Digital Signal Processing 119 (2021) 103285

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Deep learning approaches to inverse problems in imaging: Past, present and future

Santiago López-Tapia [a,*], Rafael Molina [a], Aggelos K. Katsaggelos [b]

[a] Dept. of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain
[b] Dept. of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA

ARTICLE INFO

Article history:
Available online 22 October 2021

Keywords:
Inverse imaging problems
Video super-resolution
Deep learning
Convolutional neural network

ABSTRACT

In recent years, deep learning-based models have gained momentum in imaging problems such as image and video super-resolution, image restoration or inpainting. The analytical approaches that have traditionally been used to solve image inverse problems have started to be replaced by deep learning ones, being outperformed in terms of efficacy and efficiency in many applications. However, deep learning-based models lack the adaptability of analytical models, thus making them unsuitable for dealing simultaneously with different forward image formation models. In contrast to analytical methods, deep learning models typically do not use domain knowledge and rely on learning the solution to the inverse problem from large data sets. This is making them susceptible to errors caused by the presence of degradations not seen during training. Hybrid models combining analytical and deep learning approaches have been introduced to solve such generalization issues while retaining the efficacy of deep learning models. In this work, we review deep learning and hybrid methods for solving imaging inverse problems, focusing on image and video super-resolution and image restoration. Furthermore, we discuss open problems in this area that would be of critical importance in the future, the challenges of applying deep learning models to solve them, and how future research could address them.

1. Introduction

An observed signal y can be frequently modeled as the result of applying a degradation operator A to the latent signal x. Both x and y signals are multidimensional sampled at one or more time, spatial or spectral positions. For example, x and y may represent sequences of frames in a video or still images.

The degradation operator, A, might model defocus blur introduced by the imaging device, blur due to the motion of the camera or objects in the scene, atmospheric turbulence or loss in general of spatio-temporal or spectral information. Such systems can exhibit various properties, such as linearity, shift invariance, and stability. The process of obtaining y given A and x is known as the forward model and can be written as:

$$y = A(x) + \epsilon, \tag{1}$$

where ϵ is the noise introduced during the process. At the same time, the recovery of the latent signal x given y and knowledge of A constitutes the inverse problem (if A is not known, the problem is classified as a "blind" recovery problem). There are numerous image and video processing applications requiring the solution of the inverse problem to recover x. Fig. 1 illustrates this process.

The difficulty of solving the inverse problem stems from the properties of A and ϵ. These usually determine the system to be ill-posed in the Hadamard sense [1]; that is, minor variations in the observed data result in significant variations in the solution.

The solution of inverse problems (in general referred to as recovery problems) has a long history of research and development in signal processing. These recovery problems are encountered under different names: restoration, Super-Resolution (SR), deblocking, inpainting or pansharpening to name a few. The type of recovery problem is determined by A and ϵ. For example, denoising is a particular case of image recovery where $A = I$ and $\epsilon \neq 0$. A more detailed description of the inverse problems mentioned here can be found in [2–4].

Analytical techniques for solving inverse problems have been studied for a long time. These approaches explicitly define the forward model (1), the criteria for obtaining a solution, and a solution approach, see [5–13]. Using this explicit modeling, the inverse problem is solved and an estimate of the latent signal x is obtained. At a high level, one can group recovery approaches

* Corresponding author.
E-mail addresses: sltapia@decsai.ugr.es (S. López-Tapia), rms@decsai.ugr.es (R. Molina), aggk@eecs.northwestern.edu (A.K. Katsaggelos).

https://doi.org/10.1016/j.dsp.2021.103285
1051-2004/

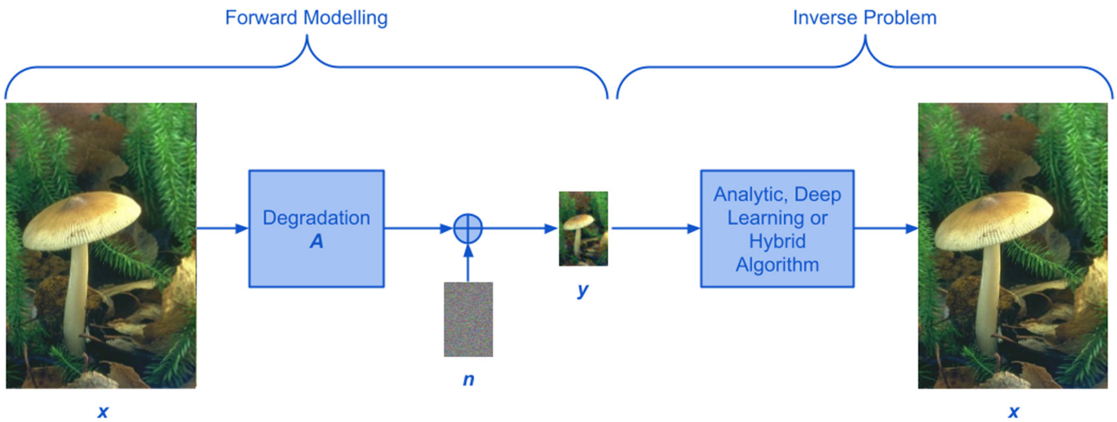

Fig. 1. Image Super-Resolution illustration of the forward modeling (1) and inverse problem. First, the degradation operator A is applied to the latent signal x and noise ϵ is added, obtaining the observation y. The objective of recovery is to solve the inverse problem and obtain an estimation of x given y. For illustration purposes, we have represented x and y as images and used as the degradation operator A blur and downsampling with scale factor 4.

into deterministic and stochastic ones. Deterministic techniques treat x as a deterministic signal and typically involve the use of an optimization criterion, such as the minimization of the l_2 error norm $||y - A(x)||_2^2$. Then, domain knowledge is incorporated into the solution process through regularization, by controlling, for example, the smoothness or the total variation of the reconstruction. In the case of stochastic approaches, the unknowns are treated as stochastic quantities and then a maximum likelihood, a maximum *a posteriori* (MAP) or a fully (hierarchical) Bayesian approach is followed. For stochastic approaches, domain knowledge is introduced into the problem through the probabilistic models used to describe the unknown quantities. Over the years, different probabilistic models and image priors have been proposed and used (see [14] for a review). Some examples of such priors are hyper-Laplacian priors [15], log-TV priors [16], mixture of Gaussians (MoG) [17], Super Gaussian (SG) [11] or Scale Mixture of Gaussian (SMG) [18]. In an analytical approach, optimization is performed for each new sample y, which provides it with high flexibility, under the generation conditions, to manage any degradation type. However, since closed-form solutions are not typically feasible, iterative optimization makes such approaches computationally expensive and tailored to a particular image.

In recent years, Deep Learning (DL) models, specifically those based on Deep Neural Networks (DNNs), have been used as a more efficient alternative to solve inverse problems. These are learning-based approaches that use an extensive training database of x and y image/sequence pairs to learn the solution to the inverse problem, i.e. they learn a mapping between the observations and the latent image [19,20]. In contrast to traditional learning-based methods mainly focused on building a dictionary or manifold space, DL-based models can learn and optimize the whole mapping process, including feature extraction. This gives them a clear advantage over traditional methods, which translates into a significant increase in performance, though at the cost of requiring large datasets to learn the mapping. However, this need for vast quantities of data has not impeded the application of DNNs to solve inverse problems since large databases can be simulated by applying the forward model.

When compared against analytical approaches, works in denoising [21], inpainting [22] and SR [23] have shown that DL models can outperform analytical ones while being significantly faster. Indeed, at a high level, learning approaches shift the computational burden to the learning phase. In contrast, the "testing" phase, i.e., the step of providing an estimate of x for a given y, is typically represented by a feed-forward network and it is, therefore, computationally efficient. Meanwhile, analytical techniques require solving an optimization problem for each new sample and are more computationally demanding. Generally speaking, the more sophisticated the modeling of the inverse problem, the more demanding the optimization process.

Incorporating domain knowledge into the solution (such as assuming a certain degree of smoothness) is far easier with analytical approaches since this modeling step represents an essential component of such approaches. This is not the case with DL models, which purely rely on the data. It is hard, in general, to incorporate such domain knowledge into the architecture of a neural network. Nevertheless, several approaches have proposed effective ways to incorporate domain knowledge into DL-based models in recent years. Such is the case of unfolding and generative modeling with neural networks [24–26].

Another disadvantage of DL-based models is that they lack the flexibility of analytical ones. More specifically, traditional DL models are not robust to degradations not seen during training, making them not well suited for problems where the DL model is expected to handle a large variety of forward image formation models, like Blind Image Deconvolution (BID). However, in recent years new approaches have been proposed to address such issue [27,11,28–33].

In this work, we first perform a comprehensive review of the literature of DL-based methods for solving image inverse problems, hoping that the reader will appreciate the technology and obtain valuable knowledge to utilize as guidance in solving such problems in their work. In this sense, our work is similar to Lucas et al. [34]. However, we also include the relevant advancements that appeared since then. Furthermore, we present a comprehensive discussion on current models limitations and challenges and how the field may evolve in the near future.

We will focus our analysis on image and video restoration and SR. We believe that these problems offer a good representation of the scenarios that emerge in other inverse problems. As such, the DL-based approximations developed to solve them should be a good fit to other image and video recovery tasks. The two problems which are the main focus of this work consist in recovering the latent signal x (image or sequence of video frames) from the observation y, which has been degraded by the operator A and noise ϵ. In the case of deconvolution or image restoration, A is a blurring matrix performing a convolution operation at each point of x with a Point Spread Function (PSF) k: $y = x \otimes k + \epsilon$, where $x \otimes k$ represents the convolution of x with the blur kernel k. Meanwhile, in SR A combines a convolution with the blur kernel k and a downsampling $\downarrow_s$ by a factor s. In this case, the variety of PSFs considered is much smaller than in image restoration, mainly of the Gaussian family. Notice that, in Video SR (VSR), the same degradation is considered to be shared by all frames in the sequence because the same system acquired them.

