the condition number of matrix $\mathbf{H}$ is large so that the solution is highly sensitive to observation noise.

1.1.2 Non-blind vs. Blind Image Restoration and SR

We can classify inverse problems as non-blind or blind depending on whether the degradation operator and observation noise level in Eqn. 1.1 and Eqn. 1.2 are known or not.

A low resolution (LR) image is modeled as down-sampled version of an ideal high resolution (HR) image. We typically model the anti-alias filtering in the down-sampling operation by a bicubic filter; hence, this process is often referred to as bicubic downsampling. In real-world applications, there are additional sources of blur in LR image formation, such as motion blur or camera shake blur, which is represented by a convolution kernel $\mathbf{k}$, given by

$$\mathbf{y} = (\mathbf{k} * \mathbf{x}) \downarrow + \mathbf{v} \tag{1.3}$$

where $\downarrow$ denotes bicubic downsampling. While the blur due to $\downarrow$ is a bicubic filter, the additional source of blur, denoted by $\mathbf{k}$ is usually unknown and image specific.

Non-blind image restoration and SR refers to the case where the blur kernel $\mathbf{k}$ and noise level in Eqn. 1.3 are known or estimated prior to the image restoration process. Most non-blind methods assume that there is no additional source of blur in LR image formation, and only model bicubic anti-alias filtering. Hence, Eqn. 1.3 simplifies as

$$\mathbf{y} = (\mathbf{x}) \downarrow + \mathbf{v} \tag{1.4}$$

Blind image restoration and SR refers to the case where the blur kernel $\mathbf{k}$ and noise level in Eqn. 1.3 are unknown and must be estimated simultaneously with the image restoration and SR process.

1.2 Model-based Regularization of Ill-Posed Inverse Problems

Since the forward model (1.1) or (1.2) is in general not invertible, one can possibly define the ordinary least squares estimate of $\mathbf{x}$ or the pseudo-inverse solution given by

$$\hat{\mathbf{x}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y} \tag{1.5}$$

However, this solution is not regularized in the sense that it is highly sensitive to small perturbations (noise) in the observation vector $\mathbf{y}$.

Finding a solution that is well-behaved in the presence of observation noise is impossible without utilizing some prior information about the ideal signal/image $\mathbf{x}$. This is called regularization of the inverse solution. Traditional model-based regularized inversion methods minimize a cost function subject to some constraints (prior) on the solution. Assuming the observation noise is additive, white Gaussian, and is independent of the signal/image $\mathbf{x}$, the regularized inverse solution can be found as:

$$\hat{\mathbf{x}}(\lambda) = arg_{\mathbf{x}}\ min\ \frac{1}{2}||\mathbf{y} - \mathbf{DHx}||^2 + \lambda\mathbf{R}(\mathbf{x}) \tag{1.6}$$

where $\mathbf{R}(\mathbf{x})$ is a regularization operator that imposes some prior on $\mathbf{x}$. Hence, the solution is the minimizer of a data-consistency cost term, which measures how well the restored image matches the observations given the degradation model, and a regularizer term, which imposes some prior knowledge or promotes images with some desirable property.

One of the first regularization methods is Tikhonov regularization, which, in the case $\mathbf{D{=}I}$, is given by (Tikhonov and Arsenin, 1977):

$$\hat{\mathbf{x}}(\lambda) = (\mathbf{H}^T\mathbf{H} + \lambda\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T\mathbf{y} \tag{1.7}$$

where $\mathbf{L}$ is a linear regularization operator expressed in matrix form and λ is a parameter that controls the tradeoff between data consistency and regularization, i.e., noise sensitivity. For example, $\mathbf{L}$ can be the Laplacian operator that estimates high frequency image components. In this case, minimizing the energy of high frequency image components can be viewed as imposing a smoothness constraint as an image prior.

Direct computation of (1.7) requires inversion of the large matrix $(\mathbf{H}^T\mathbf{H} + \lambda\mathbf{L}^T\mathbf{L})$. There are two common approaches to avoid inversion of this large matrix: i) employing an iterative solution, ii) diagonalization using the discrete Fourier transform assuming the matrix is circulant. Under certain assumptions, this regularized inverse solution can be obtained by a linear, shift-invariant regularized inverse filter.

1.3 Limitations of Linear Shift-invariant Regularized Inverse Filters

Let's express the observation model (1.2), in the case $\mathbf{D}$=$\mathbf{I}$, in scalar form as a convolution

$$y(n_1, n_2) = h(n_1, n_2) * *x(n_1, n_2) + v(n_1, n_2) \tag{1.8}$$

Taking the 2-D discrete Fourier transform of both sides, we obtain

$$Y(e^{j\omega_1}, e^{j\omega_2}) = H(e^{j\omega_1}, e^{j\omega_2})X(e^{j\omega_1}, e^{j\omega_2}) + V(e^{j\omega_1}, e^{j\omega_2}) \tag{1.9}$$

If we process the observed image by a linear, shift-invariant restoration filter $\Phi(e^{j\omega_1}, e^{j\omega_2})$, the estimated image can be expressed as

$$\hat{X}(e^{j\omega_1}, e^{j\omega_2}) = \Phi(e^{j\omega_1}, e^{j\omega_2})Y(e^{j\omega_1}, e^{j\omega_2}) \tag{1.10}$$

If we now substitute Eqn. 1.9 for $Y(e^{j\omega_1}, e^{j\omega_2})$, we get

$$\hat{X}(e^{j\omega_1}, e^{j\omega_2}) = \Phi(e^{j\omega_1}, e^{j\omega_2})[H(e^{j\omega_1}, e^{j\omega_2})X(e^{j\omega_1}, e^{j\omega_2}) + V(e^{j\omega_1}, e^{j\omega_2})] \tag{1.11}$$

In order to analyze the artifacts due to processing with a linear, shift-invariant filter $\Phi(e^{j\omega_1}, e^{j\omega_2})$, we add and subtract $X(e^{j\omega_1}, e^{j\omega_2})$ to the right hand side to obtain (Tekalp and Sezan, 1990):

$$\begin{aligned} \hat{X}(e^{j\omega_1}, e^{j\omega_2}) &= X(e^{j\omega_1}, e^{j\omega_2}) \\ &\quad +[\Phi(e^{j\omega_1}, e^{j\omega_2})H(e^{j\omega_1}, e^{j\omega_2}) - 1]X(e^{j\omega_1}, e^{j\omega_2}) \\ &\quad +\Phi(e^{j\omega_1}, e^{j\omega_2})V(e^{j\omega_1}, e^{j\omega_2}) \end{aligned} \tag{1.12}$$

The second term at the right-hand side is signal-dependent regularization error (ringing artifacts). The third term is filtered noise artifacts. If we let $\Phi(e^{j\omega_1}, e^{j\omega_2}) = H^{-1}(e^{j\omega_1}, e^{j\omega_2})$ (inverse filter) then the second term disappears, but the third term dominates and masks the signal $x(n_1, n_2)$. Hence, the trade-off between the last two terms is a theoretical limitation of LSI regularized solutions (Tekalp and Sezan, 1990).

