SEMI-UNBALANCED REGULARIZED OPTIMAL TRANSPORT FOR IMAGE RESTORATION

Simon Mignon, Bruno Galerne, Moncef Hidane, Cécile Louchet, Julien Mille

Semi-unbalanced regularized optimal transport for image restoration

Simon Mignon[a], Bruno Galerne[a,c], Moncef Hidane[b], Cécile Louchet[a], Julien Mille[b]
[a] Institut Denis Poisson – Université d'Orléans, Université de Tours, CNRS
[b] LIFAT – INSA Centre Val de Loire, Université de Tours
[c] Institut Universitaire de France (IUF)

***Abstract*—We consider in this paper the use of a penalty based on the theory of optimal transport (OT) in order to regularize inverse problems in imaging. The proposed approach is formulated in a variational setting and aims at promoting images whose patch distribution is close either to the one learned by a generative model, or to available uncorrupted reference patches. With the aid of numerical illustrations, we argue in favor of adopting an asymmetric form of unbalanced OT. We then provide details concerning the computation and the differentiation of the proposed penalty. Finally, we detail the application of our approach to a particular super-resolution setting.**

I. Introduction

Restoration problems are still today an important research topic in image processing and cover a wide range of applications. They correspond to the estimation of a target image $x^\star$ from an observation y degraded by a non-invertible or ill-conditioned linear operator $\mathcal{A}$ and an additive noise. These problems have traditionally been addressed in a variational framework where an estimate of $x^\star$ is obtained by minimizing a cost function composed of a penalty, favoring structural properties (e.g. sparsity or patch redundancy), and a data fidelity term.

Significant advances in deep learning have given rise to approaches where image restoration is addressed as a *discriminative learning problem*, *i.e.*, where the mapping from y to $x^\star$ is learned end-to-end [1], [2]. The empirical performance of these methods is good, but their lack of interpretability, as well as the need to re-train the network as soon as the degradation model changes, represent important limitations. It is also worth noting that recent works [3] have highlighted the instability of this type of approaches, especially when the forward operator $\mathcal{A}$ is not used explicitly by the neural network.

Recently, variational approaches that use a deep neural network to define the regularizer [4]–[7] have emerged, thus decoupling the learning phase from the degradation model. The regularizer can be defined explicitly or implicitly, the latter corresponding to the family of so-called plug-and-play methods [8].

When reference images or patches are available, recent works [9], [10] have proposed to use the Wasserstein distance between the empirical distribution of the features of the reference and those of the sought image as a regularizer. The authors of [9] used one-dimensional features in order to leverage the corresponding closed-form solution of OT, while OT in the patch-space is considered in [10]. The idea of using the Wasserstein distance in order to statistically constrain the features of the solution was originally proposed for texture synthesis in [11]–[14]. For the particular super-resolution problem with the presence of a reference image, that is, an image whose patch distribution is similar to that of the sought image, the aforementioned work [10] produces state-of-the-art results, improving upon variational methods that use learned deep regularizers.

The approach we adopt in this article lies in the framework of variational methods that use a generative model or uncorrupted reference examples to define the regularization term. In order to guarantee the independence of the proposed approach with respect to (wrt) the dimension of the images, the modeling is done at the patch scale. The objective is to design a penalty that favors images whose patches are consistent with a learned generative model or with a set of non-corrupted patches available at the time of the restoration.

To do so, we propose to take advantage of recent advances in numerical OT [15]–[18] and propose a penalty based on the cost of an OT between the empirical patch distribution of the restored image and that of a generative model or a reference image available at the time of restoration. It consists in finding the optimal match between the two patch distributions by minimizing the cost of transporting the mass of each patch from one distribution to the other, given a chosen cost c. Thus, unlike classical regularization approaches, our formulation explicitly controls the deviation from the prior statistical model.

With the help of numerical illustrations, we show the necessity to adopt a non-symmetric form of imbalance, subsequently called semi-unbalanced OT. We give details to calculate and differentiate this formulation. Finally, we show how the increased robustness gained with our formulation translates into improved performance for the super-resolution problem considered in [10].