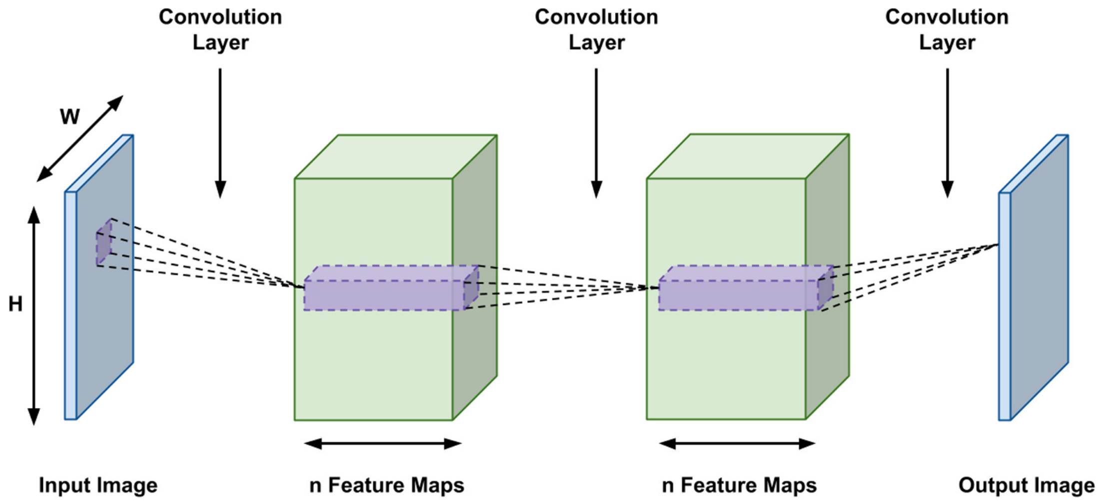

Fig. 2. A three-layer CNN with successive convolutional layers, where the spatial dimensions of the feature maps match those of the input and output images. Following each convolution there is a nonlinearity operation, not shown here.

Despite focusing on only two scenarios, due to the large amount of work in image and video restoration and SR, this article can not cover all of the approaches that use DL models for solving them. In other words, while the references described in this article by no means constitute an exhaustive list, we believe that the described techniques should provide the reader with a comprehensive overview of the ways DNNs may be employed to solve inverse problems in imaging. Finally, due to lack of space, fine details have been omitted. However, they can be acquired by referring to the original sources of the information.

The rest of the paper is organized as follows: Section 2 presents our review of DL-based models for image and video restoration and SR. Section 3 presents a discussion on open problems that we believe would be of critical importance in the future, the challenges of applying deep learning models to solve them, and how they could be addressed in the future. Finally, Sec. 4 concludes the paper.

2. Neural network architectures for image and video processing

Deep Learning (DL), also called hierarchical learning and deep structured learning, is "a class of machine learning techniques that exploits the use of many layers of non-linear information processing for supervised or unsupervised feature extraction and transformation, and pattern analysis and classification" [35]. In general, the following ML algorithms are considered part of DL [36]: Deep Neural Networks (DNN), Boltzmann Machines (deep and restricted), Deep Belief Networks and Deep Gaussian Processes [37]. From the DL models, DNNs are the most successful for high-level image and video related tasks like classification, segmentation, detection or parsing, defining the state-of-the-art for many of them [38–43]. Because of this fact, we will focus on DNN-based models for the rest of this paper.

A DNN is a parametric ML model that is formed by the stack of simple blocks. Each of these modules transforms its input to a new representation to be fed to the next block as input. The parameters of DNNs are trained end-to-end using the backpropagation algorithm [44] to optimize a loss function. In the DL literature, such modules are commonly referred to as layers. Each of these layers comprises several neurons that perform a linear transformation followed by a point-wise non-linear one. This non-linear function is called the activation function. How the neurons are connected determines the type of layer and the computations performed. The most basic of these layers connects all neurons with all the inputs and is called a fully connected layer. The most basic DNN, the multilayer perceptron (MLP), consists of two hidden fully connected layers.

The generic set of possible functions that a DNN can calculate is determined by its architecture. Thus, it is critical to pay attention to the model design when solving the inverse problem. One common architectural element chosen by most successful DNN-based models for image and video processing is the convolutional layer. Convolutional Neural Networks (CNNs) and their applications to inverse image and video processing problems are discussed in the following subsection.

2.1. End-to-end CNN for image and video processing

Although a fully connected DNN with a large number of neurons can approximate any function (universal approximation theorem [45]), in practice, the computation and data requirements make it not possible. Indeed, when dealing with highly structured data, such as images or videos, CNNs are typically the default model of choice. As we will see later in this section, CNNs are particularly suitable for processing these types of signals since they can easily extract the statistics of their input and learn the optimal features to solve the problem at hand.

CNNs were first introduced by LeCun et al. [46] in 1998 and distinguish themselves from fully connected DNNs by making use of convolutional layers. A convolutional layer does not connect all the inputs with all neurons. Instead, it applies multiple convolutions to the previous layer and learns the filter weights. The output of each convolutional layer is called a feature map. After applying the activation, these feature maps constitute the new representation of the data to be fed to the next layer. An example of a three-layer CNN architecture for image processing can be seen in Fig. 2. Notice that the spatial size of the feature maps is always the same and corresponds to the size of the input image. When using DNNs for image and video processing, the output of our model usually has to have the same spatial (and time) dimension as the input. One common approach to keep the dimensions of the output feature maps fixed to the size of the input to the convolutional layer is to pad the input with an appropriate number of zeros.

The original 2D-CNNs were designed to exploit the spatial dependencies among neighboring pixels in images. Thus, these and their application to 1D and 3D data excel at extracting useful information from local structures. They exhibit several advantages with respect to other DNNs in problems where local information is more relevant, like image and video processing:

1. The size and number of filters determine the number of parameters. Thus they require far fewer complex models to process high-dimensional data like images and videos. This reduction in the number of parameters simplifies the optimization problem.
2. The convolution operation that is the base of CNNs gives them translation invariance and locality, both advantageous properties when dealing with images and video data (see [36] for a more detailed explanation).
3. The convolution operation allows them to perform image and video feature extraction. Works in image classification, segmentation and detection have shown that these models can learn a powerful representation of images and extract multi-scale information, to the point that a simple linear classifier can solve the problem using such features.
4. CNNs can be designed to exhibit similarities with analytical methods like optimization-based iterative recovery methods or deconvolution methods, suggesting that CNN-based models can be powerful tools for solving inverse problems in imaging. Designing a CNN to perform computations similar to a well-established analytical model introduces domain knowledge in the model and increases its performance. See Section 2.5 for more details on such models.

For image and video super-resolution, the first work using CNNs applied three-layer end-to-end CNN, similar to the one in Fig. 2. Dong et al. [20] proposed a three-layer CNN for image SR that takes the interpolated LR image to estimate the HR one. Kappeler et al. [47] proposed a similar model for VSR, dealing with the temporal nature of video by feeding the network a window of motion-compensated frames at each time step. Further work in the field of SR focuses on adapting deeper architectures inspired by CNNs for object detection. This is the case of VDSR [48], which adapts the VGG architecture to SR. Meanwhile, VSR models focused on improving the extraction of inter-frame information by introducing motion-compensation into the model [49] or adding recurrent layers [50]. One of the most relevant advancements was the introduction of the interpolation step inside the network. By upsampling at the last layer of the network, be it through a transposed convolution [51], or using sub-pixel convolutions [52], the performance and time efficiency of SR and VSR models were dramatically improved [53,52,54].

Despite their earlier success with several image processing tasks [21,22,20], the use of CNN-based models for image restoration took longer to take off because of the sheer volume of various degradations that have to be tackled. Worth mentioning is the case of BID, where together with the image, the network must estimate the blur. Most CNN models proposed for image restoration are used in conjunction with other analytical methods or specialized in a specific type of blur. In fact, the first approximations limit CNN use to estimate PSF's specific parameters or extract useful image features. Such is the case of the models presented in [55,56]. The authors of [55] use a CNN to predict a motion vector per pixel and an analytical method to estimate the underlying image. In [56] a CNN is used to learn features that improve the estimation of the blurring kernel and image in the Fourier space.

More complex models are needed to increase the performance of CNN-based models for SR and tackle more complex and challenging image restoration scenarios. However, doing so with the basic CNN architecture presented above requires an exponential increase in the number of parameters and data requirements. Following previous works in image classification and object detection, residual CNNs were introduced. In the following subsection, we present these models and their applications to image and video processing.

2.2. Residual CNNs for achieving greater depth

In their work on blind deconvolution using CNNs, Hradis et al. [57] show that training deeper networks produced results of significantly better quality compared to the results obtained by shallow networks. This result is consistent with other findings in tasks such as image classification [58]. Deeper models can extract hierarchically complex features of increasing complexity. Not only that, but increasing the depth of the network increases the overall receptive field of the model, which provides more contextual information at each layer of the network.

However, training a very deep CNN was particularly challenging, primarily due to the non-existing effective methods for initializing DNNs and the vanishing of the gradient in the first layers of the network.

Recently, however, new parameter initialization strategies [60] and more efficient architectural design choices (like Batch Normalization (BN) [61]), have provided new possibilities for training deeper networks. More specifically, the use of residual blocks [40], has played a significant role in training very deep models by solving the vanishing gradient problem. The basic blocks of these models, named residual blocks, do not learn a new mapping function from one layer to the next. Instead, they learn a residual between two or more layers by adding a skip connection from the input of the residual block to its output, see Fig. 3.a.

The initial design of the residual block has been further improved. First, the authors of [62] and [63] showed that removing BN layers significantly improves the stability and performance of SR and VSR CNNs (see Fig. 3.b). Zhang et al. introduced in [64] the use of an attention mechanism to select for each image the most relevant feature maps and incorporated it into a new residual block (see Fig. 3.c). To further increase the depth of the model, the use of Residual Dense Blocks (RDBs) [65,66] was introduced for SR. An RDB is a modified Residual Block (RB) with dense connections between layers. These dense connections allow the reuse of features and provide better performance. See Fig. 3.d.

Residual CNNs have been shown to significantly outperform other CNN models in several image processing task like SR [54,62] and VSR [63]. For the particular case of BID, the first models to use end-to-end CNNs to solve this task are all residual CNNs. However, these CNNs have been applied only to specific blur types, like motion blur, without explicitly estimating the blur. In [67], a multi-scale residual CNN is used to eliminate non-uniform motion blur from images. The authors of [68] apply to the same problem a network with a dual attention mechanism. For video, a spatio-temporal residual recurrent architecture is presented in [69].