In order to overcome this theoretical limitation of LSI inverse filters, many adaptive or nonlinear methods have been proposed within the past 30 years. They are briefly discussed in the next section.

1.4 Nonlinear Model-based vs. Data-driven Approaches

Traditional nonlinear model-based regularized inversion methods have been applied to solve image/video restoration and SR problems for over 50 years. We can broadly classify available solutions as: i) iterative methods that impose deterministic constraints or priors about the ideal image, ii) methods based on statistical estimation theory, and iii) example-based methods based on machine learning (but not end-to-end deep learning). Examples of such methods include maximum a posteriori probability (MAP) estimation, sparse modeling (Papyan *et al.*, 2018), adaptive filters (Erdogmus and Principe, 2006), and example-based machine learning (Freeman *et al.*, 2002; Liu *et al.*, 2007).

Iterative methods can be used to impose constraints on the solution. Early iterative regularization methods include nonlinear Landweber iterations, iterative back-projection, or projection onto convex sets (POCS) methods. Iterative solutions to variational optimization formulations, such as the total variation (TV) regularization, have also been proposed. TV regularization suppresses oscillations (noise) in the solution while allowing for discontinuities (edges). Later, iterative solutions based on sparse and redundant image representation have become popular. Sparse redundant representations constrain the signal to the form

$$\mathbf{x} = \mathbf{A}\gamma \tag{1.13}$$

where $\mathbf{x} \in R^n$, $\gamma \in R^m$ such that $m > n$, and the $n \times m$ matrix $\mathbf{A}$ is a dictionary of atoms. The vector γ is sparse with only few (say k) nonzero elements; thus, $\mathbf{x}$ is constrained to be a linear combination of k atoms from a learned dictionary $\mathbf{A}$.

Statistical estimation methods pose image/video restoration and SR as finding the minimum mean square error (MMSE) estimate, given by

$$\hat{\mathbf{x}}_{MMSE} = arg_{\hat{\mathbf{x}}}\ min\ E\{(\mathbf{x} - \hat{\mathbf{x}})^2\} \tag{1.14}$$

or the maximum a posteriori probability (MAP) estimate, given by

$$\hat{\mathbf{x}}_{MAP} = arg_{\mathbf{x}}\ min\ \ln p(\mathbf{x}|\mathbf{y}) = arg_{\mathbf{x}}\ min\,(\ln p(\mathbf{y}|\mathbf{x}) + \ln p(\mathbf{x})) \tag{1.15}$$

Note that when the distributions are Gaussian, the first and second terms in Eqn. (1.15) correspond to those in Eqn. (1.6).

Example-based learning has also been shown to yield good results. Nevertheless, classical model-based solutions require iterations (more computation) during inference and their performance is limited since single-image SR is a severely ill-posed inverse problem.

The latest advance in the state-of-the-art in nonlinear image/video restoration and SR is based on deep learning driven by big data. It only became possible to obtain deep learned SR results that are superior to those of traditional model-based approaches within the last 5-6 years leveraging the recent advances in deep neural network architectures and training methods including optimizers, wide availability of large datasets, and powerful GPU computing.

Learned image restoration and SR tasks can be posed as a nonlinear regression problem or a generative modeling problem. We can gain insight on how deep learning helps to achieve state of the art image restoration and SR results leveraging data-driven regression paradigm by means of the following example. Suppose we want to predict the weight of a person given his/her height and age. Given a dataset with weight, height and ages of people, we can fit a surface in 3-D to given data. If we fit a linear model, this would be a plane in 3-D as depicted in Figure 1.1. A nonlinear regression framework would allow fitting an arbitrary 3-D surface to the given data. Given the height and age of a new person not in the training dataset, we can project the height and age to the 3-D prediction surface to get a reading of the predicted weight. The shape of

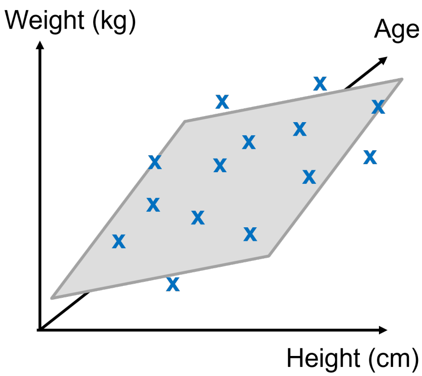

Figure 1.1: Illustration of linear regression in a 3-dimensional space.

the 3-D surface, which determines the accuracy of the predicted weight, depends on the form of the nonlinear predictor, the loss function used in fitting, and of course the goodness of the available training data.

Regressive inference for learned image restoration and SR works similarly, where we have input (LR) and output (HR) image pairs. Each corresponding LR-HR image pair is represented by a point in a very high dimensional space (each pixel is a dimension). For example, if we have 100×100 patches, that would constitute a 10,000 dimensional space. A deep learning model defines a prediction manifold that is fitted to these sample points in the very high-dimensional space. In analogy with the above example, the accuracy of the predicted HR images depends on the architecture of the neural network (the form of the predictor), the optimization criterion, and the available datasets.

Alternatively, generative inference works by first learning a model to represent the distribution of the ideal image conditioned on a given degraded image, and then sampling one or more plausible solutions from this distribution during inference.

The inference process in model-based methods and learned methods are in stark contrast. In traditional model-based methods, there is no training process, but we need to solve a different optimization problem for each test image. While this requires significantly more computation during inference, it provides flexibility to use a different degradation model for each test image. In learned methods, we typically assume all training and test images are subject to the same degradation process, and the training step requires significant computation, but the inference process is very fast. Hence, classical model-based and deep learning approaches have different strengths and weaknesses.

1.5 Three Pillars of Learned Image Restoration and SR

The three pillars of learned image restoration and SR are the network architecture, the optimization criteria, and training methodology and data. We provide a brief introduction to each of these pillars, depicted in Figure 1.2, in the following subsections.

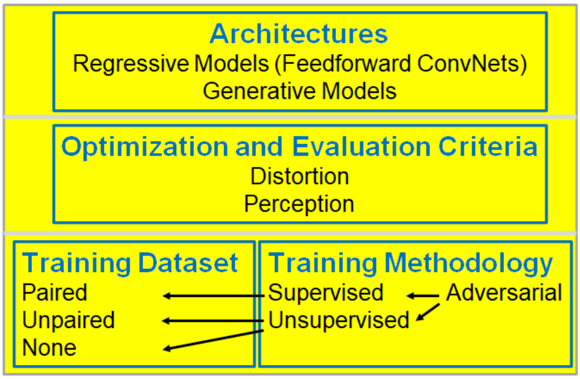

Figure 1.2: Three pillars of learned image restoration and SR.

1.5.1 Network Architectures: Regressive vs. Generative Models

In a very broad way, we can classify deep SR network architectures as regressive models and generative models. Regressive models are feedforward networks that learn a nonlinear mapping from the space of LR images to the space of HR images. They include residual networks, dense networks, and their variations. On the other hand, generative models learn the probability distribution of HR images conditioned on LR images. Thus, generative SR models enable sampling one or more HR images from the estimated conditional distribution of HR images. We provide an overview of recent advances in deep neural network architectures that contribute to achieving the state-of-the-art results in image/video restoration and SR in Section 2.

