Large Inpainting of Face Images with Trainlets

Jeremias Sulam, *Student Member, IEEE,* and Michael Elad *Fellow, IEEE*

***Abstract*—Image inpainting is concerned with the completion of missing data in an image. When the area to inpaint is relatively large, this problem becomes challenging. In these cases, traditional methods based on patch models and image propagation are limited, since they fail to consider a global perspective of the problem. In this work, we employ a recently proposed dictionary learning framework, coined Trainlets, to design large adaptable atoms from a corpus of various datasets of face images by leveraging the Online Sparse Dictionary Learning algorithm. We therefore formulate the inpainting task as an inverse problem with a sparse-promoting prior based on the learned global model. Our results show the effectiveness of our scheme, obtaining much more plausible results than competitive methods.**

I. Introduction

Image inpainting is a data completion problem that aims to recover – or fill in – missing information in an degraded image. These areas can be given by individual missing pixels distributed along the image, or by more continuous regions resulting from scratches, folding or degradation of old photographs. In the extreme case where the area to inpaint is relatively large (also known as *hole-filling*), this problem becomes challenging [1].

This ill-posed problem, whose solution is often not even well-defined, has received considerable attention in recent years. Many inpainting approaches rely on Partial Differential Equations (PDF) [2], [3], variational formulations [4], exemplar-based methods [5], sparsity-enforcing priors [6], [7] and combinations of them [8], [9]. Despite their efficient performance, all these works are restricted to either small areas or to the task of object removal, by propagating and filling-in a proper background surrounding background.

Some problems, however, require a different approach. We shall focus in the specific problem of inpainting large areas of face images, as the one depicted in Figure 1. As one could foresee, traditional patch-based methods will not be effective in recovering or estimating the missing data. Diffusion based approaches, or those of content propagation, will also find this problem too challenging. In fact, all methods that do not consider a global prior model of the image to inpaint, will fail in generating data in accordance with the specific problem at hand.

All authors are with the Computer Science Department, Technion Israel Institute of Technology.

Fig. 1. Example of a inpainted image - left: Face image with missing eyes. Right: inpainted result obtained with the proposed approach.

The task of obtaining an adaptive global model for high dimensional signals is a hard problem. Some attempts include manifold learning techniques, as in [10], where the authors propose to learn an adaptable low-dimensional manifold for images. This work includes examples of inpainting on synthetic and texture data, but it is unclear if this method could provide a feasible alternative for real world face images. The recent work in [11], on the other hand, proposes the use of convolutional neural networks to train a global model to inpaint large holes in natural images. This network, however, was trained for general (street) images and it does not apply to our specific problem.

In this work, we propose to build such a global prior employing sparse representations modeling and dictionary learning. The problem of dictionary learning consists of adaptively learning a set of atoms which are able to represent real signals as sparsely as possible, and has been a popular topic in signal and image processing over the last decade [12], [13]. However, due to the computational constraints that this problem entails, all learning methods are typically applied on small patches from the image and not the image itself [14], [15]. In other words, attempting to obtain such a global dictionary with traditional dictionary learning algorithms would be infeasible.

A novel work which has circumvented this problem is the recently proposed Trainlets framework [16]. In this work, the authors proposed an Online Sparse Dictionary Learning (OSDL) algorithm that is able to obtain large adaptable atoms from natural images. Trainlets are built as linear (sparse) combination of atoms from a fast

and analytical dictionary, that of the novel Cropped Wavelets. This work [16] presented some initial results on the ability of Trainlets to sparsely approximate face images - indicating their effectiveness in modeling high dimensional data.

This way, we will formulate the inpainting task as an inverse problem regularized by a sparse prior under a global dictionary trained from publicly available datasets. Our results indicate that the proposed approach is able to synthesize missing information which is in a accordance with the global context of the image, yielding natural reconstructed faces.