2.3. Training of end-to-end supervised CNNs

All previous end-to-end CNNs are fully supervised ML models. Therefore, to train them, they require both the input y and the desired output x. Despite needing massive amounts of data, this is not an issue in the context of solving inverse problems since these pairs can be synthetically generated by corrupting the original image x using the degradation A. Then a loss function measures the difference between the network output and x and this is used to update the model's parameters using backpropagation and a variant of gradient descend, being the most common nowadays Stochastic Gradient Descend and Adam [70]. During training, one critical step is choosing proper values for the hyperparameters that the optimization algorithm requires. Such is the case of the learning rate, learning rate decay schedule, and regularization strength.

The choice of the loss function is critical in solving an inverse problem using these types of CNNs. Because the objective of these models is to recover the original signal x, which is codified as an

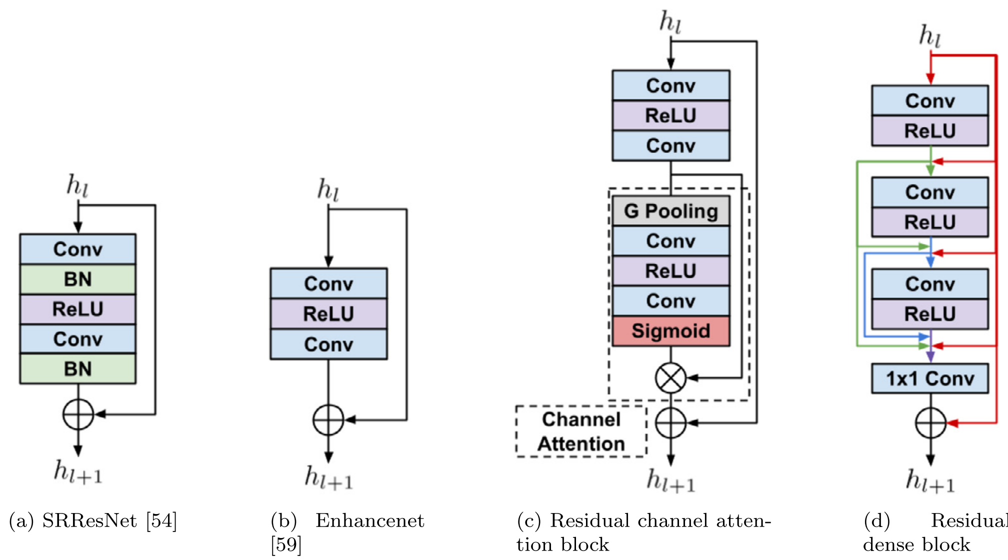

Fig. 3. Most commonly used residual blocks in image and video restoration and SR. Conv indicates a convolution layer and G Pooling refers to a global pooling operation.

image or a video, pixel-wise regression loss is most commonly used. Typically, most models employ an l2 distance, also known as Mean Squared Error (MSE), or similarly the Mean Absolute Error (MAE) and the Charbonnier loss [71].

Notice that, by training CNNs this way, the model solves the problem by risk minimization, which, coupled with the vast number of possible solutions, tends to produce blurred results. Indeed, it is well known in the image processing literature [59,54] that pixel-wise MSE does not fully reflect the difference in the visual quality between two images, favoring solutions with low frequencies. Several models solve this issue by using additional regularization terms in the loss function, such as total variation regularization. These regularizations impose image priors that discourage over-smooth solutions. Similar to this approach, several works have proposed [59,54,63,32] to train the CNN model adding different terms to the loss function together with the pixel-wise losses. One of the most successful approaches has been the introduction of a feature-based loss. This loss measures the difference between two images x and $\hat{x}$ using the features of a pre-trained CNN for object detection v_ψ:

$$l_{\text{feat}} = ||v_\psi(\hat{x}) - v_\psi(x)||_2^2. \tag{2}$$

Several works [59,54,72] have found that such features reflect properties of the human visual system and have even been used as a metric to determine the quality of the reconstruction in several image processing related tasks [72–74]. The motivation behind the use of such features is to capture a very robust representation of the input image. While the details of the input image are lost as a result of the multiple convolution operations, its content and structural information are stored in the feature maps. Thus, minimizing this loss forces the CNN to produce images with the same structure as the ground truth. In the case of BID, SR and VSR, their use has been proved to be effective in producing images with high perceptual quality [59,54,63,32]. However, they must be used in conjunction with other losses since models trained with only feature-based losses tend to produce images with high-frequency artifacts [63]. More precisely, they are usually used together with pixel-wise and adversarial losses to be described next.

2.4. Using generative adversarial networks for inverse problems

All CNN models introduced are discriminative models, meaning that they only model $\hat{p}_\theta(x|y) \sim p(x|y)$. In contrast, generative models model the data distribution $p(x)$ and, therefore, can sample from it. Generative Adversarial Networks (GANs) [75] are DL-based generative models introduced for approximating image and video distributions. GANs have proven to be able to generate realistic samples of different signal distributions, such as sounds and images of faces or bedrooms.

GANs consist of two networks: a generator g_θ and a discriminator or critic d_ϕ. The samples are produced by g_θ by mapping a random vector z (usually $z \sim N(0, I)$) to the data distribution. These models are trained to compete with each other:

$$\mathrm{L_d}(\phi;\theta) = -\mathbb{E}_{x\sim \mathrm{X}}\left[\log \mathrm{d}_\phi(x)\right] - \mathbb{E}_{\mathbf{y}\sim\mathbf{Y}}\left[\log(1-\mathrm{d}_\phi(\mathrm{g}_\theta(\mathbf{y})))\right] \tag{3}$$

$$\mathrm{L_g}(\theta;\phi) = -\mathbb{E}_{\mathbf{y}\sim\mathbf{Y}}\left[\log \frac{\mathrm{d}_\phi(\mathrm{g}_\theta(\mathbf{y}))}{1-\mathrm{d}_\phi(\mathrm{g}_\theta(\mathbf{y}))}\right]. \tag{4}$$

As it can be seen, the minimum value of the discriminator loss in (3) is obtained when a probability of 1 is assigned to all samples from the data distribution $\mathbb{E}_{x\sim \mathrm{X}}$ and 0 to the samples generated by the generator $g_\theta(y)$. Meanwhile, the generator loss (4) is minimum when the $d_\phi(g_\theta(y))$ is 1 for all generated samples. In other words, the discriminator is trained to distinguish between samples from the data distribution and those generated by the generators, while the generator must fool the discriminator. This training scheme allows GANs to learn the complex density associated with natural data distributions without defining them explicitly. Moreover, this occurs only through an indirect interaction with the training distribution via a discriminator network d_ϕ.

In the case of inverse problems, producing a sample from $p(x)$ is not enough since the problem at hand requires the ability to sample the most probable x giving an observation y. As such, instead of starting with a only random vector z, we could instead condition our generator g_θ on the observation y, which would then output a reconstruction $\hat{x} = g_\theta(y, z)$. Similarly to the generative case, the discriminator determines whether the prediction made by the generator $\hat{x}$ looks natural or not. This is the so-called conditional GAN (cGAN) framework and it is illustrated in Fig. 4.

In all cases, once the networks are trained, the discriminator is discarded and only the generator is used. In cGANs for image inverse problems, the adversarial loss is usually optimized in addition to other losses. This is because GANs, despite numerous efforts [76], are difficult to train, showing a tendency to collapse to specific patterns that are shown as high-frequency artifacts [63]. The use of supplementary losses, such as pixel-wise MSE, solves this

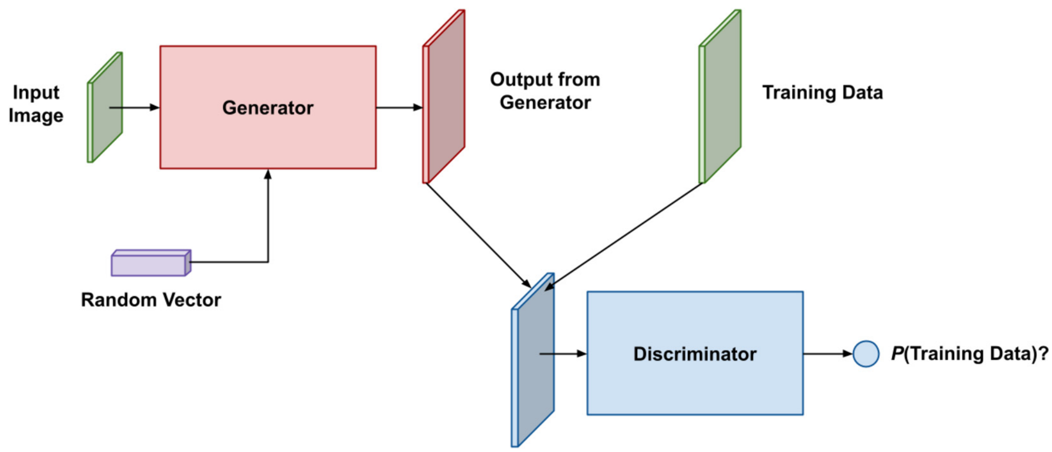

Fig. 4. The GAN framework for inverse problems. Given an observed image, the generator outputs a prediction for the output image, and the discriminator determines whether its input was synthesized by the generator, or comes from the training data.

issue and has allowed the introduction of GANs in SR [54], VSR [63] and BID [77,78]. Notice, however, that similarly to other CNN-based models, their application in BID is limited to only specific types of blurs, such as non-uniform motion blur.

As an alternative to previous approximations, Sonderby et al. [27] reformulated the SR problem as a maximum *a posteriori* estimation problem. The network's output is restricted to only those compatible with the observation and image formation model through the use of an affine projection:

$$\hat{x} = (I - A^+ A) g_\theta(y, z) + A^+ y, \tag{5}$$

where A^+ is the pseudo-inverse of the degradation A. Doing so, the relationship between x and y is explicitly introduced into the architecture of the network. This leaves only the data distribution $p(x)$ to be estimated, which can be done using a GAN. Thus, this model can be trained without the need for full supervision (data pairs y and x). However, as later shown by the authors in [32], these results are only applicable in practice for small scaling factors (smaller than two) since the GAN becomes unstable for higher ones causing too many artifacts.