1.5.2 Optimization Criteria: Distortion vs. Perception

Unlike classical model-based methods, that optimize either l_2 or l_1 distortion subject to some regularization prior, learned image restoration and SR allows optimization with respect to any differentiable loss function. Parameters of the network can be optimized purely for distortion (fidelity) or a combination of fidelity and perceptual criteria. Blau and Michaeli (2018) show that distortion and perceptual quality are at odds with each other leading to perception-distortion tradeoff. Specifically, they study the optimal probability for correctly discriminating the outputs of an image restoration algorithm from real images and show that as the mean distortion decreases, this probability increases indicating worse perceptual quality. Achieving the best trade-off between highest fidelity and perceptual quality is an interesting research problem. Fidelity and perceptual optimization criteria and perception-distortion tradeoff are reviewed in more detail in Section 3.

1.5.3 Training Methods and Data: Supervised vs. Unsupervised

A vast majority of published literature on learned image restoration and SR perform supervised training from a synthetically generated LR, HR paired image dataset. This dataset depends on a particular blur kernel and noise level that is used to generate LR images from corresponding HR images. SR models obtained this way perform incredibly well, outperforming conventional model-based methods by a large margin, when the test set of images are also generated using the same degradation process. However, if the degradation in the test set of images differ from those in the training set, then SR performance deteriorates. We call this dependence of SR performance on the degradation model used in the training set as model overfitting.

When it comes to real-world problems, this approach of training SR models based on synthetically generated LR-HR image pairs is of limited use due to model overfitting, because real LR images are degraded by blur and noise, which are unknown in the practical setting. Furthermore, in the real-world SR setting, there is no ground-truth; hence there is no paired data available for training. Hence, in the real-world setting we have blind image restoration/SR problem without ground-truth data.

Recently, more researchers have started working on blind image restoration/SR methods that require no training, or can be trained without an external training set, or can be trained by unpaired datasets. These methods can be classified as: i) two-step approaches, where the blur kernel is estimated first and then used in a non-blind SR model, or ii) methods that iteratively correct the blur kernel estimate based on the LR image and the most recent estimate of the SR image. Both supervised and unsupervised training of image and video SR models are discussed in Section 4 and Section 5, respectively.

1.6 Related Recent Survey Articles

Other survey articles have appeared in the literature while we are working on this manuscript. Some of them introduce a taxonomy for deep learned SR models grouping them into categories, some benchmark SR algorithms, and some are in preprint.

Wang *et al.* (2021) provide a nice overview of the SISR literature; however, their paper does not cover transformer-based architectures, and touches upon video SR and real-world SR issues very briefly.

In deep journey into SR (Anwar *et al.*, 2021), the authors introduce a new taxonomy of the SR algorithms based on their architectures. They also provide a systematic evaluation of more than 30 SISR algorithms on six publicly available datasets given LR-HR image pairs. However, the assessment of results was only performed in terms of PSNR and SSIM; they do not discuss perception-distortion tradeoff, and they do not address real-world SR or video SR.

Liu *et al.* (2020) propose a taxonomy and classify video SR methods into six sub-categories according to the ways they utilize inter-frame information in a preprint article. They also compare more than 30 video SR algorithms. Blind image SR (Liu *et al.*, 2021a) is another preprint article that surveys image SR methods that can deal with an unknown degradation. The authors propose a taxonomy to categorize existing methods into three different classes according to the ways they model the degradation process.

Unlike these surveys, we do not benchmark a set of algorithms or propose a new taxonomy, but we focus on the understanding of foundational ideas and provide a comprehensive overview of basic principles of regressive (predictive) and generative SR network architectures, approaches to enforce temporal consistency in video SR, full-reference and no-reference image/video quality assessment (QA) measures, and differentiable QA measures that can be used as optimization loss functions. We also discuss the real-world SR problem and survey how to deal with the cases of known degradation model and blind SR as well as unsupervised learning approaches for real-world SR in detail. We believe this monograph can be used as reference material in an advanced image processing class.

References

Agustsson, E. and R. Timofte. (2017). "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study". In: *IEEE Conf. on Comp. Vision and Patt. Recog. (CVPR) Workshops*.

Anwar, S., S. Khan, and N. Barnes. (2021). "A deep journey into super-resolution: A survey". *ACM Computing Surveys (CSUR)*. 53(3): 1–34.

Bahat, Y. and T. Michaeli. (2020). "Explorable Super Resolution". In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR)*. 2716–2725.

Baker, S. and T. Kanade. (2000). "Hallucinating faces". In: *Proc. of the IEEE Int. Conf. on Automatic Face and Gesture Recognition*.

Bell-Kligler, S., A. Shocher, and M. Irani. (2019). "Blind super-resolution kernel estimation using an internal-GAN". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 284–293.

Bello, I., B. Zoph, A. Vaswani, J. Shlens, and Q. V. Le. (2019). "Attention augmented convolutional networks". In: *Int. Conf. on Computer Vision (ICCV)*.

Bianco, S., L. Celona, P. Napoletano, and R. Schettini. (2018). "On the use of deep learning for blind image quality assessment". *Signal, Image, and Video Processing (SIViP)*. 12: 355–362.

Blau, Y., R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor. (2018). "The 2018 PIRM Challenge on Perceptual Image Super-resolution". In: *Euro. Conf. Comp. Vision (ECCV) Workshop*.

Blau, Y. and T. Michaeli. (2018). "The perception-distortion tradeoff". In: *IEEE Conf. on Comp. Vision and Patt. Recog. (CVPR)*.

Blog. (2016). "Toward A Practical Perceptual Video Quality Metric". *Netflix Technology (2016-06-06), Last Accessed: 2021-07-01*.

Brabandere, B. D., X. Jia, T. Tuytelaars, and L. V. Gool. (2016). "Dynamic filter networks". In: *Advances in Neural Information Processing Systems (NeurIPS)*.

Buades, A., B. Coll, and J.-M. Morel. (2005). "A non-local algorithm for image denoising". In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR)*.

Caballero, J., C. Ledig, A. Aitken, A. Acosta, J. Totz, Z. Wang, and W. Shi. (2017). "Realtime video super-resolution with spatio-temporal networks and motion compensation". In: *IEEE/CVF Conf. on Comp. Vision and Pattern Recog. (CVPR)*.

Cai, J., S. Gu, and R. o. Timofte. (2019a). "NTIRE 2019 Challenge on Real Image Super-Resolution: Methods and Results". In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR) Workshop*.

Cai, J., H. Zeng, H. Yong, Z. Cao, and L. Zhang. (2019b). "Toward real-world single image super-resolution: A new benchmark and a new model". In: *IEEE Int. Conf. on Computer Vision*.

Chadha, A., J. Britto, and M. M. Roja. (2020). "iSeeBetter: Spatio-temporal video super-resolution using recurrent generative back-projection networks". *Springer Jour. of Computational Visual Media, Tsinghua University Press*. 6(3): 307–317.