II. A Penalty Based on Semi-Unbalanced Optimal Transport

In this section we adopt the notations and definitions of [16], [17]. For $N \in \mathbb{N}$, let $[\![N]\!] = \{1, \dots, N\}$. We consider discrete

Acknowledgments: The authors acknowledge the support of the projects PostProdLEAP (ANR-19-CE23-0027-01) and REGETO (AAP GdR ISIS 2021).

probability measures with support in $\mathbb{R}^n$. Let $\alpha = \sum_{i=1}^N a_i \delta_{x_i}$ and $\beta = \sum_{j=1}^M b_j \delta_{y_j}$ denote two such measures, with $x_i, y_j \in \mathbb{R}^n$. Let $a = (a_i)_{i \in [\![N]\!]} \in \Sigma_N$ and $b = (b_j)_{j \in [\![M]\!]} \in \Sigma_M$ denote the associated probability vectors, Σ_p denoting the probability simplex in $\mathbb{R}^p$. Subsequently, the distributions α and β will respectively correspond the discrete uniform distribution of patches extracted from the sought image and from the reference image (or the generative model). The product measure of α and β will be denoted $\alpha \otimes \beta$. We consider the matrix $C \in \mathbb{R}^{N \times M}$ such that $c_{i,j} = c(x_i, y_j) = \|x_i - y_j\|_2^2$. For any matrix $\pi \in \mathbb{R}^{N \times M}$ let $\pi_1 = \pi \mathbf{1}_M$ and $\pi_2 = \pi^T \mathbf{1}_N$. Let ι_E be the function which is 0 on E and $+\infty$ on E^c, and let $\iota_v = \iota_{\{v\}}$ if the set contains only one element.

Definition 1 (Optimal transport): The OT between two discrete probability measures α and β is defined by

$$\mathrm{OT}(\alpha, \beta) = \min_{\pi \in \mathbb{R}_+^{N \times M}} \langle C, \pi \rangle + \iota_\alpha(\pi_1) + \iota_\beta(\pi_2). \tag{1}$$

Finding π in (1) is a constrained linear programming problem. The authors of [16], [17] proposed to make the problem strictly convex by regularizing it by the KL divergence as in Definition 2.

Definition 2 (Regularized optimal transport [16], [17]): Let $\varepsilon > 0$ be fixed. The regularized OT between two discrete probability measures α and β is defined by

$$\begin{aligned} \mathrm{OT}_\varepsilon(\alpha, \beta) = \min_{\pi \in \mathbb{R}_+^{N \times M}} & \langle C, \pi \rangle + \varepsilon \, \mathrm{KL}(\pi | \alpha \otimes \beta) \\ & + \iota_\alpha(\pi_1) + \iota_\beta(\pi_2), \end{aligned} \tag{2}$$

where $\mathrm{KL}(\alpha|\beta) = \left\langle a, \log\left(\frac{a}{b}\right) \right\rangle$ is the Kullback-Leibler divergence.

Problem (2) is associated with a dual concave maximization problem that can be solved by a block coordinate ascent, leading to an algorithm equivalent to that of Sinkhorn in logarithmic variables [16], [17, Prop. 4.4]. The gradients of $\mathrm{OT}_\varepsilon(\alpha, \beta)$ wrt a, b, $(x_i)_{i \in [\![N]\!]}$ and $(y_j)_{j \in [\![M]\!]}$ can be obtained by the envelop theorem [19, p. 124].

As explained in [16], the entropy-regularized balanced OT is biased. Rigollet and Weed [20] show that the OT_ε-projection of an empirical measure on a class of measures satisfying a so-called closure under dominance hypothesis, corresponds to a maximum likelihood estimator in a Gaussian deconvolution model whose standard deviation is precisely ε. Nevertheless, for small values of ε, OT_ε allows to compare distributions and is computable and differentiable with the Sinkhorn algorithm. These arguments make it a good candidate to define a penalty in variational image restoration methods.

