DIP-VBTV: A Color Image Restoration Model combining a Deep Image Prior and a Vector Bundle Total Variation

Thomas Batard, Gloria Haro, Coloma Ballester

DEEP IMAGE PERCEPTUAL PRIOR FOR COLOR IMAGE RESTORATION

THOMAS BATARD*, GLORIA HARO †, AND COLOMA BALLESTER †

Abstract. In this paper, we introduce a new prior for image restoration which relies on the assumption that the degradation process preserves some visual attributes of the original scene. We formulate this prior as a penalty term whose minimization expresses the invariance of the visual attributes with respect to the degradation operator. The proposed penalty term is inspired by geometric models in visual psychophysics and neuroscience. More precisely, it combines a model of the perception of colors based on the notion of covariant derivative and a model of edges/textures based on the notion of Riemannian metric, and it yields a non Euclidean extension of the Total Variation (TV) penalty term called Vector Bundle Total Variation (VBTV). Then, under the extra assumption of a Deep Image Prior (DIP), we introduce a variational model DIP-VBTV for image restoration involving the two priors. The proposed model generalizes DIP [31] and DIP-TV [26] models and we show that it outperforms them on denoising and deblurring, turning DIP-VBTV to a state-of-the-art unsupervised method for image restoration.

1. Introduction.

1.1. New perspective on image restoration. There is a growing interest in designing human vision-inspired mathematical models in image processing and computer vision (see e.g. [1],[5],[6],[7],[8],[10],[18],[29]). Dealing with restoration of natural images, this approach is justified by the fact that one aims to maintain the perception of the original scene rather than reproducing its light intensity. This is a very challenging task as the property of the Human Visual System (HVS) to be included in the restoration model depends on the degradations observed on the input image, and it is likely that the vision model describing the desired property of the HVS has to be adapted in order to fit into an image processing model.

From the observation that a clean image and a degraded version of it (noisy, blurry, downsampled,...) still share some visual content, we claim that a model for image restoration should take this information into account, which can be done by making the model preserve, or at most slightly modify, some visual attributes of the degraded image. Nonetheless, the features which should be preserved depend on the nature of the degradation. For instance, dealing with noise, the colors of the original clean image are widely altered (e.g. the hue is modified), whereas local structures (edges, textures) are still visible if the noise is not too high, which is the case in realistic situations. On the other hand, when the degradation comes from a blurring operator, local features are more degraded than colors. Hence, a model for image restoration should, on one hand, be general enough to encode some invariance of the perception of both local structures and color, but also be able to adapt the invariance to a given degradation operator.

1.2. Related work.

1.2.1. Penalty terms of variational models to express perceptual invariance. Over the last 30 years, variational models have demonstrated their efficiency to tackle several tasks in color image restoration, e.g. denoising, deblurring, inpainting, super-resolution, etc (see e.g. [32] and references therein), which are often expressed

* Computer Vision Center, Autonomous University of Barcelona, 08193 Cerdanyola del Vallès, Spain (e-mail: tbatard@cvc.uab.es),

† Department of Information and Communication Technologies, Pompeu Fabra University, 08018 Barcelona, Spain (e-mail: {gloria.haro;coloma.ballester}@upf.edu)

as a convex combination of a data term and one or more penalty terms, the latter(s) being determined by some image prior(s). The fact that the perception of local structures is almost invariant under (realistic) noise degradation has been used in many approaches for image denoising, and is implicitly encoded into a penalty term. Among the seminal penalty terms encoding such invariance, we have the Total Variation (TV) [14],[15],[28] whose minimization encourages the preservation of local structures by means of the L1 norm of the Euclidean gradient, and the Polyakov action [30] whose minimization encourages the preservation of local structures by means of the L2 norm of a Riemannian gradient, the Riemannian metric being related to the structure tensor of the image. These two penalty terms can be extended to color images in the straightforward manner, replacing the gradient of a scalar function by the Jacobian of a vector-valued function. For instance, TV extends to the so-called Vectorial Total Variation (VTV) [9],[11]. A more perceptually-based color extension is the Saturation-Value Total Variation (SVTV) [24] in which the fact that the spatial variations of the local structures of a natural image are mainly in its achromatic component is taken into account, making the model penalize the smoothing of the achromatic component of the image.

These penalty terms have also been applied to various image restoration problems as deblurring, inpainting, super-resolution. However, none of them encode some invariance of the perception of colors with respect those degradations.

1.2.2. The Vector Bundle Total Variation (VBTV). This penalty term has been introduced in [4], in which a multi-channel image $u\colon \Omega \subset \mathbb{R}^2 \longrightarrow \mathbb{R}^n$ is considered as a section of a vector bundle over a Riemannian manifold. Then, VBTV arises as the natural extension of VTV in this geometric context, and is defined by $VBTV(u) = \|Du\|_{L^1(g^{-1}\otimes h)}$, where g stands for a Riemannian metric on the base manifold, D is a covariant derivative and h is a vector bundle metric, making VBTV be determined by the geometric triplet g, D, h.

In the experiments they conducted on image denoising, the authors considered the following triplet: g, h are Euclidean metrics, and D is the flat covariant derivative whose connection 1-form vanishes in a moving frame P, introduced in [2], describing the local geometry of the corrupted image. In this context, the minimization of VBTV implicitly assumes that the moving frame P and consequently the local geometry it encodes is invariant with respect to a noise degradation. Experiments corroborated the relevance of this approach as the results showed that VBTV-L2 model outperforms various variational models for image denoising based on the minimization of a penalty term, including the VTV-L2 model [11]. Whereas such an invariance is coherent when dealing with noise, it might be less efficient in the context of other image restoration tasks where the degradation operator greatly affects the local geometry like deblurring.

1.3. Contribution. Our contribution in this paper is two-fold:

1.3.1. The minimization of VBTV can express some perceptual invariance of colors and local features with respect to a degradation operator. We propose a new interpretation of the geometric triplet g, D, h, in which each element has a perceptual meaning. Following [5], we use covariant differentiation to describe color perception. Then, we use the metric h as a weight function describing the non uniformity of colors in the perception of visual attributes. Finally, we follow the Beltrami framework [30] and use the Riemannian metric g to describe the local structures of an image. We show that this new interpretation enables to express some

perceptual invariance of colors and local features with respect to a degradation operator through the minimization of VBTV. More precisely, the main results we obtain are the following ones.

Covariant derivative to express color perception and its invariance through the minimization of VBTV. We show that, for well-chosen covariant derivatives D induced by connection 1-forms derived from the optimal connection 1-forms constructed in [3] and determined by the degraded image u_0, the sections u minimizing VBTV satisfy

$$Fu = Fu_0 \tag{1.1}$$

for some operator F. Moreover, we show that F has a perceptual interpretation, which makes equality (1.1) describe some invariance related to the perception of colors.

Weighting image components through the vector bundle metric h. By definition of VBTV, the vector bundle metric h can assign different weights to the different image components. Hence, through the minimization of VBTV, it enables to process some image components in a smaller extent than others, which can be desirable in the context of image restoration. In particular, we show that SV-TV, in which different weights are assigned to the achromatic and chromatic components, is a particular case of VBTV for a particular triplet g, D, h.

Riemannian metric derived from a perceptual structure tensor. Under the perceptual interpretations of a covariant derivative and a vector bundle metric, we introduce a generalization of the structure tensor (at the finer scale, scale 0), replacing the Jacobian operator by a covariant derivative and the Euclidean metric by a vector bundle metric, and which we call perceptual structure tensor. Then, we derive a Riemannian metric g from the perceptual structure tensor given, for $\beta > 0$, by the symmetric matrix

$$\begin{pmatrix} 1 + \beta\, h(D_{\partial_x} u, D_{\partial_x} u) & \beta\, h(D_{\partial_x} u, D_{\partial_y} u) \\ \beta\, h(D_{\partial_x} u, D_{\partial_y} u) & 1 + \beta\, h(D_{\partial_y} u, D_{\partial_y} u) \end{pmatrix} \tag{1.2}$$

in the frame (∂_x, ∂_y) induced by the Cartesian coordinates system (x, y) on Ω. This Riemannian metric generalizes the one considered in [30]. In what follows, with a slight abuse of notation, we will identify the Riemannian metric g with the matrix (1.2).
We also show in this paper that the Riemannian metric (1.2) is a critical point of an energy.