Chan, K., X. Wang, K. Yu, C. Dong, and C. Loy. (2021a). "BasicVSR: The search for essential components in video super-resolution and beyond". In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recognition (CVPR)*.

Chan, K. C., S. Zhou, X. Xu, and C. C. Loy. (2021b). "BasicVSR++: Improving Video Super-Resolution with Enhanced Propagation and Alignment".

Chandler, D. M. and S. Hemami. (2007). “VSNR: A wavelet-based visual signal-to-noise ratio for natural images”. *IEEE Trans. on Image Processing*. 16(9): 2284–2298.

Chellapilla, K., S. Puri, and P. Simard. (2006). “High performance convolutional neural networks for document processing”. In: *Int. Workshop on Frontiers in Handwriting Recognition*.

Chen, H. *et al.* (2021). “Pre-trained Image Processing Transformer”. In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR)*.

Chen, J., X. Tan, C. Shan, S. Liu, and Z. Chen. (2020). “VESR-Net: The Winning Solution to Youku Video Enhancement and Super-Resolution Challenge”. *arXiv preprint arXiv:2003.02115*.

Cheng, M., N. Lin, K. Hwang, and J. Jeng. (2012). “Fast video super-resolution using artificial neural networks”. In: *Int. Symp. Commun. Syst. Netw. Digital Signal Proc. (CSNDSP)*. 1–4.

Cheon, M., S.-J. Yoon, B. Kang, and J. Lee. (2021). “Perceptual image quality assessment with transformers”. In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR) Workshops*.

Cho, K., B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. (2014). “Learning phrase representations using RNN encoder–decoder for statistical machine translation”. In: *Empirical Methods in Natural Language Processing (EMNLP)*. 1724–1734.

Chu, M., Y. Xie, J. Mayer, L. Leal-Taixe, and N. Thuerey. (2020). “Learning temporal coherence via self supervision for GAN-based video generation”. *ACM Trans. on Graphics (TOG)*. 39(4): 75–1.

Ciresan, D. C., U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber. (2011). “Flexible, High Performance Convolutional Neural Networks for Image Classification”. In: *Int. Joint Conf. on Artificial Intelligence (IJCAI), Barcelona, Spain*.

Dai, J., H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei. (2017). “Deformable Convolutional Networks”. In: *IEEE Int. Conf. on Computer Vision (ICCV)*. 764–773. DOI: 10.1109/ICCV.2017.89.

Daly, S. (1993). *The visible difference predictor: An algorithm for the assessment of image fidelity*. Ed. by E. A.B. Watson. *Digital Images and Human Vision*. Cambridge, MA: MIT Press. 179–206.

Dendi, S. V. R. and S. S. Channappayya. (2020). "No-reference video quality assessment using natural spatiotemporal scene statistics". *IEEE Trans. Image Proc.* 29: 5612–5624.

Ding, K., K. Ma, S. Wang, and E. P. Simoncelli. (2020). "Image quality assessment: Unifying structure and texture similarity". *IEEE Trans. on Patt. Anal. Mach. Intel. (PAMI).*

Ding, K., K. Ma, S. Wang, and E. P. Simoncelli. (2021). "Comparison of Image Quality Models for Optimization of Image Processing Systems". *Int. J. Comput. Vis. (IJCV).* 129(4): 1258–1281.

Dinh, L., J. Sohl-Dickstein, and S. Bengio. (2017). "Density estimation using real NVP". In: *Int. Conf. on Learning Representations (ICLR).*

Dong, C., C. C. Loy, K. He, and X. Tang. (2014). "Learning a Deep Convolutional Network for Image Super-Resolution". In: *European Conf. on Computer Vision (ECCV).*

Dong, C., C. C. Loy, K. He, and X. Tang. (2016). "Image Super-Resolution Using Deep Convolutional Networks". *IEEE Trans. Pattern Anal. Mach. Intell.* 38(2): 295–307.

Dosovitskiy, A., L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. (2021). "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". In: *Int. Conf. Learning Representations (ICLR).*

Erdogmus, D. and J. C. Principe. (2006). "From linear adaptive filtering to nonlinear information processing". *IEEE Signal Processing Magazine*: 14–33.

Fan, Y., J. Yu, D. Liu, and T. S. Huang. (2019). "An empirical investigation of efficient spatio-temporal modeling in video restoration". In: *IEEE/CVF Int. Conf. on Comp. Vision Workshop (ICCVW).*

Feng, R., J. Gu, Y. Qiao, and C. Dong. (2019). "Suppressing Model Overfitting for Image Super-Resolution Networks". In: *IEEE/CVF Conf. on Comp. Vision and Pattern Recog. (CVPR) Workshops.*

Ferzli, R. and L. J. Karam. (2009). "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)". *IEEE Trans. on Image Processing.* 18(4): 717–728.

Freeman, W., T. Jones, and E. Pasztor. (2002). "Example-based super-resolution". *IEEE Computer Graphics and Applications*. 22(2): 56–65.

Fukushima, K. (1980). "Neocognitron : A self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position". *Biological Cybernetics*. 36(4): 193–202.

Fuoli, D. *et al.* (2020). "AIM 2020 Challenge on Video Extreme Super-Resolution: Methods and Results". In: *IEEE/CVF Int. Conf. on Comp. Vision Workshop (ICCVW)*.

Fuoli, D., S. Gu, and R. Timofte. (2019). "Efficient video super-resolution through recurrent latent space propagation". In: *Int. Conf. Comp. Vision (ICCV) Workshops*.

Gao, F., Y. Wang, P. Li, M. Tan, J. Yu, and Y. Zhu. (2017). "Deepsim: Deep similarity for image quality assessment". *Neurocomputing*. 257: 104–114.

Goodfellow, I., Y. Bengio, and A. Courville. (2016). *Deep Learning*. MIT Press, Cambridge.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. (2014). "Generative adversarial nets". In: *Advances in Neural Information Processing Systems (NIPS)*. Vol. 2. 2672–2680.

Gu, J., H. Cai, C. Dong, *et al.* (2021). "NTIRE 2021 Challenge on perceptual image quality assessment". In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR) Workshops*.

Gu, J., H. Lu, W. Zuo, and C. Dong. (2019). "Blind super-resolution with iterative kernel correction". In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR)*.

Hara, K., H. Kataoka, and Y. Satoh. (2018). "Can spatiotemporal 3D CNNs retrace the history of 2D CNNs and ImageNet?" In: *IEEE/CVF Conf. on Comp. Vis. Patt. Recog. (CVPR)*.

Haris, M., G. Shakhnarovich, and N. Ukita. (2019). "Recurrent back-projection network for video super-resolution". In: *IEEE Conf. on Comp. Vision and Patt. Recog. (CVPR)*.

Haris, M., G. Shakhnarovich, and N. Ukita. (2021). "Deep back-projection networks for single image super-resolution". *IEEE Trans. on Patt. Anal. and Mach. Intel. (PAMI)*.