Let us consider the case where we have a generative model of patches, represented by a probability measure β. This model can either be learned upstream on a set of non-degraded patches [21], or simply correspond to the empirical measure of a set of reference patches, similar to those we wish to restore. A variational approach for recovering $x^\star$ from y can be formulated as the following problem:

$$\min_x \frac{\lambda}{2} \|\mathcal{A}x - y\|^2 + \mathbb{E}_{\tilde{\beta}_M \sim \beta}[\mathrm{OT}_\varepsilon(\alpha_x, \tilde{\beta}_M)], \tag{3}$$

where $\alpha_x = \frac{1}{N} \sum_{n=1}^N \delta_{P_i x}$ is the empirical distribution of the patches $(P_i x)_{i \in [\![N]\!]}$ extracted from x, $\tilde{\beta}_M$ is the empirical distribution associated with a random sample of M patches according to β, and $\lambda, \varepsilon > 0$.

Let us illustrate the solution of this problem with a denoising example, *i.e.*, with $\mathcal{A}$ the identity operator, on an isotropic 100×100 Gaussian texture [22], the patches of which are distributed according to some Gaussian distribution $\mathcal{N}_1$. To account for the possible mismatch between the generative model β and the patch distribution of the clean image, we choose β as the mixture density $0.8\mathcal{N}_1 + 0.2\mathcal{N}_2$, with $\mathcal{N}_2$ a Gaussian distribution that generates patches with a horizontal edge. Problem (3) is minimized by stochastic gradient descent in which a sample $\tilde{\beta}_M$ from β of size $M = 24000$ is drawn at each iteration. We obtain the result labeled as OT_ε in Figure 1. We notice the creation of artifacts due to outlier patches coming from $\mathcal{N}_2$, caused by the strong constraint $\iota_{\tilde{\beta}_M}(\pi_2)$. In order to deal with this issue, we introduce a semi-unbalanced formulation for OT which is less sensitive to outliers present in the model distribution β.

Definition 3 (Regularized semi-unbalanced OT): Let $\varepsilon > 0$ and $\rho > 0$ be fixed. We define the regularized semi-unbalanced OT between the measures α and β by

$$\begin{aligned} \mathrm{OT}_{\varepsilon,\rho}(\alpha, \beta) = \min_{\pi \in \mathbb{R}_+^{N \times M}} & \langle C, \pi \rangle + \varepsilon \mathrm{KL}(\pi | \alpha \otimes \beta) \\ & + \iota_\alpha(\pi_1) + \rho \mathrm{KL}(\pi_2 | \beta). \end{aligned} \tag{4}$$

This new transport is inspired by the unbalanced version of Chizat et al. [23] and Séjourné et al. [18], denoted $\mathrm{OT}_{\varepsilon,\rho,\rho}$ in the sequel, where the two constraints $\iota_\alpha(\pi_1)$ and $\iota_\beta(\pi_2)$ are respectively replaced by $\rho \, \mathrm{KL}(\pi_1|\alpha)$ and $\rho \, \mathrm{KL}(\pi_2|\beta)$. Note that both $\mathrm{OT}_{\varepsilon,\rho}$ and $\mathrm{OT}_{\varepsilon,\rho,\rho}$ converge to the balanced regularized transport OT_ε as $\rho \to \infty$.

In Figure 1 we compare the denoising results obtained when replacing OT_ε by $\mathrm{OT}_{\varepsilon,\rho,\rho}$, resp. $\mathrm{OT}_{\varepsilon,\rho}$, in Problem (3). While, as expected, the relaxation $\mathrm{KL}(\pi_2|\tilde{\beta}_M)$ of $\iota_{\tilde{\beta}_M}(\pi_2)$ leads to decrease the sensitivity of the result wrt the outlier patches sampled from $\mathcal{N}_2$, the relaxation $\mathrm{KL}(\pi_1|\alpha_x)$ of $\iota_{\alpha_x}(\pi_1)$ produces an image where some areas are not restored. In comparison, outlier patches do not affect our proposed semi-unbalanced OT result and the denoising is spatially homogeneous, resulting in the best PSNR. To summarize, the semi-unbalanced formulation of OT proposed in (4) is a robust version of OT_ε, which allows the data distribution to only match a part of the reference distribution, the proportion of matching data being tuned by ρ. It is particularly appropriate for image restoration with reference distribution, since, in this setting, the proportion of patches in the reference and in the ground-truth are expected to differ.