1.3.2. A variational model for image restoration including VBTV as penalty term which outperforms standard methods. In order to corroborate our claim that a restoration model should take into account that a clean image and a degraded version of it share some visual content, the proposed approach considers VBTV as a penalty term of a variational problem for image restoration, yielding a model of the form

$$\arg\min_{u \in X} \frac{1}{2} \|H(u) - u_0\|_h^2 + \lambda\, VBTV(u) \tag{1.3}$$

for $\lambda > 0$, where u_0 is the observed degraded image, H is the degradation operator, and X a certain functional space. Whereas variational models assuming that X is a bounded variation space have been widely considered in the past (see [3] for a VBTV-L2 model), we consider here a new class of functional space, where X is the set of images which can be generated through a given neural network.

Recently, a Deep Image Prior (DIP) has been introduced by Ulyakov et al. [31] for image restoration, in which it is assumed that the restored image $\underline{u}$ can be generated through a neural network, yielding the following optimization problem

$$\begin{cases} \underline{\theta} = \arg\min_{\theta} \frac{1}{2}\|H(T_\theta(z)) - u_0\|_2^2 \\ \underline{u} = T_{\underline{\theta}}(z), \end{cases} \tag{1.4}$$

where T_θ is a neural network parametrized by θ whose input z is a random image of the same size as u_0. Note that the optimization is performed here on the set of parameters of the neural network rather than on a space of images. Experiments conducted in [31] showed that model (1.4) outperforms TV-based variational models in a great extent on denoising and super-resolution. The model (1.4) has then been combined with a TV prior, leading to the so-called DIP-TV model [26], given by

$$\begin{cases} \underline{\theta} = \arg\min_{\theta} \frac{1}{2}\|H(T_\theta(z)) - u_0\|_2^2 + \lambda\, TV(T_\theta(z)) \\ \underline{u} = T_{\underline{\theta}}(z), \end{cases} \tag{1.5}$$

for $\lambda > 0$. We consider here the DIP-VBTV model, which consists of replacing TV by VBTV in model (1.5).

In Sect. 4, we show that, for well-chosen covariant derivatives D and metrics h, the model DIP-VBTV outperforms DIP and DIP-VTV, a vectorial extension of DIP-TV, on denoising and deblurring. For denoising, the covariant derivative and the metric are inspired by the work [24] in which different weights are assigned to the achromatic and chromatic components of the image. For deblurring, the covariant derivatives are derived by the optimal connection 1-forms constructed in [3], and which encode some psychophysical laws.
We also compare DIP-VBTV to other methods making use of the DIP prior: DeepRED [27] and a Bayesian approach [16], and we show that better results are obtained with DIP-VBTV.

2. Connection of the geometric triplet to vision. In this section, we show that each element of the geometric triplet g, D, h can have an interpretation in terms of vision science. Let $u\colon \Omega \subset \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ be a color image, and E be the trivial vector bundle $\mathbb{R}^3 \times \Omega \longrightarrow \Omega$. In what follows, we consider u as the expression of a section of E in a moving frame.

2.1. Covariant derivative and perceptual gradient. We denote by $T^*\Omega$ the cotangent bundle of Ω and by $\mathrm{End}(E)$ the bundle of endomorphisms of E. In what follows, we denote by $\Gamma(T^*\Omega \otimes \mathrm{End}(E))$ the set of smooth sections of $T^*\Omega \otimes \mathrm{End}(E)$.

A covariant derivative on E is a differential operator $D \coloneqq d+\omega$, where d stands for the standard differential operator and $\omega \in \Gamma(T^*\Omega \otimes \mathrm{End}(E))$ is called a connection 1-form. Assuming that E is equipped with a G-associated bundle structure, where G is a Lie group acting on the fibers of E through a representation ρ, then the connection 1-form is such that is satisfies a certain transformation law under a moving frame change, and which makes the covariant derivative satisfy a G-equivariance with respect to a

moving frame change. More precisely, let ω be the expression of a connection 1-form in a moving frame, φ the expression of a section of E in the same moving frame, and $\mathcal{G}$ be another moving frame. Then, the expression of ω in the frame $\mathcal{G}$ is given by

$$\mathcal{G}d\mathcal{G}^{-1} + \mathcal{G}\omega\mathcal{G}^{-1} \tag{2.1}$$

The transformation law (2.1) makes D satisfy a G-equivariance property with respect to the moving frame change, i.e.

$$D\mathcal{G}\varphi = \mathcal{G}D\varphi \tag{2.2}$$

Note that by formula (2.1), a connection 1-form is completely determined by its value in a moving frame.
Let us assume that E is endowed with a G-associated bundle. Then, assuming that a moving frame change describes a lighting change, a covariant derivative satisfies a G-equivariance property with respect to a lighting change according to (2.2). In this context, Georgiev [22],[23] interprets a covariant derivative of an image as a perceptual gradient and relates the G-equivariance property of this perceptual gradient to the color constancy property of the HVS.

2.2. Weighting image components through the vector bundle metric h.

2.2.1. Vector bundle metric and brightness perception. It has been shown in [5] that, the brightness of the image u, according to the Helmholtz-Kohlrausch effect [20] in visual psychophysics, can be interpreted as its norm $\|u\|_h$ for a well-chosen metric h.

2.2.2. Vector bundle metric and local structures perception. Given a geometric triplet g, D, h, we have

$$\begin{aligned} VBTV(u) &\coloneqq \|Du\|_{L^1(g^{-1}\otimes h)} \\ &= \int_\Omega \sqrt{\sum_{i,j} g^{ij} h(D_{\partial_{x_i}} u, D_{\partial_{x_j}} u)}\, d\Omega, \end{aligned}$$

where g^{ij} denotes the coefficients of the inverse matrix of g. In particular, let P be a moving frame in which h is of the form

$$h = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{2.3}$$

for $\alpha > 0$, then

$$\begin{aligned} \|Du\|_{g^{-1}\otimes h} = \Big(\sum_{i,j} g^{ij} \Big[&\alpha (D_{\partial_{x_i}} P^{-1}u)_1 (D_{\partial_{x_j}} P^{-1}u)_1 + (D_{\partial_{x_i}} P^{-1}u)_2 (D_{\partial_{x_j}} P^{-1}u)_2 \\ &+ (D_{\partial_{x_i}} P^{-1}u)_3 (D_{\partial_{x_j}} P^{-1}u)_3 \Big] \Big)^{1/2}. \end{aligned} \tag{2.4}$$

Hence, a different weight is assigned to the first component $(DP^{-1}u)_1$, which enables to process this component in a greater ($\alpha > 1$) or smaller ($\alpha < 1$) extent than the other components $(DP^{-1}u)_k, k = 2, 3$, which can be desirable in image restoration.

As an example, let us consider the (constant) frame P given by the matrix

$$P = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \end{pmatrix} \tag{2.5}$$

in the RGB frame, D given by the connection 1-form $\omega \equiv 0$ and h given by (2.3) in the frame P, and g given by the Euclidean metric on Ω. Then VBTV(u) corresponds to the SVTV aforementioned, which assigns to the achromatic component (for $\alpha < 1$) less weight than the chromatic components. As a consequence, a SVTV-L2 variational model will smooth less the achromatic component than the chromatic ones. This is a desirable property in image denoising as the perception of local structures, which are mainly in the achromatic component, is less affected by noise than the perception of colors.

2.3. Riemannian metric induced by a generalization of a structure tensor and its relation to local structures perception and processing. Visual edges and textures are well described by the structure tensor which encodes image derivatives at different scales [21].

As pointed out by Chossat and Faugeras [17], there is a "strong evidence that the visual system of many species is organized in such a way that quantities related to image derivatives are extracted, and hence represented, by neuronal activity". Then, they proposed to model the processing of image edges and textures in the V1 area of the visual cortex. To this end, they suggested the presence of neuronal populations in V1 representing the structure tensor, organized as a hypercolumn at each point of the retinal plane, and whose activity evolves by equations similar to the Wilson-Cowan equations [33],[34].

With the perceptual interpretation of covariant derivatives and vector bundle metrics aforementioned, one can then generalize the structure tensor, replacing the image derivative by a covariant derivative and the Euclidean scalar product between image vectors by the one associated to a vector bundle metric. At the scale 0, it gives

$$\begin{pmatrix} h(D_{\partial_{x_1}}u, D_{\partial_{x_1}}u) & h(D_{\partial_{x_1}}u, D_{\partial_{x_2}}u) \\ h(D_{\partial_{x_1}}u, D_{\partial_{x_2}}u) & h(D_{\partial_{x_2}}u, D_{\partial_{x_2}}u) \end{pmatrix}. \tag{2.6}$$

Denoting by $\Gamma(\mathrm{SP2}(\Omega))$ the set of smooth 2x2 symmetric positive semi-definite matrix fields over Ω, we have the following result.