II. Learning the Model

Sparse representations has shown to be a powerful prior in several inverse problems in image processing (see [12] for a thorough review). This model assumes that a signal $\mathbf{y} \in \mathbb{R}^n$ can be well approximated by a decomposition of the form $\mathbf{Dx}$, where $\mathbf{D}$ is a matrix of size $n \times m$ containing signal atoms in its columns – termed dictionary –, and a sparse vector $\mathbf{x} \in \mathbb{R}^m$. The problem of finding such a sparse vector is termed sparse coding, and can formally expressed as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{Dx}\|_2 \leq \epsilon, \tag{1}$$

where ϵ is an allowed deviation in the representation, and the ℓ_0 pseudo-norm is a count on the number of non-zero elements of its argument. When the dictionary is overcomplete ($m > n$), this is an NP-hard problem in general as it is combinatorial in nature. Yet, greedy algorithms and convex relaxation alternatives allow for good approximations of its solution in practice [18], [19].

When combined with the ability to learn the dictionary from real data, and for a specific task, this model has yielded a number of state of the art results [15], [20], [21], [22]. In its general form, the dictionary learning (DL) problem reads as follows

$$\arg\min_{\mathbf{D},\mathbf{X}} \frac{1}{2}\|\mathbf{Y} - \mathbf{DX}\|_F^2 \quad \text{s.t.} \quad \|\mathbf{x}_i\|_0 \leq p \quad \forall i, \tag{2}$$

where the matrix $\mathbf{Y}$ contains signal examples ordered column-wise. This problem inherits the non-convexity induced by the ℓ_0 pseudo-norm and adds the dictionary as a minimization variable. Though a series of different algorithm have been proposed [23], [14], [15], most method undertake an alternating minimization approach minimizing over $\mathbf{X}$ and $\mathbf{D}$.

However successful, the dictionary learning problem has been traditionally restricted to the domain of modeling small image patches, thus limiting the kind of problem these methods can address. This limitation arises mainly from computational constraints, but also from the fact that the degrees of freedom of the problem – and the amount data required to tuned these adaptively – become unmanageable as the dimension increases.

Some works have attempted to provide more efficient dictionary learning algorithms. The work presented in [24] proposes to lower the complexity of using (and learning) the dictionary by suggesting a andaptable but completely separable structure. Though this is an interesting and effective idea, the complete separability constraint is often too restrictive to represent general images of high dimensions, and its batch-learning algorithm is restricted to *small* training sets.

Recently, the work in [16] proposed an Online Sparse Dictionary Learning (OSDL) algorithm which is able to manage signals of dimensions in the order of the several thousands. This approach builds on the work of [25], which models the dictionary $\mathbf{D}$ as the product of a fast and efficient *base* dictionary, and an adaptable sparse factor $\mathbf{A}$. This lowers the complexity of both, the degrees of freedom of the problem and the computational cost of applying the dictionary. This way, the dictionary learning problem is formulated as

$$\min_{\mathbf{A},\mathbf{X}} \frac{1}{2}||\mathbf{Y} - \mathbf{\Phi AX}||_F^2 \quad \text{s.t.} \quad \begin{cases} ||\mathbf{x}_i||_0 \leq p & \forall i \\ ||\mathbf{a}_j||_0 = k & \forall j \end{cases}. \tag{3}$$

In particular, the authors in [16] employ a novel Cropped Wavelets dictionary as the operator $\mathbf{\Phi}$, leveraging the multi-scale analysis properties of wavelets while achieving a completely separable and border-effects free decomposition. In order to cope with the increase of training data, on the other hand, this work proposes a dictionary learning algorithm based on ideas from stochastic optimization [26]. In a nutshell, the algorithm performs sparse coding of a mini-batch of training examples with (Sparse) OMP [27], and then updates a subset of the dictionary atoms through a variation of the Normalized Iterative Hard Thresholding algorithm [28]. The reader is referred to [16] for a detailed description of this method.