In semi-supervised and unsupervised learning, models based on GANs have established the state-of-the-art for solving image inverse problems. In these problems, we have to map the observed degraded data to clean data without knowing the formation model nor having data pairs, only samples from both distributions. In these cases, a GAN can be used to establish a mapping directly between the two distribution [79] or another intermediary one [80,81]. Notice that supervised learning models cannot be used in these scenarios since we do not have data pairs nor a way for generating them.

2.5. Combining neural networks and analytical methods

In the previous section, we have discussed how image inverse problems can be addressed using end-to-end NNs. However, as explained in Section 1, despite their advantages over analytical methods, incorporating domain-knowledge into these models is harder and, because of that, they are not robust to degradations not seen during training. Thus, these models are not well suited for problems with a high variation in degradations. This is the main reason why in image and video restoration, these models have only been applied to specific types of blurs, like motion blur. Models combining both DL and analytical approaches have been proposed to solve the issues mentioned above. Several different approaches exist that combine both methods. Here, we will focus on the most common strategies. We will refer to these models as hybrid DL models to differentiate them from other DL-based approaches for the rest of the paper, which we will refer to as simply DL models

One of the key factors of DL-based models is that they are very effective in modeling complex data distributions, being able to correct for artifacts and noise when the same type of distortions has been seen during training. Therefore, they can refine the solution of analytical approaches and correct for artifacts and noise. In the case of image and video restoration, several works [28,82,31,33] have followed this approach. They apply a two-phase strategy: First, an analytical model calculates one or several initial estimates of the reconstructed image. Then, these images are used by a DL model to produce the final estimates. Following this approach, the authors of [28] propose using a U-net to denoise the estimation obtained using filtered back projection in medical imaging problems. Similarly, the authors of [31] use a U-net to remove the noise introduced by a Tikhonov deconvolution in BID of galaxy surveys. In contrast to previous works, in [82] the authors introduce a CNN model fed with several estimates of the latent image. These estimates are obtained using the method in [83] with different prior strengths and provide complementary information that the network combines into the restored image. López-Tapia et al. [33] perform image restoration and SR with a cascade CNN that takes a very noisy estimation of the deblurred LR image. This estimate is obtained using the Zhou et al. [11] method to estimate the PSF and a Wiener filter. Then, the network cleans the deblurred LR image and upsamples it to produce the HR estimation.

Apart from the previous hybrid models, a different group of methods use NNs as data priors and integrates them directly into the optimization process of the analytical method. They use variable splitting techniques such as the alternating direction method of multipliers (ADMM) [84] and half-quadratic splitting (HQS) [85] to split the inverse problem into two subproblems [86]: a regularized recovery one (*subproblem A*) which uses as penalty the squared Euclidean distance to an image. This image is estimated using an appropriate denoising technique (*subproblem B*), which in these methods is the NN. Both subproblems are more accessible than the original one. The denoising step can be seen as projecting the estimated image into the set of “real” images, and as such, a way to model the prior term. In this case, instead of hand-engineering the prior term, NNs are used to learn it from the data. In [87] and [88] the proximal operator [89] associated to *subproblem B* is replaced by a denoising CNN and applied to image denoising, restoration, inpainting and SR. In [90] the authors opted for a denoising autoencoder network inspired by the U-Net architecture. Huang et al. [91] improve this approach for BID by using a residual CNN to estimate only the noise. Zhang et al. [92] expand on the approximation of [88] for SR by tasking the CNN with

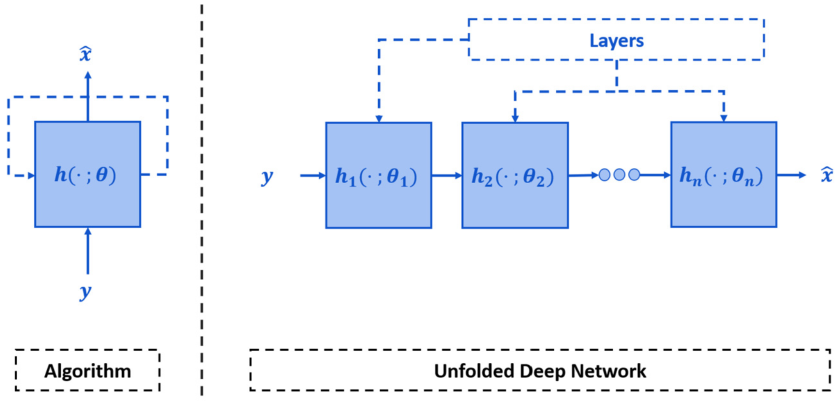

Fig. 5. A high-level illustration algorithm unfolding. A deep network (right) can be constructed from a corresponding iterative algorithm (left) by cascading its iterations $h(\cdot;\theta_i)$ a fixed number of times n.

upsampling and denoising. In contrast to the previous approaches, Chang et al. [93] use a combination of adversarial learning and a denoising autoencoder as the proximal operator.

A related approach to the previous one replaces the whole process or part of it with a randomly initialized DL model whose weights are optimized to fulfill the observation model. These internal learning methods have been introduced for SR and deblurring problems [94–96]. The success of such models lies on the use of priors naturally encoded in CNNs. This approach is known as Deep Image Prior [94]. As previously stated, these approaches consist of optimizing the weights of a network f_θ that takes the observation y as input to minimize the back-projection error $||y - \mathrm{f}_\theta(y)||_2^2$. Surprisingly, the network architecture is able to regularize the process and produce pleasing images without further regularization. Following this basic approach, Wang et al. [95] proposed a new method for image deconvolution, and Ren et al. [96] further adapted it to BID. Notice that deep image priors can be used in conjunction with explicit priors, as shown by the authors of [97,98].

Combining analytical and DL methods using the last two approaches allows for high flexibility and efficiency by integrating the DL model into the optimization. However, these methods do not provide a fully amortized approach; for each image, the solution must be found iteratively, which incurs a high computational cost since several forward passes through the NN must be performed.

Another common alternative for combining both analytical and DL-based models for image inverse problems is algorithm unfolding or unrolling. Unrolling methods were first proposed to develop fast neural network approximations for sparse coding [24]. The basic idea behind algorithm unfolding is to use the layers of a network to implement the steps of an iterative algorithm. The network parameters replace the algorithm's parameters (the model parameters and regularization coefficients). In contrast to the original algorithm, these parameters are not fixed. Instead, the model is trained end-to-end on data samples to learn them. Notice that the size of the model is fixed. Thus, passing through the network is equivalent to executing the iterative algorithm a finite number of times. See Fig. 5 for an illustration of the unfolding framework.

Following the algorithm unfolding framework to design a DL model presents a series of advantages over using a traditional DL model:

1. Since the network architecture mimics the operations of an iterative algorithm designed to solve the problem, the model's architecture introduces domain knowledge.
2. Because each layer performs a more complex operation than conventional neural network layers, the number of layers is smaller. This translates into models with far fewer parameters. Thus, these models need fewer data to train and generalize better.
3. In contrast to traditional DL models, the function of each layer is known, and the model can be easily interpreted.

These advantages, especially the potential of unrolled deep networks in developing efficient high-performance and yet interpretable network architectures, have caused the growth of this type of DL-based models' popularity. In the case of SR and restoration problems, the first unrolled deep network, LISTA [25], was proposed for image SR. The authors propose a network that implements the Iterative Shrinkage and Thresholding Algorithm (ISTA) steps, one of the most popular approaches in sparse coding. A deep network is constructed by mapping each iteration of ISTA to a layer and stacking them. Following the unfolding approach, the authors of [99] propose an end-to-end trainable unfolding network that leverages both learning-based methods and model-based methods, USRNet. They do this by integrating both subproblems of ADMM into a convolutional network with eight iterations.

For non-blind image deconvolution, RGDN [100] integrates both subproblems of ADMM into a recurrent convolutional network, where a standard CNN block replaces the gradient of the image prior unit. For the blind scenario, the architecture introduced in [56] is formed by concatenating multiple stages of essential blind image deblurring modules: feature extraction, kernel estimation and image estimation. Feature extraction is implemented with a CNN while the other two modules perform basic blind image deblurring operations using the features extracted. More recently, Li et al. [26] proposed a new network architecture product of unrolling a generalization of the traditional iterative total-variation regularization method in the gradient domain.

Up to now, we have presented a review of the most commonly used DL-based models for image and video restoration and SR. In the following subsection, we will introduce certain DL models that have yet to be applied to these problems but show promising properties.

2.6. Recent developments: inverse networks

One of the main limitations of current DL models is their lack of adaptability to degradations out of the distribution of the training data. Except for some models like [32], this issue is addressed through the combination of DL models with analytical approaches using the approximations described in Section 2.5. Several of these approximations use the DNNs to model complex data distributions. A different development related to the modelization of complex data distributions using DL-based methods needs to be mentioned.

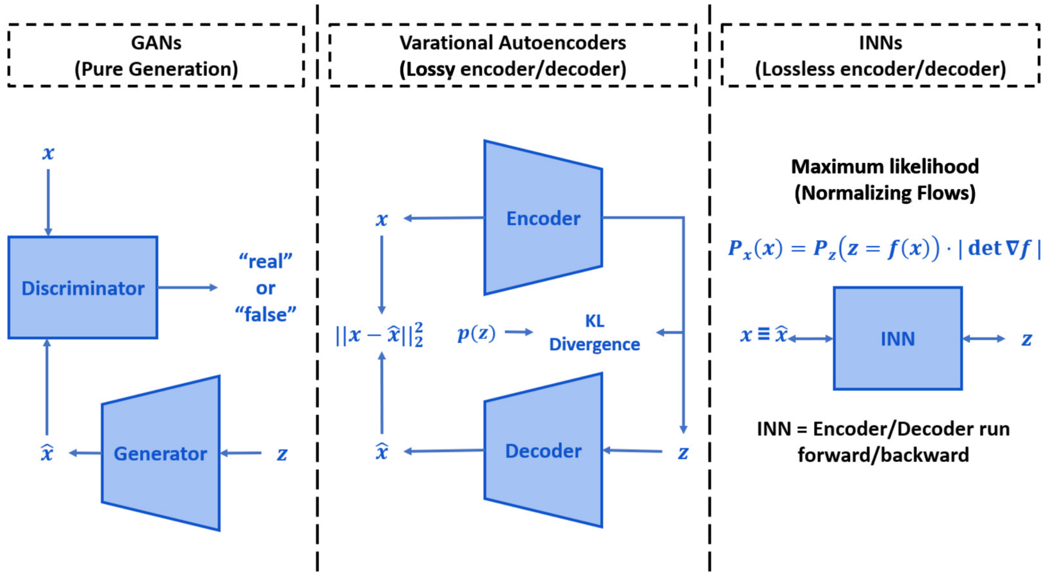

Fig. 6. A high-level illustration comparing GANs, VAEs and INNs. We focus on the overall architecture and the loss of information when performing projections between the image and latent space.