PROPOSITION 2.1. *The matrix field (2.6) is a critical point of the energy* $X : \Gamma(SP2(\Omega)) \longrightarrow \mathbb{R}^+$, *given by*

$$X(g) = \|Du\|^2_{L^2(g^{-1}\otimes h)}. \tag{2.7}$$

Proof. See Appendix A. □

From (2.6), we derive a Riemannian metric

$$g^* = \begin{pmatrix} 1 + \beta\, h(D_{\partial_{x_1}}u, D_{\partial_{x_1}}u) & \beta h(D_{\partial_{x_1}}u, D_{\partial_{x_2}}u) \\ \beta h(D_{\partial_{x_1}}u, D_{\partial_{x_2}}u) & 1 + \beta\, h(D_{\partial_{x_2}}u, D_{\partial_{x_2}}u) \end{pmatrix} \tag{2.8}$$

for $\beta > 0$, which endows (Ω, g^*) with a Riemannian manifold structure. In particular, for h represented by $\mathbb{I}_3$, the 3x3 Identity matrix, and D given by the connection 1-form $\omega \equiv 0$ both expressed in the RGB color space, then g^* corresponds to the Riemannian metric in the Beltrami framework [30] used for edge-preserving image denoising through the heat diffusion of the corresponding Laplace-Beltrami operator.

3. A class of covariant derivatives and their parallel sections. Under the assumption that the metrics g and h are positive definite, we have

$$VBTV(u) = 0 \Longleftrightarrow Du = 0, \tag{3.1}$$

i.e. the sections minimizing VBTV are the parallel sections of D, provided that D does possess parallel sections. Recall that a covariant derivative admits parallel sections if and only if the curvature of the corresponding connection 1-form vanishes identically.

In [3], a family of connection 1-forms ω^u parametrized by a set of Lie groups G acting on the pixel values of images u has been constructed. In what follows, we focus on the connection 1-forms induced by $G = \mathbb{R}^{+*}$, SO(2), SO(3) acting respectively on $\mathbb{R}$ (gray-level image), $\mathbb{R}^2$ (chrominance image), $\mathbb{R}^3$ (color image). In Sect. 3.1, we show that the corresponding covariant derivatives $D^u := d + \omega^u$ satisfy a G-equivariance with respect to u, in the sense that

$$D^{\mathcal{G}u}(\mathcal{G}u) = \mathcal{G}D^u u$$

for any G-valued moving frame $\mathcal{G}$.
Moreover, we show that these covariant derivatives can describe some invariance of a perceptual attribute with respect to a degradation operator through their parallel sections (the existence of parallel sections for these covariant derivatives has been proved in [5]). Indeed, let v be a degraded image and the corresponding covariant derivative $D^v := d + \omega^v$. We show that the parallel sections u of D^v satisfy

$$Fu = Fv \tag{3.2}$$

for some operator F describing some perceptual attribute. Finally, in Sect. 3.2 and Sect. 3.3, we construct new covariant derivatives derived from the three covariant derivatives analyzed in Sect. 3.1 and we study the existence of parallel sections for these new covariant derivatives.

3.1. Covariant derivatives whose parallel sections have a perceptual interpretation.

3.1.1. Covariant derivative induced by the action of $\mathbb{R}^{+*}$ on $\mathbb{R}$ and the corresponding parallel sections. Let us assume that u is a gray-level image. The connection 1-form induced by the action of $\mathbb{R}^{+*}$ on $\mathbb{R}$ is given by

$$\omega^u = -\frac{du}{u}. \tag{3.3}$$

The quantity du/u can be interpreted as the perceptual gradient of the image according to Weber's law in brightness perception. In this case, the $\mathbb{R}^{+*}$-equivariance of D^u with respect to u is trivial as we have

$$D^u u = 0. \tag{3.4}$$

Let v be a degraded gray-level image and $D^v \coloneqq d + \omega^v$ the covariant derivative induced by ω^v. Then, we have

$$D^v u = 0 \Longleftrightarrow \frac{du}{u} = \frac{dv}{v}. \tag{3.5}$$

In the context of image restoration, formula (3.5) together with (3.1) show that the minimization of VBTV encourages the preservation of the perceived gradient according to Weber-Fechner's law.

3.1.2. Covariant derivative induced by the action of SO(2) on $\mathbb{R}^2$ and the corresponding parallel sections. Let us assume that $u = (u_1, u_2)$ is a chrominance image. The connection 1-form induced by the action of SO(2) on $\mathbb{R}^2$ is

$$\omega^u = \begin{pmatrix} 0 & \dfrac{u_1 du_2 - u_2 du_1}{\|u\|^2} \\ -\dfrac{u_1 du_2 - u_2 du_1}{\|u\|^2} & 0 \end{pmatrix}. \tag{3.6}$$

It gives

$$D^u u = d \log(\|u\|)\, u, \tag{3.7}$$

which proves the SO(2)-equivariance of D^u with respect to u.

Let $v = (v_1, v_2)$ be a degraded $\mathbb{R}^2$-valued image, and $D^v \coloneqq d + \omega^v$ the covariant derivative induced by ω^v. We have the following result.

PROPOSITION 3.1. *The parallel sections of D^v are the sections u satisfying*

$$\begin{cases} d\|u\| = 0 \\ \dfrac{u_1 du_2 - u_2 du_1}{\|u\|^2} = \dfrac{v_1 dv_2 - v_2 dv_1}{\|v\|^2}. \end{cases} \tag{3.8}$$

Proof. See Appendix B. □

In polar coordinates $u = (r(u), \varphi(u))$ and $v = (r(v), \varphi(v))$, the coordinate r corresponds to the saturation component (up to the multiplication by a constant) and the coordinate φ to the hue. Moreover, the second equality in (3.8) reads

$$d\varphi(u) = d\varphi(v). \tag{3.9}$$

In the context of image restoration, formulas (3.8) imply that the minimization of VBTV favours the preservation of the variations of the hue of the image through the equality (3.9) and it encourages images with smooth saturation through the equality $d\|u\| = 0$.

3.1.3. Covariant derivative induced by the action of SO(3) on $\mathbb{R}^3$ and the corresponding parallel sections. Let $u = (u_1, u_2, u_3)$ be a color image. The connection 1-form ω^u induced by the action of SO(3) on $\mathbb{R}^3$ is given by

$$\omega^u = \begin{pmatrix} 0 & \frac{(u_1 du_2 - u_2 du_1)}{\|u\|^2} & \frac{(u_1 du_3 - u_3 du_1)}{\|u\|^2} \\ -\frac{(u_1 du_2 - u_2 du_1)}{\|u\|^2} & 0 & \frac{(u_2 du_3 - u_3 du_2)}{\|u\|^2} \\ -\frac{(u_1 du_3 - u_3 du_1)}{\|u\|^2} & -\frac{(u_2 du_3 - u_3 du_2)}{\|u\|^2} & 0 \end{pmatrix}. \tag{3.10}$$

As in the SO(2) case, the covariant derivative $D^u := d + \omega^u$ induced by ω^u satisfies

$$D^u u = d \log(\|u\|)\, u. \tag{3.11}$$

which proves the SO(3)-equivariance of D^u.
Let $v = (v_1, v_2, v_3)$ be a degraded color image and $D^v := d + \omega^v$ the covariant derivative induced by ω^v. We have the following result.

PROPOSITION 3.2. *The parallel sections of D^v are the sections u satisfying*

$$\begin{cases} d\|u\| = 0 \\ \omega^u u = \omega^v u \end{cases} \tag{3.12}$$

Proof. See Appendix C. □

Assuming that u, v are expressed with their RGB components and the norm is the one describing the Helmholtz-Kohlrausch effect (see [5]), formula $d\|u\| = 0$ in (3.12) implies that the minimization of VBTV encourages images with smooth brightness.

3.2. Combining the connection 1-forms induced by $\mathbb{R}^{+*}$, SO(2) and SO(3). New connection 1-forms ω^u for color images $u = (u_1, u_2, u_3)$ can be derived from the connection 1-forms (3.3),(3.6),(3.10).

Let us first consider the following combination of the connection 1-forms induced by $\mathbb{R}^{+*}$ (3.3) and SO2 (3.6)

$$\omega^u = \begin{pmatrix} -\frac{du_1}{u_1} & 0 & 0 \\ 0 & 0 & \frac{u_2 du_3 - u_3 du_2}{\|u_{2,3}\|^2} \\ 0 & -\frac{u_2 du_3 - u_3 du_2}{\|u_{2,3}\|^2} & 0 \end{pmatrix} \tag{3.13}$$

where $u_{2,3}$ denotes the $\mathbb{R}^2$-valued image (u_2, u_3). We have

$$D^u u = (0, d\log(\|u_{2,3}\|)\, u_2, d\log(\|u_{2,3}\|)\, u_3)^T,$$

which shows that the covariant derivative D^u satisfies an $\mathbb{R}^{+*}\times$ SO(2)-equivariance according to (3.4),(3.7).

