IMAGE INPAINTING THROUGH NEURAL NETWORKS HALLUCINATIONS

Alhussein Fawzi, Horst Samulowitz, Deepak Turaga, Pascal Frossard

EPFL, Switzerland & IBM Research Watson, USA

ABSTRACT

We consider in this paper the problem of *image inpainting*, where the objective is to reconstruct large continuous regions of missing or deteriorated parts of an image. Traditional inpainting algorithms are unfortunately not well adapted to handle such corruptions as they rely on image processing techniques that cannot properly infer missing information when the corrupted holes are too large. To tackle this problem, we propose a novel approach where we rely on the *hallucinations of pre-trained neural networks* to fill large holes in images. To generate globally coherent images, we further impose smoothness and consistency regularization, thereby constraining the neural network hallucinations. Through illustrative experiments, we show that pre-trained neural networks contain crucial prior information that can effectively guide the reconstruction process of complex inpainting problems.

***Index Terms*—** Inpainting, neural networks, graph-based regularization, hallucination, image completion.

1. INTRODUCTION

Image inpainting is the task of reconstructing missing or deteriorated parts of an image. This fundamental problem has received significant attention from the image processing and computer vision communities throughout the years, and led to key advances in the field (see [6] and references therein). Traditionally, image inpainting is addressed either using diffusion-based approaches that propagate local structures into the unknown parts, or examplar-based approaches that construct the missing parts one pixel (or patch) at a time, while maintaining the consistency with the neighbourhood. Unfortunately, these approaches fail when the size of the missing part is large, and an additional component providing plausible *hallucinations* is therefore needed to tackle such challenging inpainting problems. This additional information might be provided by high-order models of natural images, such as those computed by *deep neural networks*.

Deep neural networks have recently led to seminal advances in many machine learning tasks, such as supervised image classification [7]. In supervised image classification, each image has a specific label, and neural networks are learned to approximate the image-label mapping through a cascade of elementary operations. When trained on huge training datasets (millions of images with thousands of labels), deep networks have *remarkable* classification performance that can occasionally surpass the human accuracy [11]. Interestingly, the availability of free pre-trained models furthermore makes deep networks particularly easy to use, as one can readily use such networks without having to train them from scratch. State-of-the-art deep neural networks can therefore be seen as easy-to-use blackbox models that have learned valuable information on millions of different images and thousands of labels during the training procedure, and have therefore built up important prior knowledge on the statistics of natural images. This prior knowledge can be extremely useful in solving difficult tasks, such as image inpainting.

In this paper, we explore an approach that relies on the *hallucinations* of a *pre-trained* deep neural network to solve the image inpainting problem. Deep neural networks seem indeed to be a suitable candidate to guide the image reconstruction process, as these networks have gathered a significant amount of information about natural images that is extremely valuable in intricate inverse problems. We specifically build a hybrid approach where the hallucination of a deep neural network is regularized using the Total Variation (TV) norm and a graph-based regularizer to guarantee the coherence of the result. Our approach provides encouraging results, thereby showing that the hallucinations of a pre-trained neural network can be very beneficial in the solution of inverse problems.

We note that specialized deep neural networks have previously been *trained* to solve inverse problems, such as denoising and blind inpainting [12]. Our approach is however fundamentally different from these works, as we consider instead a simpler approach, where we use a *discriminatively pre-trained* neural network to guide the image reconstruction. In other words, while previous works have trained networks to specifically solve inverse problem tasks, our work treats the deep network as a system containing significant information on the statistics of natural images. *Hallucinations* of deep neural networks (i.e., searching for an image that maximizes the score of a given neuron) have previously been studied in the goal of understanding neural nets and generating "Dream"-like scenes [8, 9]. However, this line of work is different from ours, as our objective here is image inpainting. Finally, in the very recent work [2], features from deep neural

978-1-5090-0746-2/16/$31.00 ©2016 IEEE

networks are used for image super-resolution. While features from intermediate layers are used, we instead follow a simpler approach here and use directly the last layer of the deep network in the image inpainting problem.