Inverse Neural Networks (INNs) are NNs that approximate bijective functions and have a forward and backward mapping. Similarly to autoencoders, they can code and decode. However, in contrast to autoencoders, this process is always lossless. Furthermore, they can be used to implement Normalizing Flows [101]. These allow them to be trained to maximize the likelihood of the data distribution $P_x(x)$:

$$P_x(x) = P_z(z = \mathrm{f}_\theta(x))|\det\nabla \mathrm{f}_\theta|, \tag{6}$$

where f_θ is the INN, z the codification of x in the latent space of distribution P_z. Notice that the distribution P_z is fixed. Thus as long as $|\det\nabla \mathrm{f}_\theta|$ can be calculated or approximated, the model can be trained. Maximizing the likelihood of the data directly and encode and decode losslessly are the main advantages of these models. Fig. 6 shows a comparison between INNs and the other most generative DL-based architecture, GANs and Variational Autoencoders (VAEs). GANs can only generate, not encode, and they approximate the data distribution indirectly by adversarial training. VAEs can lossy encode and decode and they do not directly maximize the likelihood but a lower bound. However, it is important to notice that most current INNs are significantly slower than conventional networks. The two most common INN architectures are those base on Affine coupling layers [102] and i-ResNets [103].

Currently, very few approaches based on INNs have been proposed for SR and BID. However, for SR, SRFlow [104] has shown the remarkable capabilities of such models. This model is not deterministic, producing not a point estimation but a distribution. Because of this, this model is able to balance the introduction of high frequencies, artifacts and blur. Not only that, but existing HR images can be mapped back to the latent space and used to improve the HR estimation of similar images. All without the need for retraining the model. The authors of [105] use these models as priors for analytical approaches for compressing sensing, inpainting and denoising. The results show that they are significantly more effective than GAN priors, avoiding mode collapse and being robust to out of distribution samples.

3. Open problems and future areas of research

In the previous sections, we have introduced DL models and their most common applications to inverse problems. DL-based models have shown two important advantages when compared to analytical approaches, mainly their efficacy and efficiency. We have, however, identified two scenarios that represent open problems, far less explored than others and that we believe will be important areas of research in the future. These are: processing images and videos in the "wild" and integrating DL models for image and video processing as part of more complex systems.

3.1. Processing images and videos in the "wild"

The first scenario addresses images captured in real-world applications where not all acquisition factors are controlled. As a result, such images and videos are affected by several degradations and noise. Thus, recovering them is a more challenging problem than studied restoration or super-resolution. Although some of the models for SR and BID presented in the previous section [99,33] are able to work with several degradations at the same time, these models are still far from being able to perform well in the "wild" setting. An example of this scenario are images taken by a surveillance drone. They suffer from multiple degradations, such as blurring and aberrations caused by the camera lens, atmospheric turbulence, motion blur, low resolution for distant objects, noise and compression artifacts due to limits in storage and transmission bandwidth. If parts of the target objects are occluded, inpainting would be required. Lightning and weather conditions must also be taking into account. Another example is the improvement of satellite images. These images are affected by atmospheric disturbance and occlusions caused by clouds. Furthermore, some bands of multispectral and hyperspectral images may suffer from low spatial and temporal resolution. It is also worth mentioning the restoration and color normalization of histopathological images. Several public challenges in recent years have addressed simpler but similar scenarios to the one described here for image and video restoration, SR and denoising [106–110].

When solving inverse problems with DL one faces mainly two issues. The first one is the wide variety of possible combinations of degradations that could affect the observation. As more degradations are taken into account, the necessary number of instances to represent the data space properly grows exponentially (curse of dimensionality). Most DL models lack the generalization capabilities to handle problems with a wide variety of degradations. In other words, they are not robust to degradations not present in the training set. Moreover, even if the initial conditions are restricted and a representative training dataset can be constructed, the conditions that affect the capture of image or video may change in the future. For example, in remote sensing, the sensor can change over time or from one dataset to another, causing a modification of the parameters of the image formation model. Even if it does not introduce new noise types or degradations, the images will differ significantly from those in the training set. This is known as "data drift" [111]. Several approaches have been introduced to solve such lack of generalization issues. Except for some models like [32], these approaches focus on the combination of DL models with analytical approaches using the methods described in Section 2.5. However, it is important to notice that they have not been applied to the solution of inverse problems in "wild". The most complex degradation solved using the models presented in Section 2.5 is a combination of motion blur, downsampling and noise [99,33]. Of all hybrids models, we believe that unfolding algorithms are very promising. They are easier to integrate into more complex systems because they require significantly less memory and computation than other hybrid approaches. A system composed of a combination of several of these unrolling methods could be one possible approximation to address this problem.

When processing images and videos in the "wild", the second issue is the lack of an effective method to obtain training data for supervised learning, which could be more challenging to address than in cases described above. Since the parameters of all degradations may not be known and some of the degradations may not have an explicitly known form, the use of the image formation model to obtain synthetic pairs of training data is not possible. For example, tissues in histopathological images may present occlusions due to folds in the tissue and other geometric deformations caused by the tissue extraction and scanning process. Furthermore, in some cases obtaining a record of the target simultaneously with the input is not possible. In the case of images taken to train a surveillance drone, the turbulence of the atmosphere and light and weather conditions cannot be controlled to produce a target and an input from the same scene. Without pairs of input and target data, supervised learning cannot be used. Since most DL models for solving inverse problems are trained in a supervised fashion, they can not be used. The use of GANs to achieve semi-supervised learning could be here quite effective [79–81]. However, adapting most models for GAN training is not an easy task. Not only that, but it is vital to take into account that GANs tend to be hard to train. Without specialized architectures and careful training, they will not converge.

3.2. Integration into more complex systems

The other relevant area of research is integrating image and video processing DL models as part of a more complex system for object detection, recognition or segmentation. These models are also known as "task-specific" models. In these cases, the perceptual quality of an image is less relevant than recovering "features" necessary for the system to perform well. We can find several examples where the quality of the image is not the main focus but the performance of the system which uses such image for classification or segmentation. Such is the case of color deconvolution and normalization [112,113] and restoration [114,115] of medical images. In a more general case, we can also find some open public challenges where this is the objective [108,109].

In this context, we have identified two issues that need to be addressed. The first one relates to the robustness and reliability of the models. Most models discussed in the previous section can only produce point estimates, lacking the capacity to produce a confidence estimate. Therefore, they cannot assess the quality of their output. This is to say, they cannot determine when the image has been restored properly. This issue is especially relevant with BID models, where a wrong estimation of the PSF can cause severe artifacts. When the objective of the inverse problem is recovering the latent image or video with the purpose of visualization by a human, the human observer can easily assess the quality of the reconstruction. However, this is not the case for an automatic system. Some artifacts and noise can severely affect the accuracy of a classifier. For instance, consider a license plate recognition system dealing with atmospheric turbulence and motion blur. If the reconstruction is not precise enough, digits can be confusing (for example, a 0 for an 8). In such applications, estimating the uncertainty of the estimation is key.

Few approximations for solving image inverse problems with DL have tackled uncertainty estimation. However, there exist several approximations that can be used with DL models. First, GANs and VAEs are non-deterministic models. By sampling, the mean and variance of the image distribution can be calculated, obtaining an estimation of the uncertainty. Using the same strategy, confidence estimations can be calculated with Bayesian NNs [116] by sampling dropout masks. An alternative to NNs is to adapt Gaussian Processes (GPs) to work with the raw signal directly [117,118]. However, this is a challenging problem. Recently Morales et al. [119] have introduced an alternative method that combines both NNs and GPs by using the GPs as non-linear activations, obtaining compelling results. A good alternative is to combine the feature extraction capabilities of a DL model with GPs. Finally, INNs, together with Normalizing Flows, can be used to implement models producing not a point estimation but a distribution [104].

The second and final issue withe the integration in complex systems is the lack of proper metrics to quantify and compare the performance of these models without testing the whole system. As can be seen from an analysis of the literature presented in Section 2, most models quantify their performance using metrics designed to evaluate the perceptual quality for human observers. Such is the case for the most common full-reference quality metrics, the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM), and also the no-reference quality metrics like BRISQUE [120], or NIQE [121]. However, the features relevant for the human visual system do not have to correlate to those necessary to perform the system's task. For example, deblurring the background of a blurred portrait photo of a person increases any of the previous figures of merits without improving the performance of a system for face recognition since the face has not been restored. One alternative that has been utilized in some applications [112,113] measures the quality of the estimation by evaluating the detection performance of a set of CNN models. Since most CNNs used for object detection and classification rely on textural features [122], this approximation produces a reasonable estimation of the performance of other CNNs. However, as shown in [122–124], there exist approximations that allow CNNs to extract shape-based features and improve the robustness of the model. In these cases, and when hand-crafted features are used, this method is unreliable. Unfortunately, at the moment, there is no straightforward solution.

4. Conclusions

DL models are currently the most used approximations to solve image inverse problems. The vast majority of these methods takes an end-to-end approach to solve the inverse problem by learning the function from a dataset of input and output pairs (see Section 2.1). The most common approach for training these methods is using regression losses and synthetically generated data. However, other approaches using perceptual losses, regularizations and even semi-supervised learning have been proposed. See sections 2.3 and 2.4.