Let $v = (v_1, v_2, v_3)$ be a degraded color image, and $D^v := d + \omega^v$ the covariant derivative induced by ω^v. We deduce from (3.5) and (3.8) that

$$D^v u = 0 \Longleftrightarrow \begin{cases} \dfrac{d\,u_1}{u_1} = \dfrac{d\,v_1}{v_1} \\ d\|u_{2,3}\| = 0 \\ \dfrac{u_2 du_3 - u_3 du_2}{\|u^{2,3}\|^2} = \dfrac{v_2 dv_3 - v_3 dv_2}{\|v^{2,3}\|^2}. \end{cases} \tag{3.14}$$

Then, the existence of parallel sections is a direct consequence of the existence of parallel sections for the connection 1-forms (3.3) and (3.6).

Assuming that u, v are expressed in an opponent color space, in which u_1, v_1 is the achromatic component and $(u_2, u_3), (v_2, v_3)$ the chromatic components, we deduce from (3.14) that the minimization of VBTV encourages the preservation of the perceptual gradient of the achromatic component and the preservation of the gradient of the hue. Moreover, it encourages images with smooth saturation.

Let us now consider a combination of the connection 1-forms induced by $\mathbb{R}^{+*}$ (3.3) and SO(3) (3.10). To that purpose, we first extend the connection 1-form (3.3) to a connection 1-form on a vector bundle of rank 3 for color images

$$\begin{pmatrix} -\dfrac{du_1}{u_1} & 0 & 0 \\ 0 & -\dfrac{du_2}{u_2} & 0 \\ 0 & 0 & -\dfrac{du_3}{u_3} \end{pmatrix}, \tag{3.15}$$

which can be associated to the action of the group DC(3) of diagonal 3x3 matrices with positive entries on $\mathbb{R}^3$. Then, we combine the connection 1-form (3.15) with (3.10), giving the connection 1-form

$$\omega^u = \begin{pmatrix} -\dfrac{du_1}{u_1} & \dfrac{(u_1 du_2 - u_2 du_1)}{\|u\|^2} & \dfrac{(u_1 du_3 - u_3 du_1)}{\|u\|^2} \\ -\dfrac{(u_1 du_2 - u_2 du_1)}{\|u\|^2} & -\dfrac{du_2}{u_2} & \dfrac{(u_2 du_3 - u_3 du_2)}{\|u\|^2} \\ -\dfrac{(u_1 du_3 - u_3 du_1)}{\|u\|^2} & -\dfrac{(u_2 du_3 - u_3 du_2)}{\|u\|^2} & -\dfrac{du_3}{u_3} \end{pmatrix}. \tag{3.16}$$

The covariant derivative D^u induced by ω^u satisfies

$$D^u u = -du + d\log(\|u\|)u,$$

which shows that it does not satisfy a G-equivariance.

Let v be a color image, and $D^v := d + \omega^v$ the covariant derivative induced by ω^v.

Unlike the previous cases, this covariant derivative does not admit parallel sections. Indeed, we have the following result.

PROPOSITION 3.3. *The curvature of the connection 1-form (3.16) does not identically vanish.*

Proof. See Appendix D. □

The non existence of parallel sections for this covariant derivative makes more difficult to perform an analysis of the minimization of VBTV.

3.3. Dual connections. To any connection 1-form ω is associated a connection 1-form ω^* by means of an involution on $\Gamma(T^*\Omega \otimes \mathrm{End}(E))$. ω and ω^* are said to be dual to each other.
Let us consider the Cartan involution on $\mathfrak{gl}_n(\mathbb{R})$, given by

$$\theta(X) = -X^T. \tag{3.17}$$

The Cartan involution extends in a straightforward way to an involution Θ on $\Gamma(T^*\Omega \otimes \mathrm{End}(E))$. Then, given any connection 1-form ω, the dual connection 1-form induced by Θ is $\omega^* \coloneqq \Theta \circ \omega$.

3.3.1. Dual of the connection 1-form (3.13). The dual of the connection 1-form ω^u in (3.13) is

$$\omega^{u*} = \begin{pmatrix} \dfrac{du_1}{u_1} & 0 & 0 \\ 0 & 0 & \dfrac{u_2 du_3 - u_3 du_2}{\|u_{2,3}\|^2} \\ 0 & -\dfrac{u_2 du_3 - u_3 du_2}{\|u_{2,3}\|^2} & 0 \end{pmatrix} \tag{3.18}$$

where $u_{2,3} \coloneqq (u_2, u_3)$. It gives

$$D^u u = \left(2du_1, d\log\left(\|u_{2,3}\|\right) u_2, d\log\left(\|u_{2,3}\|\right) u_3\right)^T,$$

which shows that, unlike (3.13), the covariant derivative induced by the connection 1-form (3.18) does not satisfy a $\mathbb{R}^{+*}\times$ SO(2)-equivariance.

Let $v = (v^1, v^2, v^3)$ be a degraded color image, and $D^v \coloneqq d + \omega^{v*}$ the covariant derivative induced by ω^{v*}. We have

$$D^v u = 0 \Longleftrightarrow \begin{cases} \dfrac{d\,u_1}{u_1} = -\dfrac{d\,v_1}{v_1} \\ d\|u_{2,3}\| = 0 \\ \dfrac{u_2 du_3 - u_3 du_2}{\|u_{2,3}\|^2} = \dfrac{v_2 dv_3 - v_3 dv_2}{\|v_{2,3}\|^2}. \end{cases} \tag{3.19}$$

Let us assume that u, v are expressed in an opponent color space. Then, as the connection 1-form (3.13), the minimization of VBTV induced by (3.18) encourages the

preservation of the gradient of the hue and it favours images with smooth saturation. However, unlike (3.13), it does not encourage the preservation of the perceptual gradient of the achromatic component as it encourages the reversion of its sign.

3.3.2. Dual of the connection 1-form (3.16)). The dual of the connection 1-form ω^u (3.16) according to the involution Θ is the connection 1-form

$$\omega^{u*} := \begin{pmatrix} \dfrac{du_1}{u_1} & \dfrac{(u_1du_2 - u_2du_1)}{\|u\|^2} & \dfrac{(u_1du_3 - u_3du_1)}{\|u\|^2} \\ -\dfrac{(u_1du_2 - u_2du_1)}{\|u\|^2} & \dfrac{du_2}{u_2} & \dfrac{(u_2du_3 - u_3du_2)}{\|u\|^2} \\ -\dfrac{(u_1du_3 - u_3du_1)}{\|u\|^2} & -\dfrac{(u_2du_3 - u_3du_2)}{\|u\|^2} & \dfrac{du_3}{u_3} \end{pmatrix}, \tag{3.20}$$

and the covariant derivative induced by this connection 1-form satisfies

$$D^u u = du + d\log(\|u\|)u,$$

which shows that it does not satisfy a G-equivariance.

Let v be a degraded color image, and $D^v := d + \omega^{v*}$ the covariant derivative induced by ω^{v*}. As in the case of the covariant derivative induced by the connection 1-form (3.16), the covariant derivative D^v does not admit parallel sections. Indeed, we have the following result.

PROPOSITION 3.4. *The curvature of the connection 1-form (3.20) does not identically vanish.*

Proof. The proof is the same as the one of Proposition 3.3. □

As in the case of the covariant derivative induced by the connection 1-form (3.16), the non existence of parallel sections for this covariant derivative makes more difficult to perform an analysis of the minimization of VBTV.

4. DIP-VBTV model for image restoration. In this Section, we consider the DIP-VBTV model

$$\begin{cases} \underline{\theta} = \arg\min_{\theta} \dfrac{1}{2}\|H(T_\theta(z)) - u_0\|_2^2 + \lambda\, VBTV(T_\theta(z)) \\ \underline{u} = T_{\underline{\theta}}(z), \end{cases} \tag{4.1}$$

for different degradation operators H: noise (Sect. 4.2) and blur (Sect. 4.3). We test the model for different geometric triplets g, D, h and the results show that the best geometric triplet depends not only on the degradation considered but also on the image u_0 to process.

4.1. On the numerical scheme to solve the optimization problem. We use the same network T_θ in all the experiments conducted in this paper, i.e. an encoder-decoder with skip connections between the down and up layers. It corresponds to the default architecture in [31], which we refer to for details about the

architecture. The same network is actually used in the DIP-TV [26], DeepRED [27] and Bayesian DIP [16] models.