He, K., X. Zhang, S. Ren, and J. Sun. (2016a). “Deep residual learning for image recognition”. In: *IEEE/CVF Computer Vision and Patt. Recog. (CVPR)*.

He, K., X. Zhang, S. Ren, and J. Sun. (2016b). “Identity mappings in deep residual networks”. In: *European Conf. on Comp. Vision (ECCV)*.

Helmrich, C. R., M. Siekmann, S. Becker, S. Bosse, D. Marpe, and T. Wiegand. (2020). “XPSNR: A low-complexity extension of the perceptually weighted peak signal-to-noise-ratio for high-resolution video quality assessment”. In: *ICASSP*.

Henaff, O. J., R. L. Goris, and E. P. Simoncelli. (2019). “Perceptual straightening of natural videos”. *Nature Neuroscience*. 22(6): 984–991.

Heusel, M., H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. (2017). “GANs trained by a two time-scale update rule converge to a local Nash equilibrium”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 6626–6637.

Hochreiter, S. and J. Schmidhuber. (1997). “Long short-term memory”. *Neural Computation*. 9(8): 1735–1780.

Hornik, K., M. Tinchcombe, and H. White. (1989). “Multilayer feed-forward networks are universal approximators”. *Neural Networks, Pergamon Press*. 2: 359–366.

Hu, H., Z. Zhang, Z. Xie, and S. Lin. (2019). “Local Relation Networks for Image Recognition”. In: *IEEE/CVF Int. Conf. on Computer Vision (ICCV)*. 3464–3473.

Huang, G., Z. Liu, L. van der Maaten, and K. Q. Weinberger. (2017). “Densely connected convolutional networks”. In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR)*. Honolulu, HI.

Huang, Y., W. Wang, and L. Wang. (2015). “Bidirectional recurrent convolutional networks for multi-Frame super-resolution”. In: *Neural Information Processing Systems (NIPS)*. Vol. 28.

Huang, Y., W. Wang, and L. Wang. (2018). “Video super-resolution via bidirectional recurrent convolutional networks”. *IEEE Trans. on Patt. Anal. and Mach. Intel.* 40(4): 1015–1028.

Ignatov, A., N. Kobyshev, R. Timofte, K. Vanhoey, and L. V. Gool. (2017). "DSLR-Quality Photos on Mobile Devices with Deep Convolutional Networks". arXiv: 1704.02470 [cs.CV].

Irani, M. and S. Peleg. (1991). "Improving Resolution by Image Registration". *CVGIP: Graphical Models and Image Processing*. 53(3): 231–239.

Isobe, T., F. Zhu, X. Jia, and S. Wang. (2020). "Revisiting temporal modeling for video super-resolution". In: *British Mach. Vision Conf. (BMVC)*.

Jain, V. and S. Seung. (2008). "Natural image denoising with convolutional networks". In: *Advances in Neural Information Processing Systems (NIPS)*. 769–776.

Jang, D.-W. and R.-H. Park. (2019). "DenseNet with deep residual channel-attention blocks for single image super resolution". In: *IEEE/CVF Conf. on Comp. Vision and Pattern Recog. (CVPR) Workshops*.

Ji, X., Y. Cao, Y. Tai, C. Wang, J. Li, and F. Huang. (2020). "Real-World Super-Resolution via Kernel Estimation and Noise Injection". In: *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Jo, Y., S. Oh, J. Kang, and S. J. Kim. (2018). "Deep video superresolution network using dynamic upsampling filters without explicit motion compensation". In: *IEEE/CVF Conf. on Comp. Vis. Patt. Recog. (CVPR)*.

Jo, Y., S. Yang, and S. J. Kim. (2021). "SRFlow-DA: Super-resolution using normalizing flow with deep convolutional block". In: *IEEE/CVF Conf. Comp. Vision and Patt. Recog. (CVPR) Workshop*.

Johnson, J., A. Alahi, and L. Fei-Fei. (2016). "Perceptual losses for real-time style transfer and super-resolution". In: *European Conf. on Comp. Vision (ECCV)*.

Jolicoeur-Martineau, A. (2019). "The relativistic discriminator: A key element missing from standard GAN". In: *Int. Conf. on Learning Representations (ICLR)*.

Kancharla, P. and S. S. Channappayya. (2021). “Improving the visual quality of video frame prediction models using the perceptual straightening hypothesis”. *IEEE Signal Processing Letters*. 28: 2167–2171.

Kancharla, P. and S. S. Channappayya. (2022). “Completely blind quality assessment of user generated video content”. *IEEE Trans. on Image Processing*. 31: 263–274.

Kang, L., P. Ye, Y. Li, and D. Doermann. (2014). “Convolutional neural networks for no-reference image quality assessment”. In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR)*. 1733–1740.

Kappeler, A., S. Yoo, Q. Dai, and A. K. Katsaggelos. (2016). “Video super resolution with convolutional neural networks”. *IEEE Trans. on Computational Imaging*. 2(2): 109–122.

Keleş, O., A. M. Tekalp, J. Malik, and S. Kıranyaz. (2021a). “Self-organized residual blocks for image super-resolution”. In: *IEEE Int. Conf. on Image Processing*. Anchorage, Alaska, USA.

Keleş, O., M. A. Yılmaz, A. M. Tekalp, C. Korkmaz, and Z. Dogan. (2021b). “On the computation of PSNR for a set of images and video”. In: *Picture Coding Symp. (PCS)*.

Kim, J., J. K. Lee, and K. M. Lee. (2016). “Accurate image superresolution using very deep convolutional networks”. In: *IEEE/CVF Conf. on Comp. Vision and Pattern Recog. (CVPR)*.

Kim, S. Y., J. Lim, T. Na, and M. Kim. (2019). “3DSRNet: Video superresolution using 3D convolutional neural networks”. In: *IEEE Int. Conf. on Image Proc. (ICIP)*.

Kingma, D. P. and P. Dhariwal. (2018). “Glow: Generative flow with invertible 1x1 convolutions”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 10236–10245.

Kingma, D. P. and M. Welling. (2014). “Auto-encoding variational Bayes”. In: *Int. Conf. Learning Representations (ICLR)*.

Kiranyaz, S., J. Malik, H. B. Abdallah, T. Ince, A. Iosifidis, and M. Gabbouj. (2021). “Self-organized Operational Neural Networks with Generative Neurons”. *Neural Networks*. 140: 294–308.

Kirmemis, O. and A. M. Tekalp. (2018). “Effect of training and test datasets on image restoration and super-resolution by deep learning”. In: *Proc. of the EUSIPCO*. Rome, Italy. 514–518.

Kobyzev, I., S. Prince, and M. A. Brubaker. (2021). “Normalizing flows: An introduction and review of current methods”. *IEEE Trans. on Patt. Anal. and Mach. Intel. (PAMI)*. 43(Nov.): 3964–3979.

Köpüklü, O., N. Kose, A. Gunduz, and G. Rigoll. (2019). “Resource efficient 3D convolutional neural networks”. In: *IEEE/CVF Int. Conf. on Comp. Vision Workshop (ICCVW)*. 1910–1919.

