



HAL
open science

SEMI-UNBALANCED REGULARIZED OPTIMAL TRANSPORT FOR IMAGE RESTORATION

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

► **To cite this version:**

Simon Mignon, Bruno Galerne, Moncef Hidane, Cécile Louchet, Julien Mille. SEMI-UNBALANCED REGULARIZED OPTIMAL TRANSPORT FOR IMAGE RESTORATION. 2022. hal-03896487v1

HAL Id: hal-03896487

<https://hal.science/hal-03896487v1>

Preprint submitted on 13 Dec 2022 (v1), last revised 7 Mar 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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.

1. 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 u from an observation y degraded by a non-invertible or ill-conditioned linear operator A and an additive noise. These problems have traditionally been addressed in a variational framework where an estimate of u 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 restoration is addressed as a *discriminative learning problem*, i.e., where the mapping from y to u 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 A is not used explicitly by the neural network.

Recently, variational approaches that use a deep neural network to define the regularizer [4, 5, 6, 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 has also been adopted for texture synthesis in [11, 12, 13, 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.

1.1. Our Approach

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, 16, 17, 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 available at the time of restoration. 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].

2. A PENALTY BASED ON UNBALANCED OPTIMAL TRANSPORT

In this section we adopt the notations and definitions of [16, 17]. For $N \in \mathbb{N}$, let $\llbracket N \rrbracket = \{1, \dots, N\}$. We consider discrete 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 $a = (a_i)_{i \in \llbracket N \rrbracket} \in \Sigma_N$ and $b = (b_j)_{j \in \llbracket M \rrbracket} \in \Sigma_M$ denote the associated probability vectors, Σ_p being the probability simplex in \mathbb{R}^p . 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 (Regularized optimal transport [16, 17]). *Let $\varepsilon > 0$ be fixed. The regularized OT between two discrete probability measures α and β is defined by*

$$\text{OT}_\varepsilon(\alpha, \beta) = \min_{\pi \in \mathbb{R}_+^{N \times M}} \langle C, \pi \rangle + \varepsilon \text{KL}(\pi | \alpha \otimes \beta) + \iota_\alpha(\pi_1) + \iota_\beta(\pi_2), \quad (1)$$

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

Problem (1) 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 [17, Prop. 4.4], [16]. The gradients of $\text{OT}_\varepsilon(\alpha, \beta)$ wrt $a, b, (x_i)_{i \in \llbracket N \rrbracket}$ and $(y_j)_{j \in \llbracket M \rrbracket}$ can be obtained by implicit differentiation [19, p. 124].

As explained in [16], the entropy-regularized balanced OT is biased. Rigollet et al. [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, 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 u from y can be formulated as the following problem:

$$\min_x \frac{\lambda}{2} \|Ax - y\|^2 + \mathbb{E}_{\tilde{\beta}_M \sim \beta} [\text{OT}_\varepsilon(\alpha_x, \tilde{\beta}_M)], \quad (2)$$

where $\alpha_x = \frac{1}{N} \sum_{n=1}^N \delta_{P_n x}$ is the empirical distribution of the patches $(P_n x)_{n \in \llbracket N \rrbracket}$ 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 $A = \mathcal{I}$, on an isotropic Gaussian texture [21], 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 (2) 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 that is targeted towards applications in inverse image problems. The aim here is to decrease the sensitivity of the OT cost to outliers present in the model distribution β .

Definition 2 (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*

$$\text{OT}_{\varepsilon, \rho}(\alpha, \beta) = \min_{\pi \in \mathbb{R}_+^{N \times M}} \langle C, \pi \rangle + \varepsilon \text{KL}(\pi, \alpha \otimes \beta) + \iota_\alpha(\pi_1) + \rho \text{KL}(\pi_2, \beta). \quad (3)$$

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

In Figure 1 we compare the denoising results obtained when replacing OT_ε by $\text{OT}_{\varepsilon, \rho, \rho}$ and $\text{OT}_{\varepsilon, \rho}$ in Problem (2). While, as expected, the relaxation $\text{KL}(\tilde{\beta}_M, \pi_2)$ 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 $\text{KL}(\alpha_x, \pi_1)$ of $\iota_{\alpha_x}(\pi_1)$ produces an image where some areas have not been 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 Equation (3) is a robust version of OT and is appropriate for applications in inverse imaging problems.