2. IMAGE INPAINTING BY HALLUCINATION

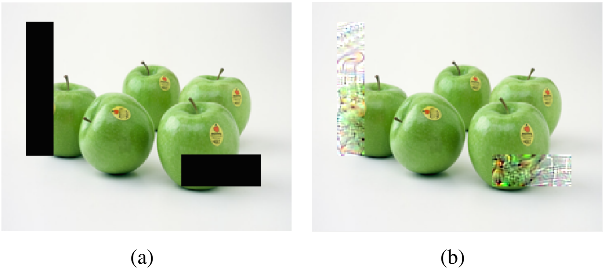
2.1. Pre-trained networks for image inpainting

We propose an approach where a pre-trained deep neural network is used to guide the reconstruction of missing pixels. Let $\mathcal{N}$ denote a neural network that is trained in a supervised fashion to *classify* images from a large number of categories. Formally, $\mathcal{N}$ is a mapping from the space of images $\mathbb{R}^{W \times H \times C}$ to the space $\mathbb{R}^L$, where L denotes the number of labels (or categories) that the neural network can classify. For a given image $I \in \mathbb{R}^{W \times H \times C}$, the network $\mathcal{N}$ outputs a vector $\mathcal{N}(I) \in \mathbb{R}^L$, where the l-th entry (denoted $\mathcal{N}_l(I)$) represents the *score* that image I is classified as label l. Our approach makes use of recent advances in deep networks to tackle the seemingly unrelated problem of image inpainting.

Denote by I an image with missing or deteriorated parts, and let I^* be the unknown image to recover. We assume that I^* has an associated label l that describes the main content of the objects in I^* (see Fig. 1 for an example image associated to label "Granny smith"). Throughout the paper, the label of an image will be estimated from the *corrupt* image I by taking the most likely class (that is, we set $l = \text{argmax}_l \mathcal{N}_l(I)$.). Hence, in this paper, the knowledge of l does *not* represent any additional burden or cost, as the label-estimation procedure is *automatic* and can be derived from the corrupt image and network. We denote by Ω the subset of $\mathbb{R}^2$ that contains the known part of the image. Our goal is to recover the pixels in the complement of Ω (denoted by Ω^c). We consider the following maximization problem:

$$\max_{\hat{I}} \mathcal{N}_l(\hat{I}) \text{ subject to } \hat{I}_\Omega = I_\Omega. \tag{1}$$

The above problem reconstructs the missing part Ω^c using the prior knowledge of the classifier, which has potentially seen millions of images during the training phase. This prior knowledge, incorporated in the network $\mathcal{N}$, makes it possible to *hallucinate* the missing pixels in order to maximize the probability of the image to be classified as l. Note however that the problem in Eq. (1) has an important caveat: the maximization of the classifier score does not necessarily result in "natural" completions of the corrupted image I. In fact, without further constraints on the reconstruction of the missing part of the image, the maximization of $\mathcal{N}_l(\hat{I})$ will result in "overfitting" the label l in the scene to maximize the probability of image classification as l. These *unrealistic hallucinations* can result in non-natural images, which maximize the number of objects labeled as l in the unknown part, but does not take into account the global structure of the image (see Fig. 1). The information provided by the neural network is therefore important to have a sketch of the shape of the missing part, but it does not necessarily result in natural completions that respect the coherence of the global image, and one should impose this constraint explicitly.


(a)

(b)

Fig. 1. Without regularization, an automatic completion of the missing part based on the neural network can result in "unrealistic hallucinations". The image on the right is found by maximizing the probability that it is classified as "Granny smith" according to Eq. (1).

2.2. Regularization strategy

We consider two regularization strategies in order to impose a natural-looking filling of the unknown part Ω^c. We first consider the Total Variation (TV) norm with the goal of removing undesirable details, while still preserving important details such as edges. Formally, the TV norm of an image is computed as follows:

$$f_{TV}(I) = \sum_{i,j} \sqrt{(I_{i+1,j} - I_{i,j})^2 + (I_{i,j+1} - I_{i,j})^2}.$$

The TV norm has been extensively used as a regularizer in several inverse problems, such as denoising [4] and super-resolution [1] due to its edge-preserving properties.