Assuming that the size of the input image u_0 (gray-level or color) is $M \times N \times$ (1 or 3), we take as input z of the network a random image of size $M \times N \times 32$, as it is done in the aforementioned papers.

4.1.1. A boosting numerical scheme. Following the approach in [31], and denoting by $E(\theta; z)$ the energy in (4.1), we consider the following numerical scheme in order to approximate a solution of the DIP-VBTV model

$$\begin{cases} n_{k+1} \sim \mathcal{N}(0, \sigma) \\ z_{k+1} = z_0 + n_{k+1} \\ \theta_{k+1} = \theta_k - lr \nabla E(\theta_k; z_{k+1}) \\ u_{k+1} = \gamma\, u_k + (1-\gamma)\, T_{\theta_{k+1}}(z_{k+1}), \end{cases} \tag{4.2}$$

where z_0 is a fixed random image, lr denotes the learning rate, ∇ stands for the gradient with respect to the first argument, and $0 < \gamma < 1$.
We can observe from (4.2) that the input z_k of the network differs at each iteration by perturbing the input random image z_0 with additive white Gaussian noise of variance σ. This technique is called noise-based regularization, and experiments showed that the restoration benefits from this type of regularization. Note that, it has been pointed out in [31] that, even if the noise-based regularization impedes the optimization process of the model (1.4), this latter eventually reaches the value 0 of the energy for a large enough number of iterations.

Last line in (4.2) reveals another boosting technique employed in the numerical scheme, which consists of using an exponential sliding window for a well-chosen weight γ.

Finally, a last boosting technique employed in [31] consists of averaging the output images of the numerical scheme (4.2) over two different runs.

4.1.2. Stopping criteria. It has been observed in [31] that the numerical scheme (4.2) applied to DIP generates noise when the number of iterations is too large. We observed the same behavior when applied to DIP-VBTV.

Another property of the numerical scheme is that it can suffer from, what is called in [31], destabilization, i.e., a significant increase of the energy $E(\theta_k; z_k)$ and blur in the generated image $T_{\theta_k}(z_k)$ can occur during the iterative procedure. Then, from such destabilization point, the energy goes down again till destabilized one more time. In order to prevent destabilization, the strategy adopted in [31] consists of tracking the optimization loss and return to parameters from the previous checkpoint iteration and stop the numerical scheme if the loss difference between two consecutive checkpoint iterations is higher than a certain threshold.

As a consequence, the stopping criteria of the numerical scheme should be carefully chosen. Indeed, the final iteration k^* should be early enough so that the image u_k^* does not possess noise and destabilization has not occurred yet, but it should also stop late enough so that the degradations generated by H are not presented in u_k^*.

In [35], an automated stopping method named Orthogonal Stopping Criterion (OSC) has been proposed, which adds a pseudo noise to the corrupted image and measures the pseudo noise component in the recovered image of each iteration based on the orthogonality between signal and noise. In [16], they avoid the need of early

stopping by conducting posterior inference using stochastic gradient Langevin dynamics.

In the experiments conducted in this paper, we follow the strategy in [31].

4.1.3. Parameters of the model and the numerical scheme. The proposed approach for image restoration has several parameters, which can be split into two categories.

1. The parameters of the model (4.1):
- The connection 1-form ω which determines the covariant derivative D.
- The vector bundle metric h.
- The trade-off parameter β between spatial and color variations in the Riemannian metric (2.8).
- The trade-off parameter λ between the data term and the penalty term in (4.1).

2. The parameters of the numerical scheme (4.2):
- The variance σ of the noise-based regularization.
- The learning rate lr.
- The weight γ of the exponential sliding window.
- The number of iterations.

4.2. Denoising. In this Section, we test DIP-VBTV on denoising, i.e. we consider the model (4.1) for H being the Identity operator. The chosen covariant derivative and vector bundle metric are the ones described in Sect. 2.2.2, i.e. the covariant derivative is given by the connection 1-form $\omega \equiv 0$ and the metric h is given by (2.3) for $\alpha = 0.3$ both in the frame (2.5). As pointed out in Sect. 2.2.2, this choice for ω and h aims at smoothing the chromatic components of the input noisy image in a greater extent than its achromatic component. Indeed, the perception of local details of the clean image, which are mainly in the achromatic component, is less affected by noise than the perception of colors.

The parameter of the Riemannian metric (2.8) is taken as $\beta = 3000$, and the trade-off parameter of the model is $\lambda = 0.1$. The parameters of the numerical scheme (4.2) are the default parameters of the DIP model [31] for denoising: $\sigma = 1/30, lr = 0.01, \gamma = 0.99$, except the number of iterations as we will see below.

We follow [31] and test our model on a database of 9 color images `http://www.cs.tut.fi/~foi/GCF-BM3D/`, which contains 8 natural images and 1 synthetic image, for additive white Gaussian noise of variance 25.

4.2.1. Dependence of the optimal number of iterations to the image size and the prior. In the denoising experiments, the numerical schemes of DIP [31] and DeepRED [27] are stopped after different numbers of iterations (1800 for DIP and 6000 for DeepRED), which shows that the number of iterations at which a model reaches its best PSNR greatly varies with the model itself.

We aim at analyzing the evolution of the PSNR of DIP-VBTV with respect to the number of iterations and compare it to the one of DIP. To this end, we run the numerical scheme (4.2) for both models applied to each of the 9 images of the dataset and stop it after 10K iterations. The results, shown on Fig. 4.1, reveal that DIP-VBTV is more stable than DIP in the sense that, for any image, the drop of the PSNR after the peak is less important. Table 4.1 indicates the highest PSNR value of both models, and the iteration at which it is reached is indicated in parenthesis. In this table,

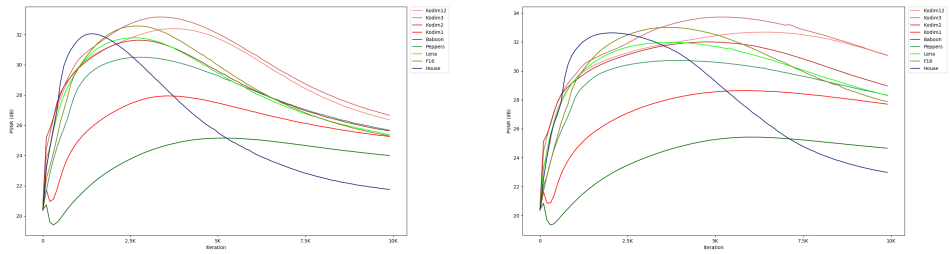

FIG. 4.1. *Evolution of the PSNR with respect to the number of iterations for two models: DIP (left plot) and DIP-VBTV (right plot).*

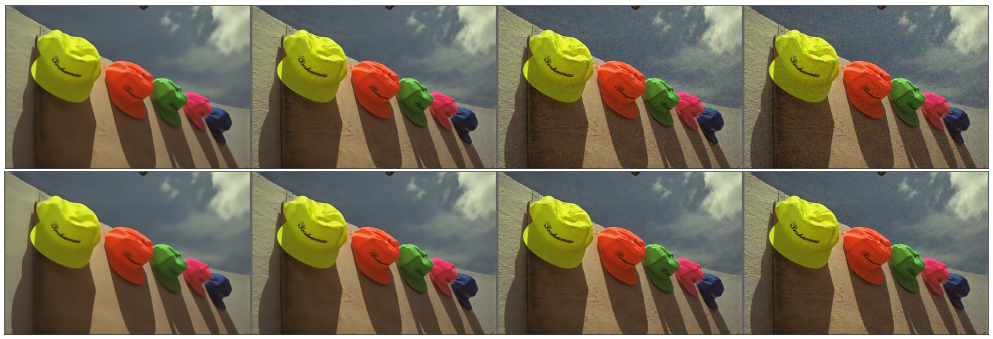

FIG. 4.2. *Evolution of the image "Kodim3" at different iterations tested on denoising with two models. Top row: DIP - Bottom row: DIP-VBTV. From left to right: 1K, 2.5K, 5K, 7.5K.*

images are ordered according to their size from left to right: 256x256 (House), 512x512 (Peppers, Lena, Baboon, F16), 768x512 (Kodim1, Kodim2, Kodim3, Kodim12). The results show that DIP reaches its highest PSNR at earlier iterations but DIP-VBTV gives better results with a mean improvement of 0.4 dB.

Fig. 4.2 compares intermediate results at the iterations 1K, 2.5K, 7.5K, 10K of the two models tested on the image "Kodim3". We observe that DIP tends to generate more noise than DIP-VBTV.