Tackling the learning of a global model for face images in parcitular, we apply OSDL on a compendium of face images taken from different datasets. To increase the variability of the training data – and to obtain a more general model – we employ images taken from the Chinese Passport dataset used in [29] (both in its aligned and not-aligned formats), the Chicago Faces Database [30], the AT&T Faces Database[1], and the Cropped Yale Database [31]. All images were rescaled to a size of 100×100 pixels, and employed *as is*; i.e., there was no coherent scaling or alignment involved. All together, these amounted to a training set of roughly 19,000

[1]Freely available from AT&T Laboratories Cambridge's website.

images. OSDL took approximately 2 days to perform 40 data-sweeps[2]. We employed the Cropped Wavelets as the base dictionary (with Daubechies Wavelets with 4 vanishing moments), which has a redundancy of ≈ 1.7. The matrix $\mathbf{A}$ was chosen to be tall (under-complete), having 6,000 atoms in it. The atom sparsity was set to 1000; i.e., these are *only* $\approx 6\%$ sparse. We present some of the obtained atoms in Figure 2, where one can see that not only they resemble faces or face-features, but also the obtained variability between different sizes and configurations.

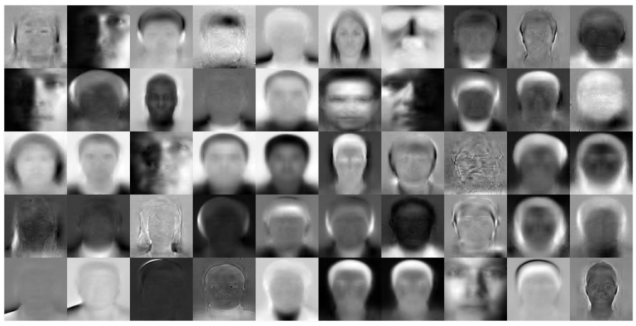

Fig. 2. Subset of the obtained atoms by OSDL.

III. Inpainting Formulation

Once the global model has been obtained, we move to describe in detail the inpainting formulation. Consider the original image $\mathbf{y}_0 \in \mathbb{R}^n$ ($n = 10,000$), and a mask $\mathbf{M}$, given by a binary matrix of size $l \times n$, where $l = c \cdot n$. This way, c denotes the fraction of the pixels that have not been removed (and remain) from the degraded image given by $\mathbf{y} = \mathbf{M}\mathbf{y}_0$.

Given this degradation model, the inpainting problem can be cast in terms of a pursuit by adding a sparse regularization term, resulting in a variant of the problem in Equation (1). In this case, however, we turn to a relaxation of this formulation moving from the ℓ_0 to the ℓ_1 norm. This way, the inpainting problem is given by the unconstrained optimization problem

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{M}\mathbf{D}\mathbf{x}\|_2 + \lambda \|\mathbf{x}\|_1, \tag{4}$$

where λ a the penalty parameter, compromising between the desired sparsity and the fidelity term. The shift from the non-convex formulation in Equation (1) to the relaxed form of the problem above is motivated by a practical aspect: in the inpainting problem, where one does not known a priori the number of non-zero elements needed to obtain a *good* reconstruction (or the equivalent ϵ threshold), it is easier to tune a penalty parameter λ. The number of non-zeros in $\mathbf{x}$ might be larger than those employed during the training, therefore making a greedy pursuit time consuming. In addition, we have found this ℓ_1 approach to yield solutions that are smoother, resulting in more naturally-looking inpainted areas.

Due to the convexity of the problem in Equation (4), a variety of algorithms can be employed to find its solution. Iterative shrinkage algorithms are particularly well-suited for this kind of problems, and we employ FISTA as the specific solver [32]. Our implementation of this method benefits from the relatively low-complexity of applying $\mathbf{D}$. Indeed, multiplying a vector by the dictionary (or its transpose) is never done explicitly. Instead, this is computed in terms of the product with the (very) sparse matrix $\mathbf{A}$ and the 1-dimensional dictionaries, which represent the separable operator $\mathbf{\Phi}$.

IV. Results

We now move to present the obtained inpainting results. For these experiments, we applied the method described in the previous section on a set of testing images, not included in the training set. In order to demonstrate the benefits of the proposed approach based on Trainlets, we compare with a number of other methods; namely: 1) the patch-propagation method of [6], which employs a sparse (patch) prior to inpaint the image, 2) a PCA (global) learned basis, and 3) the Separable Dictionary Learning Algorithm (SEDIL) [24], which also trains a global but separable dictionary. For this last method, we trained two 1-dimensional dictionaries of size 100×200 on the same training set, employing the code provided by the authors[3]. Note that both PCA and SEDIL obtain a set of global adaptive atoms by enforcing some constraints: orthogonality and separability, respectively.