Besides the smoothness of the image imposed with the TV norm, we also introduce a regularizer that leverages our knowledge of the known part Ω. Specifically, we consider a graph $G = (V, E)$ where vertices represent pixels of the image, and edges represent some similarity measure between pixels in the known part (i.e., Ω) and the unknown part (i.e., Ω^c). Assuming that (i, j) and (i', j') represent respectively elements of Ω and Ω^c, we consider the following edge weights:

$$e_{(i,j) \to (i',j')} = \exp\left(- \frac{\| I^*_{B_{(i,j)}} - I^*_{B_{(i',j')}} \|_2^2}{\sigma^2} \right), \tag{2}$$

where $B_{(i,j)}$ is a small patch (of fixed size) around (i, j), and σ is a tunable parameter. A schematic representation of the graph is shown in Fig. 2. This notion of similarity has been extensively used for solving inverse problems with non local means for example [3]. It should be noted that the current definition of the edges e is ideal in the sense that it uses

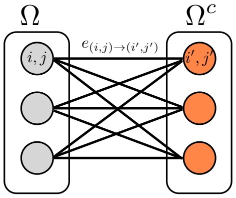


Fig. 2. Schematic representation of the graph used for the regularization.

the *unknown* image I^*. In practice, we instead opt for an iterative procedure where the graph is built from an estimate $\hat{I}$ obtained from the previous iteration (see Algorithm 1 for more details). This results in replacing I^* in Eq. (2) with the pre-computed estimate $\hat{I}$.

Having defined this graph, we define a *smoothness* prior on this graph through the following regularizer:

$$f_{\text{graph}}(I) = \mathcal{R}(I)^T L \mathcal{R}(I),$$

where L denotes the Laplacian matrix of the graph G, and $\mathcal{R}$ is the column-reshaped image. In words, the above smoothness assumption on the graph (together with our definition of similarity) encourages pixels in Ω^c to have similar pixel values to those in Ω when the neighbourhoods around the pixels are approximately similar.

2.3. Inpainting algorithm

Our final optimization problem, which involves the optimization of the hallucination term together with the graph and TV regularizers, is defined as follows:

$$\max_{\hat{I}} \mathcal{N}_l(\hat{I}) - \lambda f_{TV}(\hat{I}) - \gamma f_{\text{graph}}(\hat{I}) \quad (3)$$
$$\text{subject to } \hat{I}_\Omega = I_\Omega.$$

We solve this problem using the algorithm summarized in Algorithm 1. The algorithm specifically alternates between the construction of the graph G, and solving the optimization problem in Eq. (3). All inner optimization problems are handled using a simple projected (sub)gradient descent procedure with a fixed step size. More complex algorithms are likely to lead to better results, but we have opted for a gradient descent procedure in this work for the sake of simplicity.

3. EXPERIMENTAL RESULTS

3.1. Implementation details

In all experiments, we have used the celebrated VGG-19 deep convolutional neural network trained on the ILSVRC dataset [10]. This very deep network achieves close to state-of-the-art results on the challenging ImageNet challenge. To accelerate convergence (especially in the textureless region), we first initialize $\hat{I}$ through a simple isotropic diffusion process.

Algorithm 1 Inpainting algorithm

```
Input: corrupted image $I$,
Output: recovered image $\hat{I}$.
Compute label $l$ using the corrupted image
    $l \leftarrow \text{argmax}_l \mathcal{N}_l(I)$
Estimate $\hat{I}_1$ by solving the optimization problem
    $\hat{I}_1 \leftarrow \max_{\hat{I}} \mathcal{N}_l(\hat{I}) - \lambda f_{TV}(\hat{I})$ s.t. $\hat{I}_\Omega = I_\Omega$.
for all $i \in \{1, \dots, K\}$ do
    Construct the graph $G$ based on the estimation $\hat{I}_i$.
    Solve the optimization problem
        $\hat{I}_{i+1} \leftarrow \max_{\hat{I}} \mathcal{N}_l(\hat{I}) - \gamma f_{\text{graph}}(\hat{I}) - \lambda f_{TV}(\hat{I})$ s.t. $\hat{I}_\Omega = I_\Omega$.
end for
Return $\hat{I} \leftarrow \hat{I}_{K+1}$.
```

In what follows, the proposed approach is compared to the well-known approach for image inpainting in [5], as well as an isotropic diffusion approach to fill the pixels in Ω^c. For the approach in [5], we used the freely available MATLAB code available at `https://github.com/ikuwow/inpainting_criminisi2004`.