A closer look at Fig. 4.1 and Table 4.1 shows a correlation between the iteration at which the highest PSNR is reached and the image size, with the exception of the synthetic image Baboon whose highest PSNR is reached after a number of iterations similar to the ones of bigger images. We argue that this behavior of the models comes from:
1) the particular form of the network $T_\theta(z)$ makes the numerical scheme reconstruct the image from its lowest to highest frequencies throughout the iterative process (see Fig. 4.2).
2) natural images of bigger size possess finer details in general.
3) the synthetic image Baboon possesses a lot of fine details.

4.2.2. Automation of the stopping criteria and boosting of the results. Based on the results obtained in the previous experiment, we deduce that an automated stopping criteria should depend on the image size and the model itself. The proposed stopping criteria for the images of the database is given in Table 4.2.

Table 4.3 reports the mean and the standard deviation of the PSNR after 5 runs

TABLE 4.1
Denoising: Max PSNR over one run (in parenthesis, iteration at which the max PSNR value is reached)

Algorithm	House	Peppers	Lena	Baboon	F16	Kodak1	Kodak2	Kodak3	Kodak12	Average
DIP	32.06 (1415)	30.51 (2753)	31.79 (2587)	25.16 (5181)	32.58 (2686)	27.94 (3593)	31.63 (2735)	33.16 (3367)	32.40 (3672)	30.80 (3110)
DIP-VBTV	**32.63** (2015)	**30.72** (3883)	**31.98** (3729)	**25.42** (5976)	**33.01** (3616)	**28.63** (5880)	**32.00** (4745)	**33.72** (5143)	**32.68** (6498)	**31.20** (4609)

TABLE 4.2
Denoising: Iteration at which the numerical scheme (4.2) is stopped.

Algorithm	256×256	512×512	768×512
DIP	1500	2500	3000
DIP-VBTV	2000	4000	5000

of DIP and DIP-VBTV tested on the whole dataset with the automated stopping criteria described in Table 4.2. We observe that the improvement of DIP-VBTV over DIP (+0.4 dB) reported in Table 4.1 has increased (+0.51 dB). This small difference (+0.11 dB) shows the accuracy of the proposed stopping criteria, even if stopping the DIP model at 2500 iterations for the image Baboon seems to be too early (compare the PSNR of the image Baboon at Table 4.1 and Table 4.3). The difference is even smaller (+0.02dB) if we disregard the image Baboon as the improvement of DIP-VBTV over DIP gets +0.43dB.
Note that the standard deviation is rather small in both cases (0.03 for DIP and 0.04 for DIP-VBTV), which shows the stability of the numerical scheme for the chosen parameters.
Fig. 4.3 compares the best results over the 5 runs of DIP (left images, second and fourth row) and DIP-VBTV (right images, second and fourth row) on the images "Kodak3" and "Kodim12". In this case, the best results of DIP-VBTV have a PSNR of 33.70 dB for "Kodak3", 32.67 dB for "Kodak12", and the best results of DIP have a PSNR of 33.07 dB for "Kodak3" and 32.17 dB for "Kodak12". We observe that DIP-VBTV provides images which are perceptually closer to the clean images (left images, first and third row). Indeed, we can see for instance that the details are sharper (textures on the walls, letters on the caps for "Kodak3", textures in the sand area for "Kodak12") and the homogeneous regions are more preserved (caps for "Kodak 3" and skin of the people for "Kodak 12").

Finally, we consider the boosting technique mentioned in Sect. 4.1.1, and which consists of averaging the output images over several runs. Whereas an averaging over 2 runs has been used in [31], we noticed that an averaging over 5 runs provides a bigger improvement. Results are reported in Table 4.4, and it shows that DIP-VBTV outperforms DIP by 0.42 dB (by 0.33dB if we disregard the image Baboon), which means that DIP benefits more from the boosting than DIP-VBTV does.

4.2.3. Comparison to other methods. The same database and noise level have been used to test DeepRED [27] and Bayesian DIP [16]. Results report a score of 31.24 dB for DeepRED and 30.81dB for Bayesian DIP. The standard nonlocal methods NL-Means [12] and CBM3D [19] have also been tested on this dataset. Ulyanov et al. [31] report a score of 30.26 dB for NL-Means with the implementation [13] and

TABLE 4.3
Denoising: Mean PSNR and standard deviation over 5 runs

Algorithm	House	Peppers	Lena	Baboon	F16	Kodak1	Kodak2	Kodak3	Kodak12	Average
DIP	32.05	30.45	31.73	23.80	32.52	27.73	31.67	33.03	32.13	30.57
	± **0.02**	± **0.02**	± 0.04	± **0.06**	± **0.02**	± 0.03	± 0.04	± **0.02**	± **0.05**	± **0.03**
DIP-VBTV	**32.59**	**30.70**	**31.97**	**24.83**	**32.94**	**28.54**	**31.90**	**33.64**	**32.59**	**31.08**
	± **0.02**	± **0.02**	± **0.03**	± 0.08	± 0.03	± **0.02**	± **0.02**	± 0.04	± 0.06	± 0.04

TABLE 4.4
Denoising: PSNR of the mean image over 5 runs

Algorithm	House	Peppers	Lena	Baboon	F16	Kodak1	Kodak2	Kodak3	Kodak12	Average
DIP	32.58	30.73	32.09	23.98	33.02	28.16	32.15	33.40	32.47	30.95
DIP-VBTV	**32.84**	**30.89**	**32.20**	**25.10**	**33.28**	**29.02**	**32.26**	**33.91**	**32.81**	**31.37**

a score of 31.42 dB for CBM3D with the implementation [25].
Hence, with our score of 31.37 dB, we are approaching the state-of-the-art unsupervised denoising method CBM3D.

Finally, let us mention the DIP-TV model (1.5) introduced in [24], which has been applied and compared to DIP on another dataset containing 8 color images of size 512x512. The results, which tell that DIP-TV outperforms DIP, have been reported in SNR, which makes difficult to compare against our results. Moreover, as pointed out in [27], both DIP-TV and DIP have been stopped after 5000 iterations, in which DIP might not perform well for such image size (see e.g. Fig. 1 (left) which shows that higher PSNR is obtained at smaller iterations). Hence, the performance of DIP-TV is questionable.

4.3. Deblurring. In this Section, we test DIP-VBTV on deblurring, i.e. we consider the model (4.1) for H being a blur operator. We reproduce the experiments conducted in [27] in which two types of blur operators are considered: a 25x25 Gaussian blur with variance 1.6 and 9x9 uniform blur. In both cases, the blurry image is further contaminated by white additive Gaussian noise of variance $\sqrt{2}$. The experiments are conducted on a set of four color images of same size (256x256).

Here, we test DIP-VBTV for six different geometric triplets (see Table 4.5 and Sect. 4.1.3). Note that, unlike the denoising case, we use the Identity matrix as the metric in the opponent space (2.5). One can also observe that the trade-off parameter is smaller for DIP-VBTV Euclidean ($\lambda = 0.0001$). Indeed, preliminary experiments showed that DIP-VBTV Euclidean gives better results when applied with $\lambda = 0.0001$. In particular, it produces over-smoothed images for $\lambda = 0.001$. The parameters of the numerical scheme (4.2) are: $\sigma = 0.01$, $lr = 0.001$, $\gamma = 0.99$. The same parameters are used to test DIP.

4.3.1. Optimal number of iterations. Following the results in Sect. 4.2.1 in which it has been shown that, for a given image, DIP and DIP-VBTV reach their optimal PSNR at very different iterations, we run the numerical scheme (4.2) for both DIP and the six DIP-VBTV models for a very large number of iterations (30K). In such a way, we get some insight about the value of the iteration at which the PSNR is maximized for each model.