Korkmaz, C. (2021). *Multi-Model and Multi-Stage Learned Image Super-Resolution*. MS Thesis, Koç University, Istanbul, Turkey.

Krizhevsky, A., I. Sutskever, and G. E. Hinton. (2012). “ImageNet classification with deep convolutional neural networks”. In: *Advances in Neural Information Processing Systems (NIPS)*. 1097–1105.

Lai, W.-S., J.-B. Huang, N. Ahuja, and M.-H. Yang. (2017). “Deep Laplacian pyramid networks for fast and accurate super-resolution”. In: *IEEE/CVF Conf. on Comp. Vision and Pattern Recog. (CVPR)*.

LeCun, Y., B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. D. Jackel. (1989). “Backpropagation applied to handwritten zip code recognition”. *Neural computation*. 1(4): 541–551.

Ledig, C., L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. (2017). “Photo-realistic single image super-resolution using a generative adversarial network”. In: *IEEE Conf. on Comp. Vision and Patt. Recog. (CVPR)*.

Lee, S., M. Choi, and K. M. Lee. (2021). “DynaVSR: Dynamic adaptive blind video super-resolution”. In: *IEEE/CVF Winter Conf. on Appl. of Comp. Vision*. 2093–2102.

Li, F., H. Bai, and Y. Zhao. (2020). “Learning a Deep Dual Attention Network for Video Super-Resolution”. *IEEE Trans. on Image Proc.* 29: 4474–4488.

Liang, J., J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte. (2021). “SwinIR: Image Restoration Using Swin Transformer”. In: *IEEE Int. Conf. on Computer Vision (ICCV) Workshops*.

Liao, R., X. Tao, R. Li, Z. Ma, and J. Jia. (2015). “Video superresolution via deep draft-ensemble learning”. In: *IEEE/CVF Int. Conf. on Comp. Vision (ICCV)*. 531–539.

Lim, B., S. Son, H. Kim, S. Nah, and K. M. Lee. (2017). "Enhanced deep residual networks for single image super-resolution". In: *IEEE/CVF Conf. on Comp. Vis. Patt. Recog. (CVPR) Workshop*.

Lim, B. and K. M. Lee. (2017). "Deep recurrent Resnet for video super-resolution". In: *Proc. of APSIPA Annual Summit and Conf.*

Liu, A. *et al.* (2021a). "Blind Image Super-Resolution: A Survey and Beyond". arXiv: 2107.03055 [cs.CV].

Liu, C., H.-Y. Shum, and W. Freeman. (2007). "Face hallucination: Theory and practice". *Int. J. Comput. Vision*. 75(1): 115–134.

Liu, D., Z. Wang, Y. Fan, X. Liu, Z. Wang, S. Chang, X. Wang, and T. Huang. (2018). "Learning temporal dynamics for video super-resolution: A deep learning approach". *IEEE Trans. on Image Processing*. 27(7): 3432–3445.

Liu, H. *et al.* (2020). "Video super resolution based on deep learning: A comprehensive survey". arXiv: 2007.12928 [cs.CV].

Liu, X., J. van de Weijer, and A. D. Bagdanov. (2017). "RankIQA: Learning from Rankings for No-reference Image Quality Assessment". In: *Int. Conf. on Comp. Vision (ICCV)*.

Liu, Z., H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, F. Wei, and B. Guo. (2022). "Swin Transformer V2: Scaling Up Capacity and Resolution". In: *IEEE/CVF Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*.

Liu, Z., Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. (2021b). "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows". In: *IEEE/CVF Int. Conf. on Computer Vision (ICCV)*.

Lucas, A., S. L. Tapia, R. Molina, and A. K. Katsaggelos. (2019). "Generative Adversarial Networks and Perceptual Losses for Video Super-Resolution". *IEEE Trans. on Image Proc.* 28(7): 3312–3327.

Lugmayr, A. *et al.* (2019). "AIM 2019 Challenge on real-world image super-resolution: Methods and results". In: *Int. Conf. Comp. Vision (ICCV) Workshops*.

Lugmayr, A. *et al.* (2020a). "NTIRE 2020 Challenge on Real-World Image Super-Resolution: Methods and Results". arXiv: 2005.01996 [eess.IV].

Lugmayr, A., M. Danelljan, L. V. Gool, and R. Timofte. (2020b). "SRFlow: Learning the super-resolution space with onrmalizing flow". In: *Euro. Conf. Comp. Vision (ECCV)*.

Luo, Z., Y. Huang, S. Li, L. Wang, and T. Tan. (2020). "Unfolding the alternating optimization for blind super resolution". In: *Advances in Neural Information Processing Systems (NeurIPS)*. Vol. 34.

Ma, C. *et al.* (2017). "Learning a no-reference quality metric for single-image super-resolution". *Comp. Vision and Image Understand (CVIU)*.

Makansi, O., E. Ilg, and T. Brox. (2017). "End-to-end learning of video super-resolution with motion compensation". In: *German Conf. on Pattern Recognition (GCPR)*. Basel, Switzerland.

Manasa, K. and S. S. Channappayya. (2016). "An optical flow-based full reference video quality assessment algorithm". *IEEE Trans. Image Proc.* 25(June): 2480–2492.

Mannos, J. and D. Sakrison. (1974). "The effects of a visual fidelity criterion on the encoding of images". *IEEE Trans. Inform. Theory*. 20(4): 525–536.

McCulloch, W. and W. Pitts. (1943). "A logical calculus of the ideas immanent in nervous activity". *Bulletin of Mathematical Biophysics*. 5: 115–133.

Minsky, M. and S. Papert. (1969). *Perceptrons : An Introduction to Computational Geometry*. Cambridge, Mass., USA: M.I.T. Press.

Mittal, A., A. K. Moorthy, and A. C. Bovik. (2012). "No-Reference Image Quality Assessment in the Spatial Domain". *IEEE Trans. on Image Processing*. 21(12): 4695–4708.

Mittal, A., R. Soundararajan, and A. C. Bovik. (2013). "Making a completely blind image quality analyzer". *IEEE Signal Processing Letters*. 20(3): 209–212.

Monga, V., Y. Li, and Y. C. Eldar. (2021). "Algorithm unrolling". *IEEE Signal Processing Magazine*. Mar.: 18–44.

Nah, S. *et al.* (2019a). "NTIRE 2019 Challenge on Video Deblurring and Super-Resolution: Dataset and Study". In: *IEEE/CVF Conf. on Comp. Vision and Pattern Recog. (CVPR) Workshops*.

Nah, S. *et al.* (2019b). "NTIRE 2019 Challenge on Video Super-Resolution: Methods and Results". In: *IEEE/CVF Conf. on Comp. Vision and Pattern Recog. (CVPR) Workshops.*

Oh, K.-S. and K. Jung. (2004). "GPU implementation of neural networks". *Pattern Recognition.* 37(6): 1311–1314.

Pan, J., H. Bai, J. Dong, J. Zhang, and J. Tang. (2021). "Deep blind video super-resolution". In: *IEEE/CVF Int. Conf. on Comp. Vision (ICCV).* 4811–4820.