We detail below how a dual formulation allows us to compute the functional $\mathrm{OT}_{\varepsilon,\rho}(\alpha, \beta)$ and its gradient [16], [18], a key practical point to solve (3) numerically.

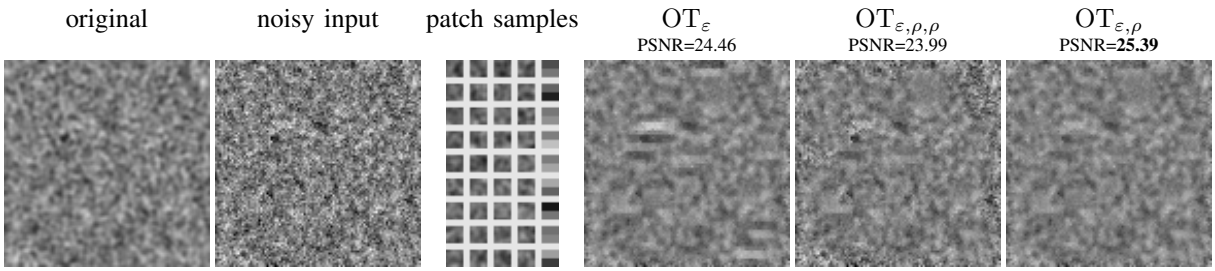

Fig. 1. Denoising of a 100×100 Gaussian texture with an imperfect patch model. From left to right: original image, noisy input, samples from the corrupted generative model, restoration by $\mathrm{OT}_{\varepsilon}$, $\mathrm{OT}_{\varepsilon,\rho,\rho}$ and $\mathrm{OT}_{\varepsilon,\rho}$ ($\lambda = 0.0192, \varepsilon = 2\times10^{-4}, \rho = 0.1$, patch size = 8, stride = 2). The generative model simulates patches suitable for image restoration in 80% of cases, and patches with an horizontal edge unsuitable for restoration in 20% of the cases.

Proposition 1 (Dual formulation): For $\varepsilon > 0$ fixed,

$$\begin{aligned}\mathrm{OT}_{\varepsilon,\rho}(\alpha,\beta) = &\max_{(f,g)\in\mathbb{R}^N\times\mathbb{R}^M} \langle a, f\rangle - \langle b, \Phi^*(-g)\rangle \\ &- \varepsilon\left\langle a\otimes b, \exp\left(\frac{f\oplus g - C}{\varepsilon}\right) - 1\right\rangle\end{aligned} \tag{5}$$

with $\Phi^*(q) = \rho(\exp(\frac{q}{\rho}) - 1)$, applied to each component.

Problem (5) is a concave maximization problem which can be solved by alternate maximization wrt f and g, as specified by the following theorem.

Theorem 1 (Sinkhorn's algorithm [18]): Starting from any $f^0 \in \mathbb{R}^N$, the following algorithm converges to a solution of Problem (5):

$$g_j^{t+1} = -\tfrac{\varepsilon}{1+\frac{\varepsilon}{\rho}} \log\left(\textstyle\sum_{i=1}^N a_i \exp\left(\frac{f_i^t - c_{i,j}}{\varepsilon}\right)\right),\ j\in[\![M]\!],$$
$$f_i^{t+1} = -\varepsilon \log\left(\textstyle\sum_{j=1}^M b_j \exp\left(\frac{g_j^{t+1} - c_{i,j}}{\varepsilon}\right)\right),\ i\in[\![N]\!].$$

This sequence of vectors (f^t, g^t) satisfies

$$F(f^t, g^t) = \langle a, f^t\rangle - \langle b, \Phi^*(-g^t)\rangle \tag{6}$$

where $F(f,g)$ is the function to maximise in Problem (5).