3.2. Results

We illustrate results on example images in Fig. 3. It can be observed that the diffusion process results in losing the edges, and the approach in [5] does not properly manage to reconstruct the features of the corrupted objects, as the masked regions are quite large. Note for example that the round shape of the apple, or the backwheel of the kart are particularly challenging shapes, and hence not correctly reconstructed by the technique in [5]. On the other hand, the proposed approach in many cases correctly *completes* the shape of the images by leveraging the concepts (e.g., of apples, karts, ...) learned by the deep network. We believe that, without this prior knowledge, it is extremely difficult for an algorithm to recover the correct shape. In that sense, the deep network can be very beneficial when the masks are large, and hallucination is therefore needed. To emphasize this point, we show in Fig. 4 a zoom on the backwheel of the kart by solving the problem in Eq. (3) *with* and *without* the presence of the hallucination term $\mathcal{N}_l$. Without the hallucination term $\mathcal{N}_l$, the procedure does not succeed in generating the correct round-shaped wheel, thereby showing that the neural-network term $\mathcal{N}_l$ is crucial in the global optimization. We stress however that relying *only* on the deep network hallucination term without

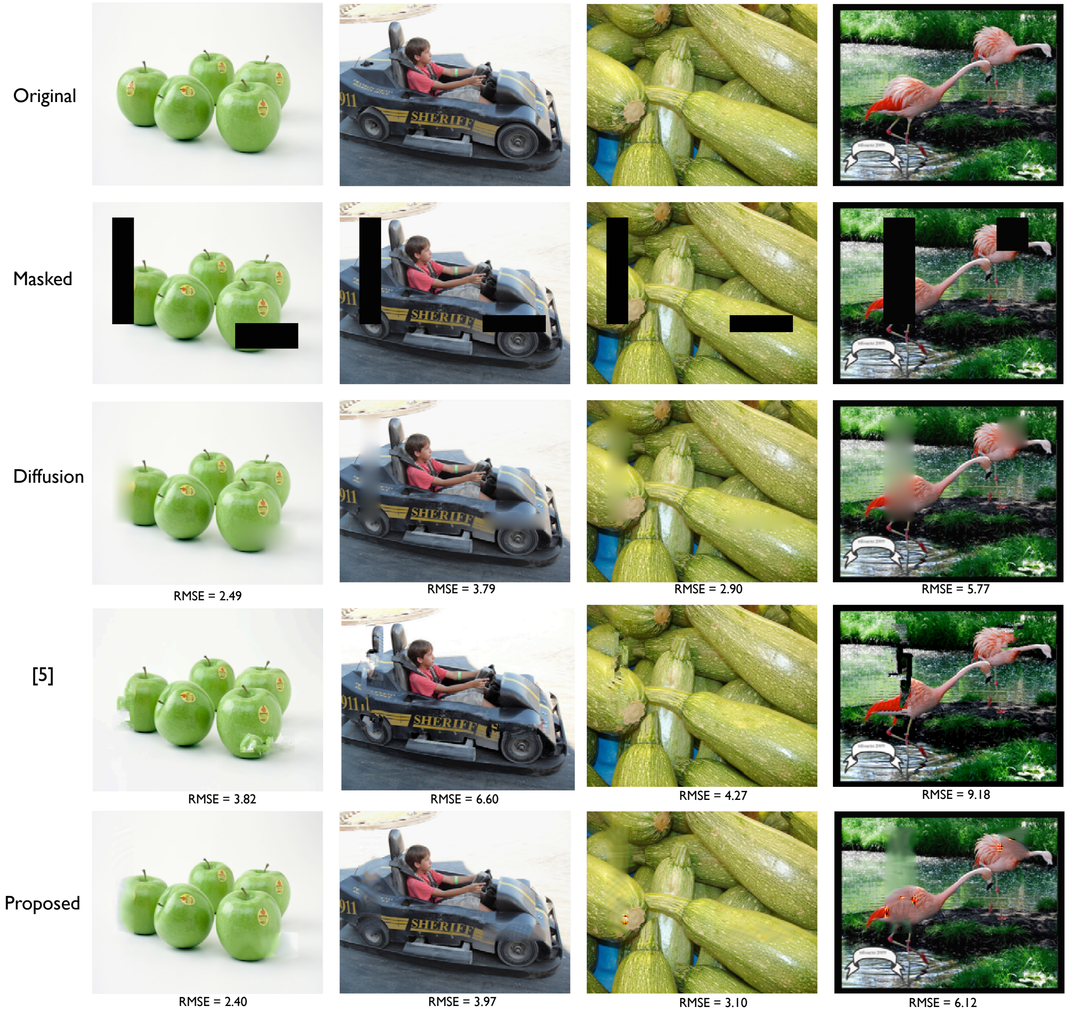
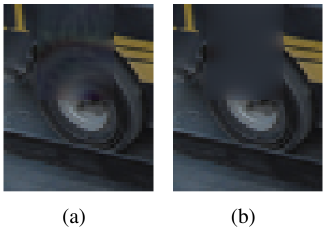