The first approximations to increase the performance of these models focused mainly on increasing the complexity and improving the optimization of the DL model. However, as researchers started to tackle more complex inverse problems, like BID, the focus shifted towards improving these models by introducing domain knowledge. Although some DL models, like [27,32], have shown remarkable advances by introducing this domain knowledge through clever restrictions and manipulations of the architecture, the most successful approximations are hybrid ones. These new models have successfully combined the flexibility of analytical approximations and the effectiveness of DL models. Hybrid models can tackle inverse problems with a wide variety of degradations while outperforming analytical methods and DL models. However, this has come at the cost of higher computational requirements. We have presented these models in Section 2.5.

While the techniques described in this article are successful at solving inverse problems in imaging, surpassing the analytical state of the art, there still exist challenges that they have yet to solve. As we explain in Section 3.1, current approximations are not yet capable enough to be used in the "wild". In such a scenario, images and video are degraded by a combination of multiple degradations and noise types. The image formation model may even be only partially known. The interest in such cases is growing, and we believe that obtaining such models is going to be one of the main lines of research in solving image inverse problems in the future. To solve the problem, the combination of hybrid models, especially unfolded ones, and semi-supervised learning, like the one used with Cycle GANs in [79–81], will be of critical importance.

The second open problem that we have identified in Section 3.2 is the integration of DL models within other systems for object detection, recognition or segmentation. Most DL models produce only a point estimate without any uncertainty estimation. Thus, there is no way of knowing if the estimation is reliable. After all, by their nature, the solution to an inverse problem cannot be deterministic since there is an infinite number of possible solutions. Moreover, current metrics and techniques for assessing the performance of these models only consider the perceptual quality in regards to a human observer. We believe that, together with the development of new metrics, the integration of uncertainty estimation techniques, like the one proposed in [119], would be essential lines of future research.

Finally, and because of the two open problems presented in Section 3, we expect future research in DL-based approaches for inverse problems will depart from the traditional end-to-end mapping approach and instead focus on solving a concrete step of the formulated inverse problem. Therefore, they will converge to hybrid models between traditional analytical approaches and DL. We expect that with the introduction of new approaches, like inverse networks and Gaussian processes, the current limitations of these models, like their computation requirements, will be mitigated.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by project PID2019-105142RB-C22 funded by MCIN/AEI/10.13039/501100011033 and project D2-P20_00286 funded by Junta de Andalucía and FEDER "A way to make Europe".

References

[1] J. Hadamard, Lectures on Cauchy's Problem in Linear Partial Differential Equations, Yale University Press, New Haven, CT, 1923.
[2] A. Bovik, The Essential Guide to Image Processing, Academic, New York, 2009.
[3] A. Bovik, The Essential Guide to Video Processing, 2 ed., Academic, New York, 2009.
[4] A.K. Katsaggelos, Fundamentals of digital image and video processing, https://www.coursera.org/learn/digital, 2015.
[5] A.K. Katsaggelos, R. Molina, J. Mateos, Super resolution of images and video, Synth. Lect. Image Video Multimed. Proc. 1 (1) (2007) 1–134.
[6] S.D. Babacan, R. Molina, A.K. Katsaggelos, Variational Bayesian super resolution, IEEE Trans. Image Process. 20 (4) (2011) 984–999.
[7] H. Zhang, L. Zhang, H. Shen, A super-resolution reconstruction algorithm for hyperspectral images, Signal Process. 92 (9) (2012) 2082–2096.
[8] C. Liu, D. Sun, On Bayesian adaptive video super resolution, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2) (2014) 346–360.
[9] P. Ruiz, X. Zhou, J. Mateos, R. Molina, A. Katsaggelos, Variational bayesian blind image deconvolution: a review, Digit. Signal Process. 47 (December 2015) 116–127.
[10] X. Zhou, J. Mateos, F. Zhou, R. Molina, A.K. Katsaggelos, Variational Dirichlet blur kernel estimation, IEEE Trans. Image Process. 24 (12) (2015) 5127–5139.
[11] X. Zhou, M. Vega, F. Zhou, R. Molina, A.K. Katsaggelos, Fast Bayesian blind deconvolution with Huber super Gaussian priors, Digit. Signal Process. 60 (2017) 122–133.
[12] J. Serra, J. Mateos, R. Molina, A. Katsaggelos, Variational em method for blur estimation using the spike-and-slab image prior, Digit. Signal Process. 88 (May 2019) 116–129.
[13] F. Pérez-Bueno, M. López-Pérez, M. Vega, J. Mateos, V. Naranjo, R. Molina, A. Katsaggelos, A tv-based image processing framework for blind color deconvolution and classification of histological images, Digit. Signal Process. 101 (June 2020) 102727.
[14] Z. Chen, S.D. Babacan, R. Molina, A.K. Katsaggelos, Variational bayesian methods for multimedia problems, IEEE Trans. Multimed. 16 (4) (2014) 1000–1017.
[15] B. Zhang, R. Liu, H. Li, Q. Yuan, X. Fan, Z. Luo, Blind image deblurring using adaptive priors, in: Internet Multimedia Computing and Service, Communications in Computer and Information Science, Springer, Singapore, Aug. 2017, pp. 13–22.
[16] D. Perrone, R. Diethelm, P. Favaro, Blind deconvolution via lower-bounded logarithmic image priors, in: International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR), 2015.
[17] R. Fergus, B. Singh, A. Hertzmann, S.T. Roweis, W.T. Freeman, Removing camera shake from a single photograph, ACM Trans. Graph. 25 (3) (2006).
[18] S.D. Babacan, R. Molina, M.N. Do, A.K. Katsaggelos, Bayesian blind deconvolution with general sparse image priors, in: A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, C. Schmid (Eds.), Computer Vision – European Conference on Computer Vision 2012, Berlin, Heidelberg, Springer Berlin Heidelberg, 2012, pp. 341–355.
[19] W.T. Freeman, T.R. Jones, E.C. Pasztor, Example-based super-resolution, IEEE Comput. Graph. Appl. 22 (2) (2002) 56–65.
[20] C. Dong, C.C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in: European Conference on Computer Vision, Springer, 2014, pp. 184–199.
[21] V. Jain, S. Sebastian, Natural image denoising with convolutional networks, in: D. Koller, D. Schuurmans, Y. Bengio, L. Bottou (Eds.), Advances in Neural Information Processing Systems 21, Curran Associates, Inc., 2009, pp. 769–776.
[22] J. Xie, L. Xu, E. Chen, Image denoising and inpainting with deep neural networks, in: Conference on Neural Information Processing Systems, 2012, pp. 1–9.
[23] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (Feb. 2016) 295–307.
[24] K. Gregor, Y. LeCun, Learning fast approximations of sparse coding, in: Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, pp. 399–406.
[25] Z. Wang, D. Liu, J. Yang, W. Han, T. Huang, Deep networks for image super-resolution with sparse prior, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 370–378.
[26] Y. Li, M. Tofighi, J. Geng, V. Monga, Y.C. Eldar, Efficient and interpretable deep blind image deblurring via algorithm unrolling, IEEE Trans. Comput. Imaging 6 (2020) 666–681.
[27] C. Sonderby, J. Caballero, L. Theis, W. Shi, F. Huszar, Amortised MAP inference for image super-resolution, in: International Conference on Learning Representations, 2017.
[28] K.H. Jin, M.T. McCann, E. Froustey, M. Unser, Deep convolutional neural network for inverse problems in imaging, IEEE Trans. Image Process. 26 (9) (2017) 4509–4522.