Alternatively, the solution vectors f and g can be computed by iterating a symmetric fixed-point method [19]: starting from any $(\tilde f^0, \tilde g^0) \in \mathbb{R}^N \times \mathbb{R}^M$, the following algorithm also converges to a solution of (5):

$$\tilde g_j^{t+1} = \tfrac{1}{2}\left(\tilde g_j^t - \tfrac{\varepsilon}{1+\frac{\varepsilon}{\rho}} \log\left(\textstyle\sum_{i=1}^N a_i \exp\left(\frac{\tilde f_i^t - c_{i,j}}{\varepsilon}\right)\right)\right),$$
$$\tilde f_i^{t+1} = \tfrac{1}{2}\left(\tilde f_i^t - \varepsilon \log\left(\textstyle\sum_{j=1}^M b_j \exp\left(\frac{\tilde g_j^t - c_{i,j}}{\varepsilon}\right)\right)\right),$$

with $(i,j) \in [\![N]\!] \times [\![M]\!]$.

In our experiments, we use the latter symmetric scheme with an additional last non-symmetric iteration in order to guarantee Eq. (6) and (7) below. We use the initializations $(f^0, g^0) = (\mathbf{0}_N, \mathbf{0}_M)$ and a scaling of the regularisation ε as in [19, p.120].

In order to use gradient-based optimization, one should be able to differentiate $\mathrm{OT}_{\varepsilon,\rho}(\alpha,\beta)$ wrt the (discrete) support of α, that is, wrt the coordinates x_i. This is possible by assuming that the convergence of Sinkhorn's algorithm is reached [16]: if $(f^\star, g^\star)$ is a solution of Problem (5), then

$$\partial_{x_i} \mathrm{OT}_{\varepsilon,\rho}\left(\textstyle\sum_{i=1}^N a_i\delta_{x_i}, \beta\right) = a_i \nabla\varphi(x_i) \tag{7}$$

TABLE I

AVERAGE PSNR AND LPIPS OBTAINED ON TWO DATASETS OF 18 IMAGE PAIRS FROM THE MVTEC IMAGE DATABASE [24], [25]. REFERENCE IMAGES FROM THE FIRST DATASET ARE DEFECT FREE WHILE THOSE IN THE SECOND DATASET CONTAIN DEFECTS LIKE IN FIGURE 2. FROM TOP TO BOTTOM: RESTORATION BY WPP [10], $\mathrm{OT}_{\varepsilon}$, $\mathrm{OT}_{\varepsilon,\rho}$.

	Defect-free		With defects	
	PSNR ↑	LPIPS ↓	PSNR ↑	LPIPS ↓
WPP [10]	30.08	0.095	30.13	0.099
$\mathrm{OT}_{\varepsilon}$	30.12	**0.086**	29.38	0.113
$\mathrm{OT}_{\varepsilon,\rho}$ (Ours)	**30.54**	0.091	**30.93**	**0.092**

where $\varphi : \mathbb{R}^n \to \mathbb{R}$ has the expression

$$\varphi(x) = -\varepsilon \log\left(\textstyle\sum_{j=1}^M b_j \exp\left(\frac{g_j^\star - c(x,y_j)}{\varepsilon}\right)\right). \tag{8}$$

III. SUPER-RESOLUTION WITH A REFERENCE IMAGE

In [10], super-resolution with a reference image is described as the restoration of an image $x^\star$ given its low resolution (LR) version y^{LR} and a reference image x_{ref}, the patch distribution of which is assumed to be similar to the one of the ground-truth $x^\star$. This setting is relevant when working on specific classes of images, *e.g.* textures or material images. The forward model $y^{\mathrm{LR}} = SHx^\star + \eta$, where H is a convolution operator, S a dowsampling operator, and η an additive Gaussian noise, is assumed to be known. The authors of [10] estimate the ground truth HR image $x^\star$ by solving the minimization problem

$$\min_x \frac{\lambda}{2}\|SHx - y^{\mathrm{LR}}\|^2 + \frac{1}{2}\sum_{\ell=0}^{L} \mathrm{OT}_0(\alpha_{x^\ell}, \beta_{x_{\mathrm{ref}}{}^\ell}), \tag{9}$$

where OT_0 refers to non-regularized OT, $\alpha_{x^\ell} = \frac{1}{N^\ell}\sum_{i=1}^{N^\ell}\delta_{P_i x^\ell}$ and $\beta_{x_{\mathrm{ref}}{}^\ell} = \frac{1}{M^\ell}\sum_{j=1}^{M^\ell}\delta_{P_j x_{\mathrm{ref}}^\ell}$ are respectively the empirical distributions associated with the patches of $x^\ell = A^\ell x$ and $x_{\mathrm{ref}}^\ell = A^\ell x_{\mathrm{ref}}$, A being a convolution with a 4×4 Gaussian blur kernel with a standard deviation of 1 followed by a $\times2$ downsampling. Thus, the regularization term in Problem (9), coined Wasserstein Patch Prior (WPP) [10], favors images x whose patch distributions at different scales $\ell = 0,\dots,L$ are close to those of the reference image x_{ref}. We chose $L = 1$.