Fig. 3. Inpainting results on sample images.

further regularization, we obtain poor inpainting results (e.g., compare the inpainting result in Fig. 3 *with* regularization to Fig. 1 *without* regularization). Finally, for completeness, we also report the Root Mean Squared Errors (RMSE) between the original and recovered images in Fig. 3 obtained using the different approaches; note however that such a metric favors *smooth* solutions and is therefore only mildly indicative of the quality of the inpainting solution. Nevertheless, the quantitative measures confirm the visual observations showing that the proposed method compares favorably to the reference inpainting method in [5].

4. CONCLUSIONS

In this paper, we have proposed a novel approach for image inpainting through hallucinations of neural networks. To control the quality of the hallucination, we have considered simple regularizers that are key to guarantee the global smoothness and consistency of the image. However, other regularizers might provide results with better quality. We believe nevertheless that the main idea of this paper, which consists in using the knowledge of a pre-trained network to guide the reconstruction task, can be very beneficial in a broad number of image processing tasks. We hope this paper will lead to an exploration of this simple idea in challenging imaging tasks.



(a) (b)

Fig. 4. Reconstructed wheel *with* (left) and *without* (right) the hallucination term $\mathcal{N}_l$. Notice the wheel is correctly reconstructed in the former case, but not in the latter.

References

[1] S. D. Babacan, R. Molina, and A. K. Katsaggelos. "Total variation super resolution using a variational approach". In: *IEEE International Conference on Image Processing (ICIP)*. 2008, pp. 641–644.

[2] J. Bruna, P. Sprechmann, and Y. LeCun. "Super-Resolution with Deep Convolutional Sufficient Statistics". In: *arXiv preprint arXiv:1511.05666* (2015).

[3] A. Buades, B. Coll, and J.-M. Morel. "A non-local algorithm for image denoising". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 2. 2005, pp. 60–65.

[4] A. Chambolle. "An algorithm for total variation minimization and applications". In: *Journal of Mathematical imaging and vision* 20.1-2 (2004), pp. 89–97.

[5] A. Criminisi, P. Pérez, and K. Toyama. "Region filling and object removal by exemplar-based image inpainting". In: *IEEE Transactions on Image Processing* 13.9 (2004), pp. 1200–1212.

[6] C. Guillemot and O. Le Meur. "Image inpainting: Overview and recent advances". In: *IEEE Signal Processing Magazine* 31.1 (2014), pp. 127–144.

[7] A. Krizhevsky, I. Sutskever, and G. Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in Neural Information Processing Systems (NIPS)*. 2012, pp. 1106–1114.

[8] A. Mahendran and A. Vedaldi. "Understanding deep image representations by inverting them". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 5188–5196.

[9] A. Mordvintsev, M. Tyka, and C. Olah. "Inceptionism: Going deeper into neural networks, Google Research Blog". In: *Retreived June* 17 (2015).

[10] K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

[11] Y. Taigman et al. "Deepface: Closing the gap to human-level performance in face verification". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2014, pp. 1701–1708.

[12] J. Xie, L. Xu, and E. Chen. "Image denoising and inpainting with deep neural networks". In: *Advances in Neural Information Processing Systems (NIPS)*. 2012, pp. 341–349.