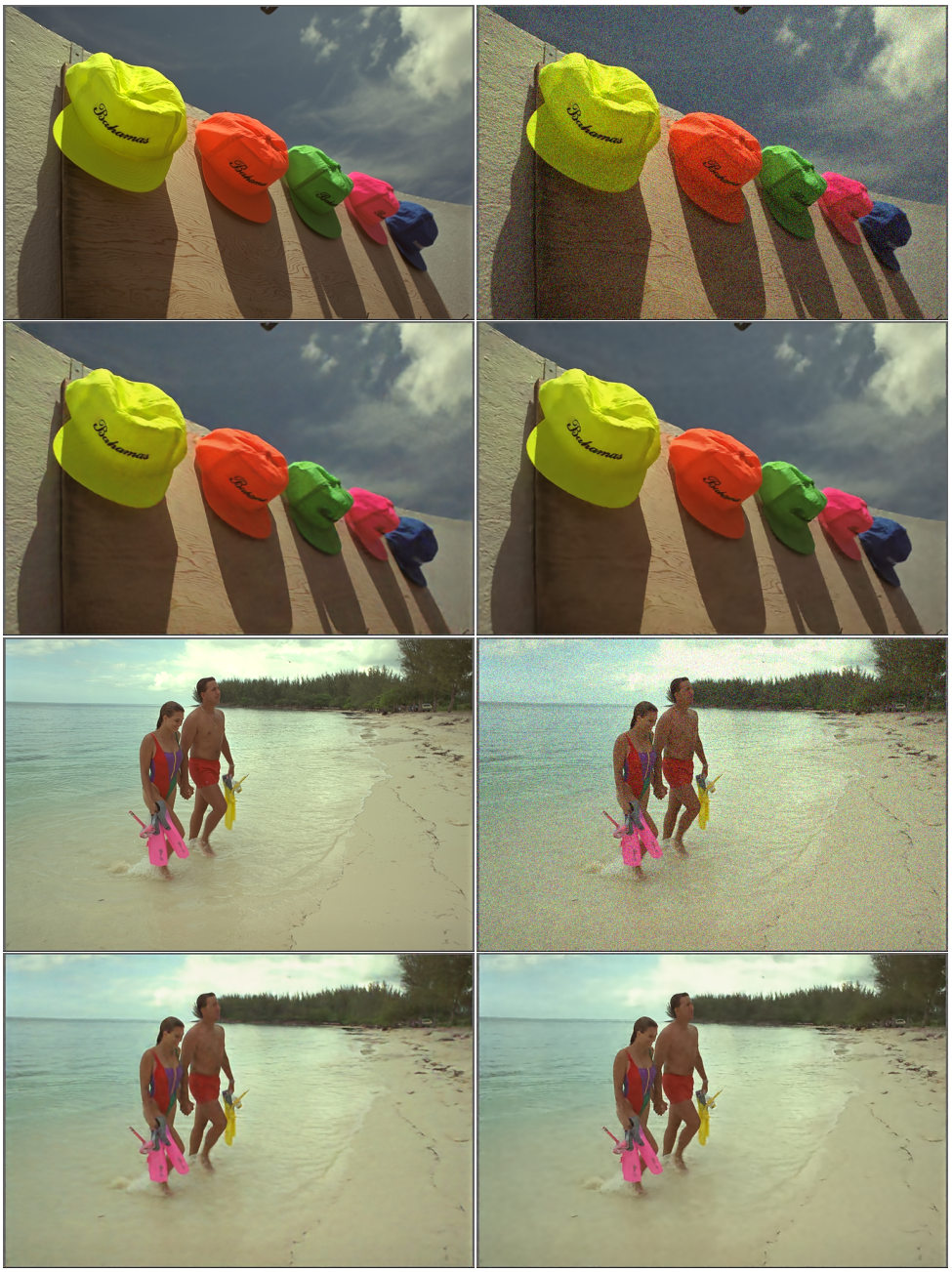

FIG. 4.3. *Comparison of DIP and DIP-VBTV models for denoising. Clockwise from top-left to bottom-left for each image: Clean ground truth images "Kodak3" and "Kodak12" - Input noisy image obtained by adding white Gaussian noise of variance 25 to the clean images - Result of DIP-VBTV model - Result of DIP model. In both cases, best results, in terms of PSNR, over the 5 runs.*

In Table 4.6 and Table 4.7, we report the results for Gaussian and uniform blur respectively. Results show that, for each image and degradation operator, the best

TABLE 4.5
Configurations of the model (4.1) tested for deblurring

Name	ω	Color space	h	β	λ
DIP-VBTV Euclidean	0	RGB	$\mathbb{I}_3$	0	0.0001
DIP-VBTV Riemannian	0	RGB	$\mathbb{I}_3$	3000	0.001
DIP-VBTV (3.13)	(3.13)	Opponent space (2.5)	$\mathbb{I}_3$	3000	0.001
DIP-VBTV (3.18)	(3.18)	Opponent space (2.5)	$\mathbb{I}_3$	3000	0.001
DIP-VBTV (3.16)	(3.16)	RGB	$\mathbb{I}_3$	3000	0.001
DIP-VBTV (3.20)	(3.20)	RGB	$\mathbb{I}_3$	3000	0.001

result is obtained by a DIP-VBTV model for $\omega \neq 0$, this latter varying with both the image and the degradation. The average columns confirm the improvement of DIP-VBTV models for $\omega \neq 0$ with respect to DIP. Indeed, the improvement goes from +0.34 dB for the connection 1-form (3.16) to +0.49 dB for the connection 1-form (3.18) in the Gaussian blur case. In the uniform blur case, the improvement goes from +0.38 dB for the connection 1-form (3.20) to +0.54 dB for the connection 1-form (3.16). Note that DIP-VBTV Euclidean and DIP-VBTV Riemannian models, where $\omega \equiv 0$, perform worse than DIP-VBTV for $\omega \neq 0$ but outperform DIP in average (except DIP-VBTV Euclidean in the uniform blur case).
By observing the iterations at which the highest PSNR are attained, we observe a strong correlation between the values of the iteration and the values of the PSNR. Indeed, in most of the cases, the model providing the best PSNR is the one where the highest PSNR is obtained at the latest iteration. On the other hand, in most of the cases, (if we disregard DIP-VBTV Euclidean which uses a different trade-off parameter), the model providing the worse PSNR is the one where the highest PSNR is obtained at the earliest iteration. Finally, we observe that the four DIP-VBTV models with $\omega \neq 0$ reach their highest PSNR later than DIP. The models DIP-VBTV Euclidean and Riemannian reach their highest PSNR at intermediate iterations.

Fig. 4.4 and Fig. 4.5 compare the evolution of the PSNR throughout the iterative process for DIP (left plots) and DIP-VBTV (3.16) (right plots) models tested for both Gaussian and uniform blur on the whole dataset. We observe that DIP-VBTV is more stable than DIP in the sense that the PSNR decreases slower after the peak in all the cases. Fig. 4.6 compares intermediate results at the iterations 2.5K, 5K, 10K, 30K of the two models tested on the image "Parrots" with uniform blur. The images confirm the difference in the behavior of the PSNR curves observed in Fig. 4.5. Indeed, the images at 10K (which corresponds more or less to the peak of the PSNR in both models) and at 30K are very different in the case of DIP and much less in the case of DIP-VBTV. In particular, DIP generates a very noisy image at 30K (top row, right column) which coincides with the low PSNR observed in the red curve at 30K (left plot in Fig. 4.5).

4.3.2. Automation of the stopping criteria. Following the results shown in Sect. 4.3.1, we propose the following stopping criteria for DIP and the six DIP-VBTV models (see Table 4.8).

Table 4.9 and Table 4.10 report the mean and standard deviation of the PSNR of DIP and the six DIP-VBTV models over 5 runs, tested on the whole dataset with the numerical scheme (4.2) stopped after the optimal number of iterations reported in Table 4.8.

As in the previous experiment, we observe that all the DIP-VBTV models, except

TABLE 4.6

Deblurring (Gaussian): Max PSNR over one run (in parenthesis, iteration at which the max PSNR value is reached)

Algorithm	Butterfly	Leaves	Parrots	Starfish	Average
DIP	33.39 (9363)	32.37 (12317)	35.50 (10259)	35.35 (10630)	34.15 (10642)
DIP-VBTV Euclidean	33.35 (8650)	31.92 (10565)	35.88 (12731)	35.76 (12544)	34.23 (11122)
DIP-VBTV Riemannian	33.19 (6920)	32.01 (10075)	35.77 (14740)	35.86 (11608)	34.21 (10836)
DIP-VBTV (3.13)	33.37 (8686)	32.32 (11193)	**36.55** (17547)	**36.20** (13142)	34.61 (12642)
DIP-VBTV (3.18)	**34.19** (13818)	32.51 (20789)	36.01 (13146)	35.84 (11097)	**34.64** (14713)
DIP-VBTV (3.16)	33.31 (9405)	32.11 (11515)	36.40 (14207)	36.12 (11992)	34.49 (11780)
DIP-VBTV (3.20)	33.74 (13683)	**32.75** (20751)	36.02 (13487)	35.76 (11652)	34.57 (14893)

TABLE 4.7

Deblurring (Uniform): Max PSNR over one run (in parenthesis, iteration at which the max PSNR value is reached)