The inpainting algorithm resulting from the minimization of Equation (4) depends on the parameter λ, which needs to be tuned for each particular case. In our experiments, and for a legitimate comparison, we run each method for a series of values of this parameter and then selected the most plausible results for each method separately[4]. The comparison with [6], on the other hand, is not entirely fair: inpaiting methods based on patch propagation are not expected to perform well in this challenging problem. Yet, we include them for completion and in order to demonstrate the intrinsic need of a global model.

[2]We run our experiment on a Windows computer with an Intel Xeon E5 CPU, with 64 Gb of RAM running Windows 64 bits. However, no parallel processing was used, and memory consumption did not exceed 16 Gb.

[3]Note that this is a batch method, and we employed 2,000 iterations. Training with SEDIL took approximately 2.5 days, resulting in both dictionary learning algorithms running for about the same time.

[4]Note that the selection of the best (most plausible) result is somewhat subjective, for which we have used our most fair judgment.

Fig. 3. Inpainting results. From left to right: masked image, patch propagation [6], PCA, SEDIL [24], Trainlets [16], and the original image.

We present a subset of our results in Figure 3, and more examples can be found in the supplementary material. As can be seen, Trainlets provide the best results – often making it hard to distinguish between the original and the synthetic inpainted image. As expected, the local method of [6] provides results that are not in agreement with the global context. The performance of SEDIL is limited, while PCA sometimes manages to recover somewhat of a natural result. Still, the constraints imposed by both of these two methods appear to be too restrictive for this problem. Some cases are particularly interesting: in the second image, where the glare in the glasses occlude the left eye, our approach manages to restore it; in the third image, we inpaint an eye which was not originally there due lighting conditions, still in a plausible manner.

V. Conclusion

We have presented a simple inpainting algorithm which exploits the representation power of Trainlets. By leveraging the OSDL algorithm, we were able to train a global model for a diverse collection of images. When this model is deployed with a sparse enforcing prior, we are able to inpaint large areas in face images obtaining very plausible reconstructions. An interesting observation is the fact that once a good global model is at our disposal, there is no need of any extra algorithmic manipulation of the data: there is no symmetry, exemplar-based copying or other form of external regularization enforced in the reconstruction. All this information is naturally captured by the learning process, alleviating the reconstruction stage. Exploring the ability of a similar approach in other kinds of inverse problems is an interesting direction of research, and constitutes part of ongoing work.

Acknowledgment

The research leading to these results has received funding from the European Research Council under European Unions Seventh Framework Programme, ERC Grant agreement no. 320649. The authors would like to thank Michael Zibulevsky for the insightful discussions that helped shape this work.

REFERENCES

[1] C. Guillemot and O. Le Meur, "Image inpainting: Overview and recent advances," *Signal Processing Magazine, IEEE*, vol. 31, no. 1, pp. 127–144, 2014. 1

[2] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pp. 417–424, ACM Press/Addison-Wesley Publishing Co., 2000. 1

[3] D. Tschumperlé, "Fast anisotropic smoothing of multi-valued images using curvature-preserving pde's," *International Journal of Computer Vision*, vol. 68, no. 1, pp. 65–82, 2006. 1

[4] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, and J. Verdera, "Filling-in by joint interpolation of vector fields and gray levels," *Image Processing, IEEE Transactions on*, vol. 10, no. 8, pp. 1200–1211, 2001. 1

[5] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *Image Processing, IEEE Transactions on*, vol. 13, no. 9, pp. 1200–1212, 2004. 1

[6] Z. Xu and J. Sun, "Image inpainting by patch propagation using patch sparsity," *Image Processing, IEEE Transactions on*, vol. 19, no. 5, pp. 1153–1165, 2010. 1, 3, 4

[7] O. G. Guleryuz, "Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part ii: adaptive algorithms," *Image Processing, IEEE Transactions on*, vol. 15, no. 3, pp. 555–571, 2006. 1

[8] A. Bugeau, M. Bertalmío, V. Caselles, and G. Sapiro, "A comprehensive framework for image inpainting," *Image Processing, IEEE Transactions on*, vol. 19, no. 10, pp. 2634–2645, 2010. 1