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

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

$$\text{OT}_{\varepsilon, \rho}(\alpha, \beta) = \max_{(f, g) \in \mathbb{R}^N \times \mathbb{R}^M} \langle a, f \rangle - \langle b, \Phi^*(-g) \rangle - \varepsilon \left\langle a \otimes b, \exp\left(\frac{f \oplus g - C}{\varepsilon}\right) - 1 \right\rangle \quad (4)$$

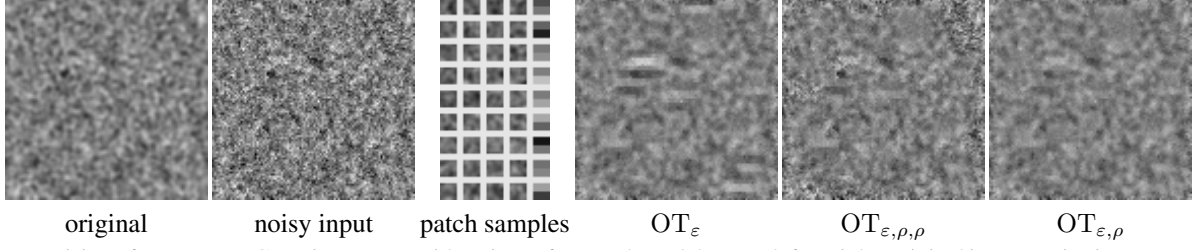


Fig. 1. Denoising of a 100×100 Gaussian texture with an imperfect patch model. From left to right: original image, noisy input, samples from the corrupted generative model, restoration by OT_ε , $\text{OT}_{\varepsilon, \rho, \rho}$ and $\text{OT}_{\varepsilon, \rho}$ ($\lambda = 0.0192$, $\varepsilon = 10^{-4}$, $\rho = 0.1$, patch size = 8). 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. While the optimal balanced OT creates aberrant patches in the restored image and the unbalanced OT does not restore the entire image, the semi-unbalanced OT eliminates the artifacts and restores the entire image. The PSNRs are respectively 24.46, 23.99 and 25.39.

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

Problem (4) 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 (4):*

$$g_j^{t+1} = -\frac{\varepsilon}{1+\varepsilon} \log \left(\sum_{i=1}^N a_i \exp \left(\frac{f_i^t - c_{i,j}}{\varepsilon} \right) \right), \quad j \in \llbracket M \rrbracket,$$

$$f_i^{t+1} = -\varepsilon \log \left(\sum_{j=1}^M b_j \exp \left(\frac{g_j^{t+1} - c_{i,j}}{\varepsilon} \right) \right), \quad i \in \llbracket N \rrbracket.$$

Alternatively, the solution vectors f and g can be computed by iterating a symmetric fixed-point method [19]: Starting from any $(f^0, g^0) \in \mathbb{R}^N \times \mathbb{R}^M$, the following algorithm also converges to a solution of (4):

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

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

with $(i, j) \in \llbracket N \rrbracket \times \llbracket M \rrbracket$.

The latter symmetric iterations, with initialization $f^0 = \mathbf{0}_N$ and $g^0 = \mathbf{0}_M$, are the ones we use in all our numerical experiments.

Proposition 2. *Let $F(f, g)$ be the function to be maximized in Problem (4). The sequence of vectors (f^t, g^t) defined by the Sinkhorn algorithm satisfies:*

$$F(f^t, g^t) = \langle a, f^t \rangle - \langle b, \Phi^*(-g^t) \rangle. \quad (5)$$

In order to use gradient-based optimization, one should be able to differentiate $\text{OT}_{\varepsilon, \rho}(\alpha, \beta)$ 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^*, g^*) is a solution of Problem (4), then

$$\partial_{x_i} \text{OT}_\varepsilon \left(\sum_{i=1}^N a_i \delta_{x_i}, \beta \right) = a_i \nabla \varphi(x_i) \quad (6)$$