Papamakarios, G., E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan. (2021). "Normalizing flows for probabilistic modeling and inference". *Jou. of Machine Learning Research.* 22(Nov.): 1–64.

Papyan, V., Y. Romano, J. Sulam, and M. Elad. (2018). "Theoretical foundations of deep learning via sparse representations". *IEEE Signal Processing Magazine*: 72–89.

Patti, A. J., M. I. Sezan, and A. M. Tekalp. (1997). "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time". *IEEE Trans. on Image Processing.* 6(8): 1064–1076.

Pérez-Pellitero, E., M. S. M. Sajjadi, M. Hirsch, and B. Schölkopf. (2019). "Perceptual video super-resolution with enhanced temporal consistency". *arXiv preprint arXiv:1807.07930.*

Pérez-Pellitero, E., M. S. M. Sajjadi, M. Hirsch, and B. Schölkopf. (2018). "Photorealistic video super-resolution". In: *ECCV Workshops (PIRM).*

Ponomarenko, N., F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin. (2007). "On Between-coefficient Contrast Masking of DCT Basis Functions". In: *Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM), Scottsdale, Arizona.*

Prashnani, E., H. Cai, Y. Mostofi, and P. Sen. (2018). "PieAPP: Perceptual image-error assessment through pairwise preference". In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR).*

Rad, M. S., T. Yu, C. Musat, H. K. Ekenel, B. Bozorgtabar, and J.-P. Thiran. (2021). "Benefiting from Bicubically Down-Sampled Images for Learning Real-World Image Super-Resolution". In: *IEEE/CVF WACV.*

Ramachandran, P., N. Parmar, A. Vaswani, I. Bello, A. Levskaya, and J. Shlens. (2019). "Stand-Alone Self-Attention in Vision Models". In: *(NeurIPS)*.

Ranftl, R., A. Bochkovskiy, and V. Koltun. (2021). "Vision Transformers for Dense Prediction". In: *IEEE Int. Conf. on Computer Vision (ICCV)*.

Rezende, D. J. and S. Mohamed. (2015). "Variational inference with normalizing flows". In: *Int. Conf. on Machine Learning (ICML)*. Vol. 37. 1530–1538.

Rippel, O. and R. P. Adams. (2013). "High-dimensional probability estimation with deep density models". arXiv: 1302.5125 [cs.CV].

Ronneberger, O., P. Fischer, and T. Brox. (2015). "U-Net: Convolutional networks for biomedical image segmentation". In: *Medical Image Computing and Comp. Assisted Interven. (MICCAI)*. Vol. 9351. 234–241.

Rosenblatt, F. (1958). "The perceptron: A probabilistic model for information storage and organization in the brain". *Psychological Review*. 65(6): 386–388.

Rumelhart, D., G. Hinton, and R. Williams. (1986). "Learning representation by back-propagating errors". *Nature*. 323: 533–536.

Sajjadi, M. S., R. Vemulapalli, and M. Brown. (2018). "Frame-recurrent video super-resolution". In: *IEEE/CVF Conf. on Comp. Vis. Patt. Recog. (CVPR)*.

Salimans, T., I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen. (2016). "Improved Techniques for Training GANs". In: *Advances in Neural Information Processing Systems (NeurIPS)*.

Seshadrinathan, K. and A. C. Bovik. (2010). "Motion tuned spatio-temporal quality assessment of natural videos". *IEEE Trans. on Image Processing*. 19(2): 335–350.

Sheikh, H. R. and A. C. Bovik. (2006). "Image information and visual quality". *IEEE Trans. on Image Processing*. 15(2): 430–444.

Sheikh, H. R., A. C. Bovik, and G. de Veciana. (2005). "An information fidelity criterion for image quality assessment using natural scene statistics". *IEEE Trans. on Image Processing*. 14(12): 2117–2128.

Shen, Z., M. Zhang, H. Zhao, S. Yi, and H. Li. (2021). “Efficient attention: Attention with linear complexities”. In: *IEEE/CVF Winter Conf. on Applications of Computer Vision (WACV)*. 3531–3539.

Shi, W., J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. (2016). “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network”. In: *IEEE/CVF Conf. on Comp. Vis. Patt. Recog. (CVPR)*. 1874–1883.

Shi, X., Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo. (2015). “Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting”. In: *Int. Conf. on Neural Information Processing Systems*. Vol. 1. *NIPS’15*. Montreal, Canada: MIT Press. 802–810.

Shocher, A., N. Cohen, and M. Irani. (2018). “"Zero-Shot" Super-Resolution using Deep Internal Learning”. In: *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*.

Srivastava, R. K., K. Greff, and J. Schmidhuber. (2015). “Training very deep networks”. In: *Int. Conf. on Neural Information Processing Systems (NIPS)*. Vol. 2. 2377–2385.

Su, D., H. Wang, L. Jin, X. Sun, and X. Peng. (2020). “Local-global fusion network for video super-resolution”. *IEEE Access*. 8(Sept.): 172443–172456.

Tao, X., H. Gao, R. Liao, J. Wang, and J. Jia. (2017). “Detail-revealing deep video superresolution”. In: *IEEE/CVF Int. Conf. on Comp. Vision (ICCV)*.

Tekalp, A. M. (2015). *Digital Video Processing*. 2nd. USA: Prentice Hall. ISBN: 0133991008.

Tekalp, A. M. and I. Sezan. (1990). “Quantitative analysis of artifacts in linear space-invariant image restoration”. *Multidimensional Systems and Signal Processing*. 1: 143–177.

Tian, Y., Y. Zhang, Y. Fu, and C. Xu. (2020). “TDAN: Temporally deformable alignment network for video super-resolution”. In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recognition (CVPR)*. 3360–3369.

Tikhonov, A. and V. Arsenin. (1977). *Solution of Ill-posed Problems*. Winston.

Timofte, R. *et al.* (2017). “NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study”. In: *IEEE Conf. on Comp. Vision and Patt. Recog. (CVPR) Workshops*.

Timofte, R. *et al.* (2018). “NTIRE 2018 Challenge on single image superresolution: Methods and results”. In: *IEEE/CVF Conf. on Comp. Vision and Pattern Recog. (CVPR) Workshops*.

Timofte, R., R. Rothe, and L. J. V. Gool. (2016). “Seven ways to improve example-based single image super resolution”. In: *IEEE Conf. on Comp. Vision and Patt. Recog. (CVPR)*.

Tong, T., G. Li, X. Liu, and Q. Gao. (2017). “Image super-resolution using dense skip connections”. In: *Int. Conf. on Comp. Vision (ICCV)*.

Tran, D., L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. (2015). “Learning spatiotemporal features with 3D convolutional networks”. In: *Int. Conf. Comp. Vision (ICCV)*. 4489–4497.

Trussell, H. J. and R. Civanlar. (1984). “The feasible solution in signal restoration”. *IEEE Trans. on Acoustics, Speech, and Signal Processing*. 32(2): 201–212.

Ulyanov, D., A. Vedaldi, and V. Lempitsky. (2020). “Deep image prior”. *Int. Jour. of Computer Vision*. 128(Mar.): 1867–1888.