[9] O. Le Meur and C. Guillemot, "Super-resolution-based inpainting," in *Computer Vision–ECCV 2012*, pp. 554–567, Springer, 2012. 1

[10] G. Peyré, "Manifold models for signals and images," *Computer Vision and Image Understanding*, vol. 113, no. 2, pp. 249–260, 2009. 1

[11] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. Efros, "Context encoders: Feature learning by inpainting," 2016. 1

[12] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images," *SIAM Review.*, vol. 51, pp. 34–81, Feb. 2009. 1, 2

[13] J. Mairal, F. Bach, and J. Ponce, "Sparse modeling for image and vision processing," *Foundations and Trends® in Computer Graphics and Vision*, vol. 8, no. 2-3, pp. 85–283, 2014. 1

[14] M. Aharon, M. Elad, and A. M. Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *IEEE Trans. on Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006. 1, 2

[15] J. Mairal, F. Bach, and G. Sapiro, "Non-local Sparse Models for Image Restoration," *IEEE International Conference on Computer Vision.*, vol. 2, pp. 2272–2279, 2009. 1, 2

[16] J. Sulam, B. Ophir, M. Zibulevsky, and M. Elad, "Trainlets: Dictionary learning in high dimensions," *IEEE Transactions on Signal Processing*, vol. 64, no. 12, pp. 3180–3193, 2016. 1, 2, 4

[17] S. Mallat and Z. Zhang, "Matching Pursuits With Time-Frequency Dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, 1993.

[18] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal Matching Pursuit: Recursive Function Approximat ion with Applications to Wavelet Decomposition," *Asilomar Conf. Signals, Syst. Comput. IEEE.*, pp. 40–44, 1993. 2

[19] J. A. Tropp, "Just Relax : Convex Programming Methods for Identifying Sparse Signals in Noise," *IEEE Transactions on In*, vol. 52, no. 3, pp. 1030–1051, 2006. 2

[20] Y. Romano, M. Protter, and M. Elad, "Single image interpolation via adaptive nonlocal sparsity-based modeling," *IEEE Trans. on Image Process.*, vol. 23, no. 7, pp. 3085–3098, 2014. 2

[21] J. Sulam and M. Elad, "Expected patch log likelihood with a sparse prior," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, Lecture Notes in Computer Science, pp. 99–111, Springer International Publishing, 2015. 2

[22] R. Giryes and M. Elad, "Sparsity-based poisson denoising with dictionary learning," *Image Processing, IEEE Transactions on*, vol. 23, no. 12, pp. 5057–5069, 2014. 2

[23] K. Engan, S. O. Aase, and J. H. Husoy, "Multi-frame Compression: Theory and Design," *Signal Processing*, vol. 80, pp. 2121–2140, 2000. 2

[24] S. Hawe, M. Seibert, and M. Kleinsteuber, "Separable dictionary learning," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 438–445, 2013. 2, 3, 4

[25] R. Rubinstein, M. Zibulevsky, and M. Elad, "Double Sparsity : Learning Sparse Dictionaries for Sparse Signal Approximation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1553–1564, 2010. 2

[26] L. Bottou, "Online algorithms and stochastic approximations," in *Online Learning and Neural Networks*, Cambridge University Press, 1998. revised, Oct 2012. 2

[27] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient Implementation of the K-SVD Algorithm using Batch Orthogonal Matching Pursuit," *Technion - Computer Science Department - Technical Report.*, pp. 1–15, 2008. 2

[28] T. Blumensath and M. E. Davies, "Normalized iterative hard thresholding: Guaranteed stability and performance," *IEEE Journal on Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 298–309, 2010. 2

[29] O. Bryt and M. Elad, "Compression of facial images using the K-SVD algorithm," *J. Vis. Commun. Image Represent.*, vol. 19, pp. 270–282, May 2008. 2

[30] D. S. Ma, J. Correll, and B. Wittenbrink, "The chicago face database: A free stimulus set of faces and norming data," *Behavior research methods*, vol. 47, no. 4, pp. 1122–1135, 2015. 2

[31] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol. 23, no. 6, pp. 643–660, 2001. 2

[32] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009. 3