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

$$\varphi(x) = -\varepsilon \log \left(\sum_{j=1}^M b_j \exp \left(\frac{g_j^* - c(x, y_j)}{\varepsilon} \right) \right). \quad (7)$$

3. SUPER-RESOLUTION WITH A REFERENCE IMAGE

Super-resolution with a reference image is a particular problem of image restoration. The authors of [10] describe it as the restoration of an image x given its low resolution (LR) version y^{LR} and a reference image \tilde{x} , the patch distribution of which is assumed to be similar to the one of the ground-truth x . This setting is relevant when working on specific classes of images, e.g. textures or material images. The forward model $y^{\text{LR}} = SHx + \eta$, where H is a convolution operator, S a subsampling operator, and η an additive Gaussian noise, is assumed to be known.

The authors of [10] estimate the ground truth HR image x by solving the minimization problem

$$\min_x \frac{\lambda}{2} \|SHx - y^{\text{LR}}\|^2 + \sum_{\ell=0}^L \text{OT}_0(\alpha_{x^\ell}, \beta_{\tilde{x}^\ell}), \quad (8)$$

where OT_0 refers to non-regularized OT, $\alpha_{x^\ell} = \frac{1}{N} \sum_{i=1}^N \delta_{P_i x^\ell}$ and $\beta_{\tilde{x}^\ell} = \frac{1}{M} \sum_{j=1}^M \delta_{P_j \tilde{x}^\ell}$ are respectively the empirical distributions associated with the patches of $x^\ell = A^\ell x$ and $\tilde{x}^\ell = A^\ell \tilde{x}$, A being a $\times 2$ downsampling operator. Thus, the regularization term in Problem (8) favors images x whose patch distributions at different scales $\ell = 0, \dots, L$ are close to those of the reference image \tilde{x} . It is chosen $L = 1$.

In practice, to compensate for the differences between α_x and $\beta_{\tilde{x}}$, the authors of [10] propose to apply the regularization term of Problem (8) 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 \tilde{x} to be aggregated into the artificial bounds of $\mathcal{P}x$.

We propose to solve the same problem by replacing $\text{OT}_0(\mathcal{P}x)$ by $\text{OT}_{\varepsilon, \rho}(x)$, 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 [22]. Results reported in Figure 2 confirm that the increased robustness obtained when replac-