Unterthiner, T., S. van Steenkiste, K. Kurach, R. Marinier, M. Michalski, and S. Gelly. (2019). “FVD : A new metric for video generation”. In: *Int. Conf. Learn. Represent. (ICLR) Workshop on Deep Generative Models for Highly Structured Data*.

Wang, X., K. C. Chan, K. Yu, C. Dong, and C. Change Loy. (2019). “EDVR: Video restoration with enhanced deformable convolutional networks”. In: *IEEE Conf. on Comp. Vision and Patt. Recog. (CVPR) Workshops*.

Wang, X., R. Girshick, A. Gupta, and K. He. (2018a). “Non-local neural networks”. In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR)*. 7794–7803.

Wang, X., K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy. (2018b). “ESRGAN: Enhanced super-resolution generative adversarial networks”. In: *European Conf. Comp. Vision (ECCV) Workshop*.

Wang, Z. and A. C. Bovik. (2009). "Mean squared error: Love it or leave it?" *IEEE Signal Processing Magazine*. 26(1): 98–117.

Wang, Z., A. Bovik, H. Sheikh, and E. Simoncelli. (2004). "Image quality assessment: from error visibility to structural similarity". *IEEE Trans. on Image Processing*. 13(4): 600–612.

Wang, Z., J. Chen, and S. C. H. Hoi. (2021). "Deep learning for image super-resolution: A survey". *IEEE Trans. on Patt. Anal. and Mach. Intel. (PAMI)*. 43(10): 3365–3387.

Wang, Z., E. Simoncelli, and A. Bovik. (2003). "Multiscale Structural Similarity for Image Quality Assessment". In: *Asilomar Conf. on Signals, Systems, and Computers, Pacific Grove, CA, USA*. 1398–1402.

Wang, Z., X. Cun, J. Bao, and J. Liu. (2022). "Uformer: A General U-Shaped Transformer for Image Restoration". In: *IEEE/CVF Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*.

Werbos, P. J. (1990). "Backpropagation through time: What it does and how to do it". *Proceedings of the IEEE*. 78(10): 1550–1560.

Winkler, C., D. Worrall, E. Hoogeboom, and M. Welling. (2019). "Learning likelihoods with conditional normalizing flows".

Xie, Y., E. Franz, M. Chu, and N. Thuerey. (2018). "TempoGAN: A temporally coherent, volumetric GAN for super-resolution fluid flow". *ACM Trans. on Graphics (TOG)*. 37(4): 1–15.

Xu, Y.-S., S.-Y. R. Tseng, Y. Tseng, H.-K. Kuo, and Y.-M. Tsai. (2020). "Unified dynamic convolutional network for super-resolution with variational degradations". In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR)*. 12 496–12 505.

Xue, T., B. Chen, J. Wu, D. Wei, and W. T. Freeman. (2019). "Video Enhancement with Task-Oriented Flow". *Int. Jour. of Computer Vision (IJCV)*. 127(8): 1106–1125.

Yang, F., H. Yang, J. Fu, H. Lu, and B. Guo. (2020). "Learning Texture Transformer Network for Image Super-Resolution". In: *CVPR*.

Yang, W., J. Feng, G. Xie, J. L. Liu, Z. Guo, and S. Yan. (2018). "Video super-resolution based on spatial-temporal recurrent residual networks". *Computer Vision and Image Understanding*. 168: 79–92.

Yang, X., W. Xiang, H. Zeng, and L. Zhang. (2021). "Real-world video super-resolution: A benchmark dataset and a decomposition based learning scheme". In: *IEEE/CVF Int. Conf. on Comp. Vision (ICCV) Workshop*. 4781–4790.

Ye, P. and D. Doermann. (2011). "No-reference image quality assessment using visual codebook". In: *IEEE Int. Conf. Image Process. (ICIP)*.

Ying, X., L. Wang, Y. Wang, W. Sheng, W. An, and Y. Guo. (2020). "Deformable 3D convolution for video super-resolution". *IEEE Signal Processing Letters*. 27: 1500–1504.

Yu, J., Y. Fan, and T. Huang. (2019). "Wide activation for efficient image and video super-resolution". In: *British Mach. Vision Conf., Cardiff Univ.*

Yuan, Y., S. Liu, J. Zhang, Y. Zhang, C. Dong, and L. Lin. (2018). "Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks". In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR) Workshops*.

Zhang, H., M. Cisse, Y. N. Dauphin, and D. Lopez-Paz. (2018a). "mixup: Beyond Empirical Risk Minimization". In: *Int. Conf. on Learning Representations (ICLR)*.

Zhang, K. *et al.* (2020a). "NTIRE 2020 Challenge on perceptual extreme super-resolution: Methods and results". In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR) Workshops*.

Zhang, K., Y. Li, W. Zuo, L. Zhang, L. van Gool, and R. Timofte. (2021). "Plug-and-play image restoration with deep denoiser prior". *IEEE Trans. on Patt. Anal. and Mach. Intel. (PAMI)*. June.

Zhang, K., W. Zuo, and L. Zhang. (2018b). "Learning a single convolutional super-resolution network for multiple degradations". In: *EEE/CVF Conf. CVPR*.

Zhang, K., L. van Gool, and R. Timofte. (2020b). "Deep unfolding network for image super-resolution". In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR)*.

Zhang, K., W. Zuo, S. Gu, and L. Zhang. (2017). "Learning deep CNN denoiser prior for image restoration". In: *IEEE Conf. on Comp. Vision and Patt. Recog. (CVPR)*. 3929–3938.

Zhang, K., M. Sun, T. X. Han, X. Yuan, L. Guo, and T. Liu. (2018c). "Residual Networks of Residual Networks: Multilevel Residual Networks". *IEEE Trans. Circuits and Systems for Video Tech.* 28(6): 1303–1314.

Zhang, L., L. Zhang, and A. C. Bovik. (2015). "A feature-enriched completely blind image quality evaluator". *IEEE Trans. on Image Processing*. 24(8): 2579–2591.

Zhang, R., P. Isola, A. A. Efros, E. Shechtman, and O. Wang. (2018d). "The unreasonable effectiveness of deep features as a perceptual metric". In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recog.(CVPR)*. 586–595. DOI: 10.1109/CVPR.2018.00068.

Zhang, W., Y. Liu, C. Dong, and Y. Qiao. (2019). "RankSRGAN: Generative adversarial networks with ranker for image super-resolution". In: *Int. Conf. on Comp. Vision (ICCV)*.

Zhang, Y., K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu. (2018e). "Image Super-Resolution Using Very Deep Residual Channel Attention Networks". In: *IEEE/CVF ECCV*.

Zhang, Y., Y. Tian, Y. Kong, B. Zhong, and Y. Fu. (2018f). "Residual dense network for image super-resolution". In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR)*. 2472–2481.

Zhua, H., L. Lia, J. Wua, W. Donga, and G. Shia. (2020). "MetaIQA: Deep Meta-learning for No-Reference Image Quality Assessment". In: *IEEE/CVF Conf. on Comp. Vision and Patt. Recog. (CVPR)*.