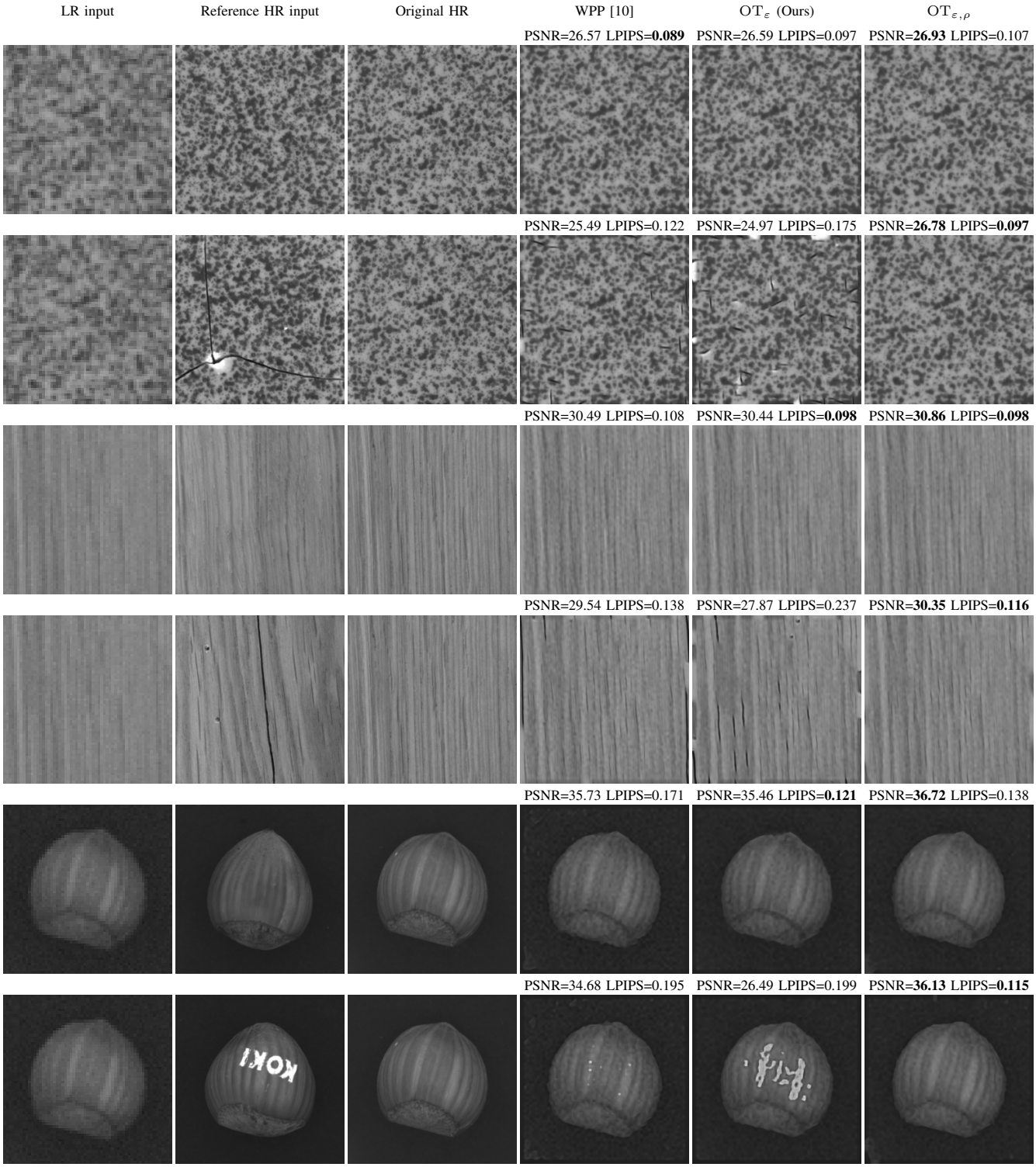

Fig. 2. Application to super-resolution with a reference image. From left to right: LR input, reference HR input, original HR image, restoration by WPP [10], $\mathrm{OT}_{\varepsilon}$ and $\mathrm{OT}_{\varepsilon,\rho}$.