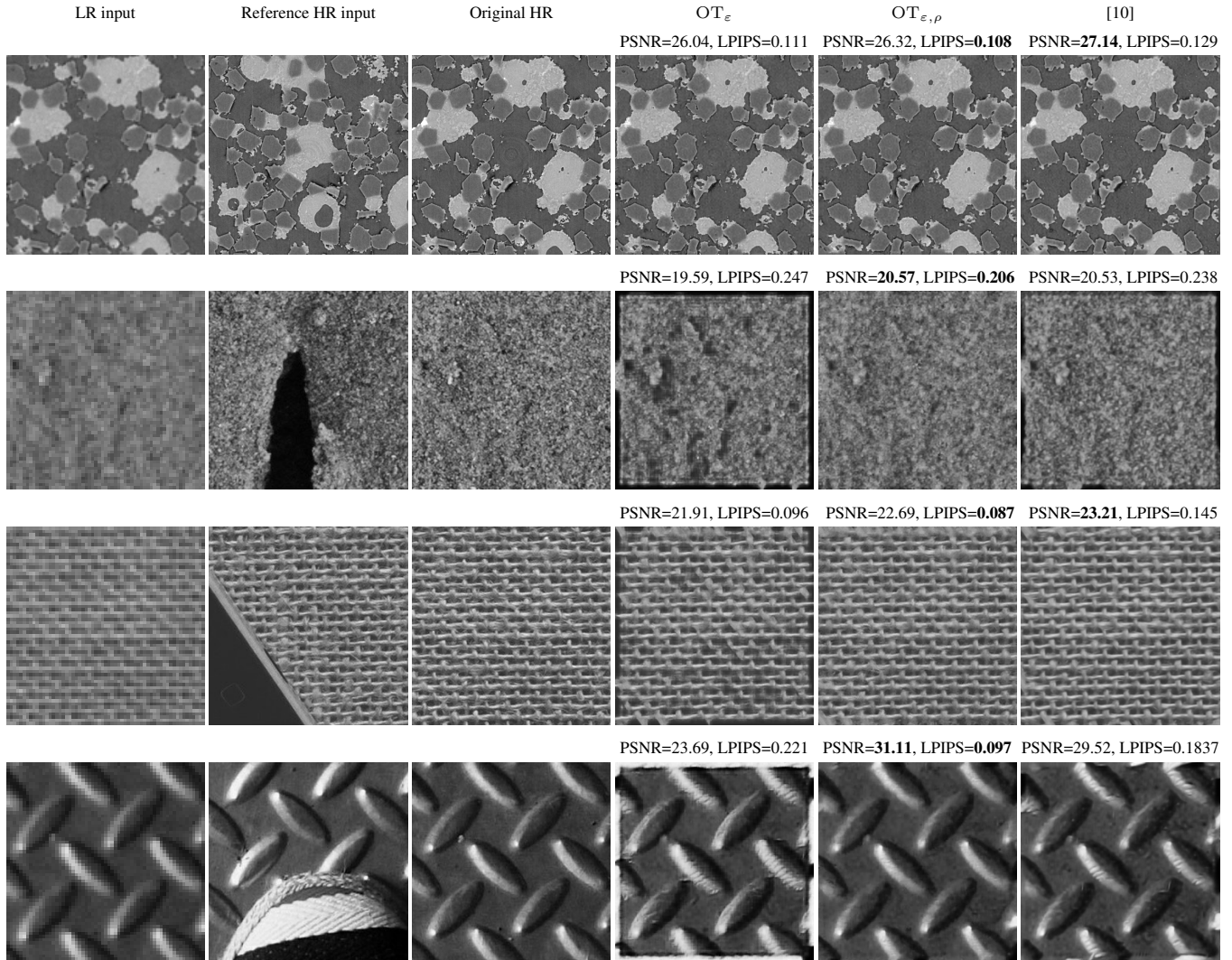


Fig. 2. Application to super-resolution with a reference image. From left to right: LR input (convolution with a 16×16 Gaussian blur kernel with standard deviation 2 and stride 4), reference HR input, original HR image, restoration by OT_ε , $OT_{\varepsilon,\rho}$ and [10]. Results obtained with $\varepsilon = 4 \times 10^{-4}$, $\lambda = 0.033$, $\rho = 0.8$, patch size = 6 for the first row and with $\varepsilon = 4 \times 10^{-4}$, $\lambda = 0.006$, $\rho = 0.01$, patch size = 6 for rows 2, 3, 4. The downsampling operator A is a convolution with a 4×4 Gaussian blur kernel with standard deviation 1 and stride 2. With $OT_{\varepsilon,\rho}$, we gain in PSNR, LPIPS and visually on OT_ε and [10]. For the last three 256×256 images, the restoration time is 110s for our method against 140s for theirs. For the first 600×600 image, our restoration time passes to 30 minutes against 7 minutes for [10]. PSNRs and LPIPS are calculated 6 pixels away from the edges.

ing $OT_0(\mathcal{P}x)$ by $OT_{\varepsilon,\rho}(x)$ yields improved LPIPS¹ [23] and competitive PSNR scores, especially when the reference image is not ideally close to the original image.

4. CONCLUSION

We have presented a new methodological framework for image restoration, based on an asymmetrically unbalanced notion of OT. Through two experiments, we have shown that

¹<https://github.com/richzhang/PerceptualSimilarity>, version 0.1.4

semi-unbalanced regularized transport is a robust alternative to the usual balanced and unbalanced formulations for patch-based image restoration. This work opens the door to several perspectives such as its extension to a semi-discrete setting, allowing to incorporate deep generative models into image restoration problems.

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

5. 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. Venkatakrisnan, 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] 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.
- [22] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint:1412.6980*, 2014.
- [23] 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.