[29] J.C. Ye, Y. Han, E. Cha, Deep convolutional framelets: a general deep learning framework for inverse problems, SIAM J. Imaging Sci. 11 (2) (2018) 991–1048.
[30] K. Zhang, W. Zuo, L. Zhang, Learning a single convolutional super-resolution network for multiple degradations, in: IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[31] F. Sureau, A. Lechat, J.-L. Starck, Deep learning for a space-variant deconvolution in galaxy surveys, A & A 641 (2020) A67.
[32] S. López-Tapia, A. Lucas, R. Molina, A.K. Katsaggelos, A single video super-resolution gan for multiple downsampling operators based on pseudo-inverse image formation models, Digit. Signal Process. 104 (2020) 102801.
[33] S. López-Tapia, N.P. de la Blanca, Fast and robust cascade model for multiple degradation single image super-resolution, IEEE Trans. Image Process. 30 (2021) 4747–4759.
[34] A. Lucas, M. Iliadis, R. Molina, A.K. Katsaggelos, Using deep neural networks for inverse problems in imaging, IEEE Signal Process. Mag. 35 (1) (2018) 20–36.
[35] L. Deng, D. Yu, Deep learning: Methods and applications, Tech. Rep. MSR-TR-2014-21, Microsoft, May 2014.
[36] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016, http://www.deeplearningbook.org.
[37] A. Damianou, N. Lawrence, Deep gaussian processes, in: Proceedings of Machine Learning Research (PMLR), vol. 31, in: C.M. Carvalho, P. Ravikumar (Eds.), (Scottsdale, Arizona, USA), 29 Apr–01 May 2013, pp. 207–215.
[38] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 25, Curran Associates, Inc., 2012, pp. 1097–1105.
[39] F. Liu, G. Lin, C. Shen, CRF learning with CNN features for image segmentation, Pattern Recognit. 48 (Oct. 2015) 2983–2992.
[40] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
[41] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: a deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2017) 2481–2495.
[42] R. Girshick, Fast r-cnn, in: 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440–1448.
[43] S. Valipour, M. Siam, M. Jagersand, N. Ray, Recurrent fully convolutional networks for video segmentation, in: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), 2017, pp. 29–36.
[44] R. Rojas, The Backpropagation Algorithm, Springer Berlin Heidelberg, Berlin, Heidelberg, 1996, pp. 149–182.
[45] K. Hornik, M. Stinchcombe, H. White, Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, Neural Netw. 3 (5) (1990) 551–560.
[46] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, in: Proceedings of the IEEE, 1998, pp. 2278–2324.
[47] A. Kappeler, S. Yoo, Q. Dai, A.K. Katsaggelos, Video super-resolution with convolutional neural networks, IEEE Trans. Comput. Imaging 2 (2) (2016) 109–122.
[48] J. Kim, J. Kwon Lee, K. Mu Lee, Accurate image super-resolution using very deep convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1646–1654.
[49] J. Caballero, C. Ledig, A. Aitken, A. Acosta, J. Totz, Z. Wang, W. Shi, Real-time video super-resolution with spatio-temporal networks and motion compensation, arXiv preprint arXiv:1611.05250, 2016.
[50] Y. Huang, W. Wang, L. Wang, Bidirectional recurrent convolutional networks for multi-frame super-resolution, in: Advances in Neural Information Processing Systems, 2015, pp. 235–243.
[51] C. Dong, C.C. Loy, X. Tang, Accelerating the super-resolution convolutional neural network, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Computer Vision – ECCV 2016, in: Cham, Springer International Publishing, 2016, pp. 391–407.
[52] W. Shi, J. Caballero, F. Huszár, J. Totz, A.P. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1874–1883.
[53] C. Dong, C.C. Loy, X. Tang, Accelerating the super-resolution convolutional neural network, in: Proceedings of European Conference on Computer Vision, 2016.
[54] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al., Photo-realistic single image super-resolution using a generative adversarial network, arXiv preprint arXiv:1609.04802, 2016.
[55] J. Sun, W. Cao, Z. Xu, J. Ponce, Learning a convolutional neural network for non-uniform motion blur removal, in: Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on, IEEE, 2015, pp. 769–777.
[56] C.J. Schuler, M. Hirsch, S. Harmeling, B. Schölkopf, Learning to deblur, IEEE Trans. Pattern Anal. Mach. Intell. 38 (7) (2016) 1439–1451.
[57] M. Hradiš, J. Kotera, P. Zemčík, F. Šroubek, Convolutional neural networks for direct text deblurring, in: Proceedings of the British Machine Vision Conference (BMVC), BMVA Press, September 2015, pp. 6.1–6.13.
[58] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint, arXiv:1409.1556, 2014.
[59] M.S.M. Sajjadi, B. Schölkopf, M. Hirsch, EnhanceNet: single image super-resolution through automated texture synthesis, in: Computer Vision (ICCV), 2017 IEEE International Conference on, IEEE, 2017, pp. 4501–4510.
[60] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing human-level performance on imagenet classification, in: IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1026–1034.
[61] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: International Conference on Machine Learning, 2015, pp. 448–456.
[62] B. Lim, S. Son, H. Kim, S. Nah, K.M. Lee, Enhanced deep residual networks for single image super-resolution, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, vol. 1, 2017, p. 3.
[63] A. Lucas, S. López-Tapia, R. Molina, A.K. Katsaggelos, Generative adversarial networks and perceptual losses for video super-resolution, IEEE Trans. Image Process. 28 (July 2019) 3312–3327.
[64] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu, Image super-resolution using very deep residual channel attention networks, in: European Conference on Computer Vision (ECCV), 2018.
[65] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual dense network for image super-resolution, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2018, pp. 2472–2481.
[66] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, C.C. Loy, Esrgan: enhanced super-resolution generative adversarial networks, in: The European Conference on Computer Vision Workshops (ECCVW), September 2018.
[67] S. Nah, T.H. Kim, K.M. Lee, Deep multi-scale convolutional neural network for dynamic scene deblurring, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017, pp. 257–265.
[68] Y. Zhang, W. Li, Z. Li, T. Ning, Dual attention per-pixel filter network for spatially varying image deblurring, Digit. Signal Process. 113 (2021) 103008.
[69] T.H. Kim, K.M. Lee, B. Schölkopf, M. Hirsch, Online video deblurring via dynamic temporal blending network, in: 2017 IEEE International Conference on Computer Vision (ICCV), Oct. 2017, pp. 4058–4067.
[70] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, CoRR, arXiv:1412.6980 [abs], 2014.
[71] W.-S. Lai, J.-B. Huang, N. Ahuja, M.-H. Yang, Fast and accurate image super-resolution with deep laplacian pyramid networks, IEEE Trans. Pattern Anal. Mach. Intell. (2018).
[72] R. Zhang, P. Isola, A.A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep networks as a perceptual metric, in: CVPR, 2018.
[73] S. Anwar, S. Khan, N. Barnes, A deep journey into super-resolution: a survey, ACM Comput. Surv. 53 (May 2020).
[74] K. Zhang, M. Danelljan, Y. Li, R. Timofte, J. Liu, J. Tang, G. Wu, Y. Zhu, X. He, W. Xu, C. Li, C. Leng, J. Cheng, G. Wu, W. Wang, X. Liu, H. Zhao, X. Kong, J. He, Y. Qiao, C. Dong, X. Luo, L. Chen, J. Zhang, M. Suin, K. Purohit, A.N. Rajagopalan, X. Li, Z. Lang, J. Nie, W. Wei, L. Zhang, A. Muqeet, J. Hwang, S. Yang, J. Kang, S.-H. Bae, Y. Kim, Y. Qu, G.-W. Jeon, J.-H. Choi, J.-H. Kim, J.-S. Lee, S. Marty, E. Marty, D. Xiong, S. Chen, L. Zha, J. Jiang, X. Gao, W. Lu, H. Wang, V. Bhaskara, A. Levinshtein, S. Tsogkas, A. Jepson, X. Kong, T. Zhao, S. Zhao, P.S. Hrishikesh, D. Puthussery, C.V. Jiji, N. Nan, S. Liu, J. Cai, Z. Meng, J. Ding, C.M. Ho, X. Wang, Q. Yan, Y. Zhao, L. Chen, L. Sun, W. Wang, Z. Liu, R. Lan, R.M. Umer, C. Micheloni, Aim 2020 challenge on efficient super-resolution: methods and results, in: A. Bartoli, A. Fusiello (Eds.), Computer Vision – ECCV 2020 Workshops, Springer International Publishing, Cham, 2020, pp. 5–40.
[75] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in Neural Information Processing Systems, 2014, pp. 2672–2680.
[76] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. Courville, Improved training of Wasserstein gans, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Red Hook, NY, USA, Curran Associates Inc., 2017, pp. 5769–5779.
[77] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, J. Matas, Deblurgan: blind motion deblurring using conditional adversarial networks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 8183–8192.
[78] O. Kupyn, T. Martyniuk, J. Wu, Z. Wang, Deblurgan-v2: deblurring (orders-of-magnitude) faster and better, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 8877–8886.
[79] G. Lin, Q. Wu, L. Chen, L. Qiu, X. Wang, T. Liu, X. Chen, Deep unsupervised learning for image super-resolution with generative adversarial network, Signal Process. Image Commun. 68 (2018) 88–100.
[80] Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, L. Lin, Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018, pp. 814–81409.
[81] S. Chen, Z. Han, E. Dai, X. Jia, Z. Liu, X. Liu, X. Zou, C. Xu, J. Liu, Q. Tian, Unsupervised image super-resolution with an indirect supervised path, in: 2020

IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, pp. 1924–1933.
[82] S. Vasu, V.R. Maligireddy, A.N. Rajagopalan, Non-blind deblurring: handling kernel uncertainty with CNNs, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 3272–3281.
[83] D. Krishnan, R. Fergus, Fast image deconvolution using hyper-laplacian priors, in: Conference on Neural Information Processing Systems, 2009.
[84] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (2011) 1–122.
[85] D. Geman, C. Yang, Nonlinear Image Recovery with Half-Quadratic Regularization, 1995.
[86] M.V. Afonso, J.M. Bioucas-Dias, M.A.T. Figueiredo, Fast image recovery using variable splitting and constrained optimization, IEEE Trans. Image Process. 19(9) (3) (2010) 2345–2356.
[87] T. Meinhardt, M. Möller, C. Hazirbas, D. Cremers, Learning proximal operators: using denoising networks for regularizing inverse imaging problems, in: IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22–29, 2017, 2017, pp. 1799–1808.
[88] K. Zhang, W. Zuo, S. Gu, L. Zhang, Learning deep CNN denoiser prior for image restoration, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3929–3938.
[89] N. Parikh, S. Boyd, Proximal algorithms, Found. Trends Optim. 1 (Jan. 2014) 127–239.
[90] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, X. Lu, Denoising prior driven deep neural network for image restoration, in: Computing Research Repository, 2018, arXiv:1801.06756 [abs].
[91] L. Huang, Y. Xia, Joint blur kernel estimation and CNN for blind image restoration, Neurocomputing 396 (2020) 324–345.
[92] K. Zhang, W. Zuo, L. Zhang, Deep plug-and-play super-resolution for arbitrary blur kernels, in: IEEE Conference on Computer Vision and Pattern Recognition, 2019.
[93] J.H.R. Chang, C. Li, B. Póczos, B.V.K.V. Kumar, One network to solve them all — solving linear inverse problems using deep projection models, in: 2017 IEEE International Conference on Computer Vision (ICCV), Oct. 2017, pp. 5889–5898.
[94] D. Ulyanov, A. Vedaldi, V. Lempitsky, Deep image prior, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 9446–9454.
[95] Z. Wang, Z. Wang, Q. Li, H. Bilen, Image deconvolution with deep image and kernel priors, in: The IEEE International Conference on Computer Vision (ICCV) Workshops, Oct 2019.
[96] D. Ren, K. Zhang, Q. Wang, Q. Hu, W. Zuo, Neural blind deconvolution using deep priors, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), jun 2020, pp. 3338–3347.
[97] G. Mataev, P. Milanfar, M. Elad, Deepred: deep image prior powered by red, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, Oct 2019.
[98] R. Hyder, H. Mansour, Y. Ma, P.T. Boufounos, P. Wang, A consensus equilibrium solution for deep image prior powered by red, in: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 1380–1384.
[99] K. Zhang, L. Van Gool, R. Timofte, Deep unfolding network for image super-resolution, in: IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 3217–3226.
[100] D. Gong, Z. Zhang, Q. Shi, A. van den Hengel, C. Shen, Y. Zhang, Learning deep gradient descent optimization for image deconvolution, IEEE Trans. Neural Netw. Learn. Syst. (2020), https://doi.org/10.1109/TNNLS.2020.2968289, in press.
[101] D. Rezende, S. Mohamed, Variational inference with normalizing flows, in: F. Bach, D. Blei (Eds.), Proceedings of the 32nd International Conference on Machine Learning in, Lille, France, in: Proceedings of Machine Learning Research, vol. 37, PMLR, 07–09 Jul 2015, pp. 1530–1538.
[102] L. Dinh, J. Sohl-Dickstein, S. Bengio, Density estimation using real nvp, in: International Conference on Learning Representations (ICLR), 2017.
[103] J. Behrmann, W. Grathwohl, R.T.Q. Chen, D. Duvenaud, J.-H. Jacobsen, Invertible residual networks, in: K. Chaudhuri, R. Salakhutdinov (Eds.), Proceedings of the 36th International Conference on Machine Learning, Long Beach, California, USA, in: Proceedings of Machine Learning Research, vol. 97, PMLR, 09–15 Jun 2019, pp. 573–582.
[104] A. Lugmayr, M. Danelljan, L. Van Gool, R. Timofte, Srflow: learning the super-resolution space with normalizing flow, in: European Conference on Computer Vision (ECCV), 2020.
[105] M. Asim, M. Daniels, O. Leong, P. Hand, A. Ahmed, Invertible generative models for inverse problems: mitigating representation error and dataset bias, in: Proceedings of Machine Learning and Systems 2020, 2020, pp. 4577–4587.
[106] A. Lugmayr, M. Danelljan, R. Timofte, N. Ahn, D. Bai, J. Cai, Y. Cao, J. Chen, K. Cheng, S. Chun, W. Deng, M. El-Khamy, C.M. Ho, X. Ji, A. Kheradmand, G. Kim, H. Ko, K. Lee, J. Lee, H. Li, Z. Liu, Z.-S. Liu, S. Liu, Y. Lu, Z. Meng, P.N. Michelini, C. Micheloni, K.P. Prajapati, H. Ren, Y.H. Seo, W. Siu, K.-Ah Sohn, Y. Tai, R.M. Umer, S. Wang, H. Wang, T.H. Wu, H. Wu, B. Yang, F. Yang, J. Yoo, T. Zhao, Y. Zhou, H. Zhuo, Z. Zong, X. Zou, Ntire 2020 challenge on real-world image super-resolution: methods and results, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, pp. 2058–2076.
[107] A. Abdelhamed, M. Afifi, R. Timofte, M.S. Brown, Y. Cao, Z. Zhang, W. Zuo, X. Zhang, J. Liu, W. Chen, C. Wen, M. Liu, S. Lv, Y. Zhang, Z. Pan, B. Li, T. Xi, Y. Fan, X. Yu, G. Zhang, J. Liu, J. Han, E. Ding, S. Yu, B. Park, J. Jeong, S. Liu, Z. Zong, N. Nan, C. Li, Z. Yang, L. Bao, S. Wang, D. Bai, J. Lee, Y. Kim, K. Rho, C. Shin, S. Kim, P. Tang, Y. Zhao, Y. Zhou, Y. Fan, T. Huang, Z. Li, N.A. Shah, W. Liu, Q. Yan, Y. Zhao, M. Możejko, T. Latkowski, L. Treszczotko, M. Szafraniuk, K. Trojanowski, Y. Wu, P.N. Michelini, F. Hu, Y. Lu, S. Kim, W. Kim, J. Lee, J.-H. Choi, M. Zhussip, A. Khassenov, J.H. Kim, H. Cho, P. Kansal, S. Nathan, Z. Ye, X. Lu, Y. Wu, J. Yang, Y. Cao, S. Tang, Y. Cao, M. Maggioni, I. Marras, T. Tanay, G. Slabaugh, Y. Yan, M. Kang, H.-S. Choi, K. Song, S. Xu, X. Lu, T. Wang, C. Lei, B. Liu, R. Gupta, V. Kumar, Ntire 2020 challenge on real image denoising: dataset, methods and results, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, pp. 2077–2088.
[108] W. Scheirer, R. VidalMata, S. Banerjee, B. RichardWebster, M. Albright, P. Davalos, S. McCloskey, B. Miller, A. Tambo, S. Ghosh, S. Nagesh, Y. Yuan, Y. Hu, J. Wu, W. Yang, X. Zhang, J. Liu, Z. Wang, H.-T. Chen, T.-W. Huang, W.-C. Chin, Y.-C. Li, M. Lababidi, C. Otto, Bridging the gap between computational photography and visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. (2020) 1.
[109] P. Wei, H. Lu, R. Timofte, L. Lin, W. Zuo, Z. Pan, B. Li, T. Xi, Y. Fan, G. Zhang, J. Liu, J. Han, E. Ding, T. Xie, L. Cao, Y. Zou, Y. Shen, J. Zhang, Y. Jia, K. Cheng, C. Wu, Y. Lin, C. Liu, Y. Peng, X. Zou, Z. Luo, Y. Yao, Z. Xu, S.W. Zamir, A. Arora, S. Khan, M. Hayat, F. Khan, K.-H. Ahn, J.-H. Kim, J.-H. Choi, J.-S. Lee, T. Zhao, S. Zhao, Y. Han, B.-H. Kim, J. Baek, H. Wu, D. Xu, B. Zhou, W. Guan, X. Li, C. Ye, H. Li, H. Zhong, Y. Shi, Z. Yang, X. Yang, X. Li, X. Jin, Y. Wu, Y. Pang, S. Liu, Z.-S. Liu, L.-W. Wang, C.-T. Li, M.-P. Cani, W. Siu, Y. Zhou, R.M. Umer, C. Micheloni, X. Cong, R. Gupta, F. Almasri, T. Vandamme, O. Debeir, Aim 2020 challenge on real image super-resolution: methods and results, in: ECCV Workshops, 2020.
[110] A. Ignatov, R. Timofte, Z. Zhang, M. Liu, H. Wang, W. Zuo, J. Zhang, R. Zhang, Z. Peng, S. Ren, L. Dai, X. Liu, C. Li, J. Chen, Y. Ito, B. Vasudeva, P. Deora, U. Pal, Z. Guo, Y. Zhu, T. Liang, C. Li, C. Leng, Z. Pan, B. Li, B.-H. Kim, J. Song, J.C. Ye, J. Baek, M. Zhussip, Y. Koishekenov, H.C. Ye, X. Liu, X. Hu, J. Jiang, J. Gu, K. Li, P. Tan, B. Hou, Aim 2020 challenge on learned image signal processing pipeline, in: A. Bartoli, A. Fusiello (Eds.), Computer Vision – ECCV 2020 Workshops, Springer International Publishing, Cham, 2020, pp. 152–170.
[111] J. Quionero-Candela, M. Sugiyama, A. Schwaighofer, N.D. Lawrence, Dataset Shift in Machine Learning, The MIT Press, 2009.
[112] D. Tellez, G. Litjens, P. Bándi, W. Bulten, J. Bokhorst, F. Ciompi, J.V.D. Laak, Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology, Med. Image Anal. 58 (2019) 101544.
[113] F. Pérez-Bueno, M. López-Pérez, M. Vega, J. Mateos, V. Naranjo, R. Molina, A. Katsaggelos, A tv-based image processing framework for blind color deconvolution and classification of histological images, Digit. Signal Process. 101 (June 2020) 102727.
[114] T. Kohlberger, Y. Liu, M. Moran, P.-H.C. Chen, T. Brown, C. Mermel, J. Hipp, M.C. Stumpe, Whole-slide image focus quality: automatic assessment and impact on ai cancer detection, J. Pathol. Inform. 10 (2019).
[115] D. He, D. Cai, J. Zhou, J. Luo, S.-L. Chen, Restoration of out-of-focus fluorescence microscopy images using learning-based depth-variant deconvolution, IEEE Photonics J. 12 (2) (2020) 1–13.
[116] Y. Gal, Z. Ghahramani, Dropout as a Bayesian approximation: representing model uncertainty in deep learning, in: Proceedings of the 33rd International Conference on Machine Learning (ICML-16), 2016.
[117] H. He, W.-C. Siu, Single image super-resolution using gaussian process regression, in: CVPR 2011, 2011, pp. 449–456.
[118] H. Wang, X. Gao, K. Zhang, J. Li, Fast single image super-resolution using sparse gaussian process regression, Signal Process. 134 (2017) 52–62.
[119] P. Morales-Alvarez, D. Hernández-Lobato, R. Molina, J.M. Hernández-Lobato, Activation-level uncertainty in deep neural networks, in: International Conference on Learning Representations, 2021.
[120] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708.
[121] A. Mittal, R. Soundararajan, A. Bovik, Making a "completely blind" image quality analyzer, IEEE Signal Process. Lett. 20 (03 2013) 209–212.
[122] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F.A. Wichmann, W. Brendel, Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness, in: International Conference on Learning Representations, 2019.
[123] Y. Li, Q. Yu, M. Tan, J. Mei, P. Tang, W. Shen, A. Yuille, Cihang Xie, Shape-texture debiased neural network training, in: International Conference on Learning Representations, 2021.
[124] M.A. Islam, M. Kowal, P. Esser, S. Jia, B. Ommer, K.G. Derpanis, N. Bruce, Shape or texture: understanding discriminative features in {cnn}s, in: International Conference on Learning Representations, 2021.

Santiago López-Tapia received the bachelor's, master's and Ph.D.'s degrees in computer science from the University of Granada in 2014, 2015 and 2021, respectively. He is currently with the Visual Information Processing Group, Department of Computer Science and Artificial Intelligence, University of Granada. His research mainly focuses in the use of deep learning models for image restoration and classification.

Rafael Molina received the degree in mathematics and the Ph.D. degree in optimal design in linear models from the University of Granada, Granada, Spain, in 1979 and 1983, respectively. He was the Dean of the Computer Engineering School, University of Granada, from 1992 to 2002. In 2000, he joined the University of Granada, as a Professor of computer science and artificial intelligence. He was the Head of the Computer Science and Artificial Intelligence Department, University of Granada, from 2005 to 2007. His research focuses on using Bayesian modeling and inference in problems like image restoration, active learning, and machine learning.

Aggelos K. Katsaggelos received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Greece, in 1979, and the M.S. and Ph.D. degrees in electrical engineering from the Georgia Institute of Technology in 1981 and 1985, respectively. In 1985, he joined the Department of Electrical Engineering and Computer Science, Northwestern University, where he is currently a professor. He was the Ameritech Chair of information technology and the AT&T Chair, and he is the Joseph Cummings Chair. He has published extensively in the areas of signal processing and communications, computational imaging, and machine learning (over 300 journal papers).