In practice, to compensate for the differences between α_x and $\beta_{x_{\mathrm{ref}}}$, the authors of [10] propose to apply the regularization term of Problem (9) to a padded version $\mathcal{P}x$ of x, while the fidelity term is applied directly to x. This enables to alleviate the rigidity of OT by allowing outlier patches of x_{ref} to be aggregated into the artificial bounds of $\mathcal{P}x$.

We propose to solve the same problem by replacing $\mathrm{OT}_0(\mathcal{P}x)$ by $\mathrm{OT}_{\varepsilon,\rho}(x)$, that is, we apply our robust semi-unbalanced formulation of OT directly to x, instead of artificially padding x and then using balanced OT. We numerically

solve our minimization problem by gradient descent with the Adam optimizer [26] with $\mathrm{lr} = 0.01$.

We consider for our experiments images from the MVTec image database [24], [25] tranformed to grayscale and resized to 256×256 size. This is a dataset that includes a training set of defect-free images and a test set of images with different types of defects. We selected 18 HR images from the MVTec image database. Following Hertrich et al. [10], for each HR image, we simulated its corresponding LR version by convolving it with a 16×16 Gaussian kernel with a standard deviation of 2, then we applied a subsampling $\times 4$ and added a Gaussian noise $\eta \sim \mathcal{N}(0, 0.01^2)$. For each of the LR images, we chose two reference HR images: one with defects and one without. We created two datasets, the first contains 18 pairs of LR images and reference images without defects, while the reference images in the second dataset all have some defects. We restored each of the LR images in both datasets using their associated reference images with $\varepsilon = 4 \times 10^{-4}$, $\lambda = 0.006$, $\rho = 0.01$, patch size $= 6$. All patches are extracted from the image to be restored and 10000 patches are randomly drawn from the reference image. We calculated 6 pixels away from the edges the average PSNR and the average LPIPS [27] per dataset and grouped them in Table I. We show in Figure 2 some of the results from Table I. For each image in Figure 2, the first (resp. second) row restores the LR image with a reference HR image without defects (resp. with defects). The results in Figure 2 and Table I confirm that the increased robustness obtained by replacing $\mathrm{OT}_0(\mathcal{P}x)$ with $\mathrm{OT}_{\varepsilon,\rho}(x)$ leads to better PSNR and LPIPS scores, especially when the reference image is not ideally close to the original image.

IV. Conclusion

We have presented a new methodological framework for image restoration, based on an asymmetrically unbalanced notion of OT. Through two sets of experiments, we have shown that semi-unbalanced regularized transport is a robust alternative to the usual balanced and unbalanced formulations for patch-based image restoration. While we have evaluated our method on a variational SR problem with a reference image, future work will focus on other image restoration problems that can take advantage of our semi-unbalanced OT penalty.

References

[1] A. Lucas, M. Iliadis, R. Molina, and A.K. Katsaggelos, "Using deep neural networks for inverse problems in imaging: beyond analytical methods," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 20–36, 2018.

[2] V. Monga, Y. Li, and Y.C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, 2021.

[3] N.M. Gottschling, V. Antun, B. Adcock, and A.C. Hansen, "The troublesome kernel: why deep learning for inverse problems is typically unstable," *arXiv preprint:2001.01258*, 2020.

[4] A. Bora, A. Jalal, E. Price, and A.G Dimakis, "Compressed sensing using generative models," in *ICML*, 2017, pp. 537–546.

[5] M. González, A. Almansa, and P. Tan, "Solving inverse problems by joint posterior maximization with autoencoding prior," *arXiv preprint:2103.01648*, 2021.

[6] R. Laumont, V. De Bortoli, A. Almansa, J. Delon, A. Durmus, and M. Pereyra, "On maximum-a-posteriori estimation with plug & play priors and stochastic gradient descent," *arXiv preprint:2201.06133*, 2022.

[7] J. Prost, A. Houdard, A. Almansa, and N. Papadakis, "Learning local regularization for variational image restoration," *arXiv preprint:2102.06155*, 2021.

[8] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 945–948.

[9] M. El Gheche, J.-F. Aujol, Y. Berthoumieu, and C.-A. Deledalle, "Texture reconstruction guided by a high-resolution patch," *IEEE TIP*, vol. 26, no. 2, pp. 549–560, 2016.

[10] J. Hertrich, A. Houdard, and C. Redenbach, "Wasserstein patch prior for image superresolution," *IEEE Transactions on Computational Imaging*, vol. 8, pp. 693–704, 2022.

[11] G. Tartavel, G. Peyré, and Y. Gousseau, "Wasserstein loss for image synthesis and restoration," *SIAM Journal on Imaging Sciences*, vol. 9, no. 4, pp. 1726–1755, 2016.

[12] J. Gutierrez, B. Galerne, J. Rabin, and T. Hurtut, "Optimal patch assignment for statistically constrained texture synthesis," in *SSVM 2017*, 2017, pp. 172–183.

[13] B. Galerne, A. Leclaire, and J. Rabin, "A texture synthesis model based on semi-discrete optimal transport in patch space," *SIAM Journal on Imaging Sciences*, vol. 11, no. 4, pp. 2456–2493, 2018.

[14] A. Houdard, A. Leclaire, N. Papadakis, and J. Rabin, "A generative model for texture synthesis based on optimal transport between feature distributions," *Journal of Mathematical Imaging and Vision*, pp. 1–25, 2022.

[15] B. Charlier, J. Feydy, J. Glaunès, F.-D. Collin, and G. Durif, "Kernel operations on the GPU, with autodiff, without memory overflows," *Journal of Machine Learning Research*, vol. 22, no. 74, pp. 1–6, 2021.

[16] J. Feydy, T. Séjourné, F. Vialard, S. Amari, A. Trouvé, and G. Peyré, "Interpolating between optimal transport and MMD using Sinkhorn divergences," *AiSTATS*, pp. 2681–2690, 2019.

[17] G. Peyré and M. Cuturi, "Computational optimal transport: With applications to data science," *Foundations and Trends in Machine Learning*, vol. 11, no. 5-6, pp. 355–607, 2019.

[18] T. Séjourné, J. Feydy, F. Vialard, A. Trouvé, and G. Peyré, "Sinkhorn divergences for unbalanced optimal transport," *arXiv preprint:1910.12958*, 2019.

[19] J. Feydy, *Geometric data analysis, beyond convolutions*, Ph.D. thesis, University of Paris-Saclay, 2020.

[20] P. Rigollet and J. Weed, "Entropic optimal transport is maximum-likelihood deconvolution," *Comptes Rendus Mathématique*, vol. 356, no. 11-12, pp. 1228–1235, 2018.

[21] Daniel Zoran and Yair Weiss, "From learning models of natural image patches to whole image restoration," in *2011 International Conference on Computer Vision*, 2011, pp. 479–486.

[22] B. Galerne, Y. Gousseau, and J.-M. Morel, "Random phase textures: Theory and synthesis," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 257 – 267, 2011.

[23] Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard, "Scaling algorithms for unbalanced optimal transport problems," *Mathematics of Computation*, vol. 87, no. 314, pp. 2563–2609, 2018.

[24] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "MVTec AD — a comprehensive real-world dataset for unsupervised anomaly detection," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9584–9592.

[25] P. Bergmann, K. Batzner, M. Fauser, D. Sattlegger, and C. Steger, "The MVTec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection.," in *International Journal of Computer Vision*, 2021, vol. 129(4), pp. 1038–1059.

[26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint:1412.6980*, 2014.

[27] R. Zhang, P. Isola, A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE Conference on computer vision and pattern recognition*, 2018, pp. 586–595.