Algorithm	Butterfly	Leaves	Parrots	Starfish	Average
DIP	31.81 (9997)	30.57 (11910)	33.00 (10611)	31.74 (9558)	31.78 (10519)
DIP-VBTV Euclidean	31.80 (11237)	29.67 (13041)	32.67 (13478)	31.94 (13289)	31.52 (12761)
DIP-VBTV Riemannian	31.85 (11134)	30.19 (11204)	33.56 (13055)	32.12 (11591)	31.93 (11746)
DIP-VBTV (3.13)	**32.42** (11851)	30.29 (12230)	33.64 (11919)	32.42 (13055)	32.19 (12264)
DIP-VBTV (3.18)	32.16 (12063)	**30.78** (14645)	33.78 (11953)	32.24 (11159)	32.24 (12455)
DIP-VBTV (3.16)	32.31 (11172)	30.45 (12969)	**34.08** (15911)	**32.43** (12073)	**32.32** (13031)
DIP-VBTV (3.20)	32.18 (10855)	30.65 (13788)	33.59 (11365)	32.20 (11941)	32.16 (11987)

DIP-VBTV Euclidean for uniform blur, outperform DIP in average over the dataset. Moreover, focusing on the DIP-VBTV models with $\omega \neq 0$, the improvement goes from +0.35 dB to +0.49 dB for Gaussian blur and from +0.4dB to +0.5dB for uniform blur. Those results are similar to the results obtained in the previous experiment, which shows the accuracy of proposed stopping criteria. Let us also mention that DIP-VBTV (3.16) gives the most stable results for both degradation operators as it has the smallest standard deviation in average. Finally, let us mention one difference with respect to the previous experiment. In the Gaussian blur case, the best average score is now given by the model induced by the connection 1-form (3.20) whereas it was given by the model DIP-VBTV (3.18) in the previous experiment (Table 4.6).

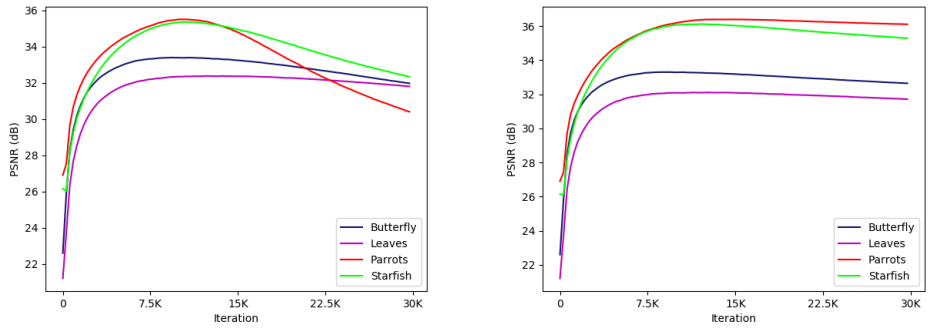

FIG. 4.4. *Evolution of the PSNR for two different models tested on the whole dataset corrupted with Gaussian blur. DIP (left) and DIP-VBTV (3.16) (right).*

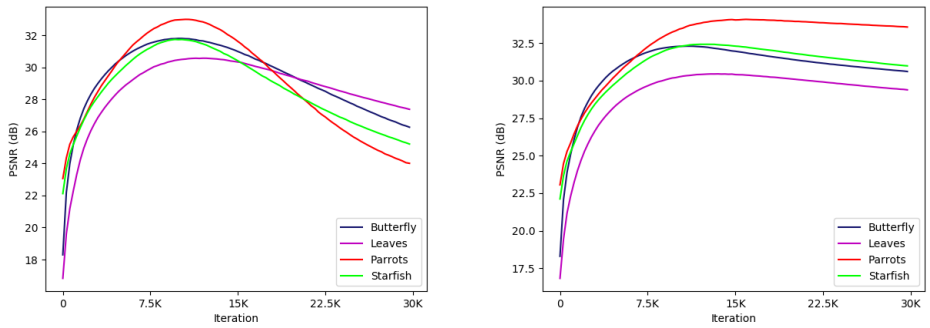

FIG. 4.5. *Evolution of the PSNR for two different models tested on the whole dataset corrupted with uniform blur. DIP (left) and DIP-VBTV (3.16) (right).*

In Fig. 4.7, we compare the images providing the best results over the 5 runs (in terms of PSNR) of DIP and DIP-VBTV (3.16) tested on the four images for uniform blur. Whereas the best PSNR for DIP are 32.01 dB for "Butterfly", 30.31 dB for "Leaves", 31.93 dB for "Starfish" and 32.94 dB for "Parrots" (second column from the right), DIP-VBTV (3.16) reaches 32.24 dB for "Butterfly", 30.36 dB for "Leaves", 32.57 dB for "Starfish" and 33.93 dB for "Parrots" (first column from the right). We observe that DIP provides much noisier images, which might explain the difference in PSNR.

4.3.3. Boosting of the results. We apply the boosting technique aforementioned by considering the mean of the output images of each model over the 5 runs. Table 4.11 and Table 4.12 report the PSNR of the mean images. We observe that the DIP-VBTV models which gave the best mean PSNR among the 6 DIP-VBTV models in Table 4.9 and Table 4.10 do not provide the best results after averaging the 5 images anymore. Indeed, the DIP-VBTV (3.18) model now gives the best results in both cases according to the results in the columns "Average".
As in the denoising case, DIP benefits more from this boosting technique than the DIP-VBTV models. Indeed, DIP has increased its average PSNR by an amount of 1.12 dB on Gaussian blur and 1.03 dB on uniform blur. On the other hand, the best

FIG. 4.6. *Evolution of the image "Parrots" at different iterations tested with two models for uniform blur: Top row: DIP - Bottom row: DIP-VBTV (3.16). From left to right: 2.5K, 5K, 10K, 30K.*

TABLE 4.8
Deblurring: Iteration at which the numerical scheme (4.2) is stopped.

Algorithm	Stopping criteria (number of iterations)
DIP	10K
DIP-VBTV for $\omega = 0$	11K
DIP-VBTV for $\omega \neq 0$	12K

increase of PSNR among the DIP-VBTV models occurs with the DIP-VBTV (3.18) model and represents 0.92 dB for Gaussian blur and 0.66 dB for uniform blur. This result makes sense as the averaging of the 5 images does reduce the noise, which is much more present in DIP results (compare the last two columns in Fig. 4.7). Hence, we intuit that we could improve the results of DIP-VBTV by considering a greater number of iterations in the numerical scheme (4.2). It would provide images with more details and noise, this latter being removed afterwards by this boosting technique.

4.3.4. Comparison to DeepRED. Because DeepRED [27], even when applied to color images, is evaluated on their luminance channel, we evaluate DIP-VBTV on the luminance channel as well. However, for some reasons we ignore, we do not obtain the results reported in [27] when computing the PSNR of the luminance channel of the input blurred images with respect to the luminance channel of the ground truth images, which makes unfair a direct comparison of the PSNR of the results of both methods. For this reason, we made the choice of comparing the increase of PSNR of the two models with respect to the PSNR of the blurred images obtained in each case. Here, we consider the luminance channel of the first result over 5 runs of the model DIP-VBTV (3.18). Results are reported in Table 4.13 for Gaussian blur and Table 4.14 for uniform blur. We observe in both cases that DIP-VBTV outperforms DeepRED in a great extent. Note also that DeepRED reports an improvement of +0.92dB on uniform blur and +0.89dB on Gaussian blur with respect to DIP. However, the authors do not mention the number of iterations used to evaluate DIP in that experiment.

TABLE 4.9
Deblurring (Gaussian): Mean PSNR and standard deviation over 5 runs

Algorithm	Butterfly	Leaves	Parrots	Starfish	Average
DIP	33.30 ± 0.119	32.06 ± 0.101	35.51 ±0.151	35.38 ± 0.073	34.06 ± 0.111
DIP- VBTV Euclidean	33.36 ± 0.11	31.82 ± 0.197	35.74 ± 0.156	35.85 ± 0.081	34.19 ± 0.136
DIP- VBTV Riemannian	32.88 ± 0.174	32.08 ± 0.208	35.87 ± 0.198	35.70 ± 0.113	34.13 ± 0.173
DIP-VBTV (3.13)	33.41 ± 0.159	31.91 ± 0.126	**36.20** ± 0.177	**36.11** ± 0.216	34.41 ± 0.17
DIP-VBTV (3.18)	33.74 ±0.238	**32.57** ± 0.117	35.88 ± **0.05**	35.70 ± 0.107	34.47 ± 0.128
DIP- VBTV (3.16)	33.57 ± **0.111**	32.19 ± **0.076**	36.15 ± 0.133	36.08 ± **0.069**	34.50 ± **0.097**
DIP- VBTV (3.20)	**33.78** ± 0.114	32.53 ± 0.19	36.06 ± 0.091	35.83 ± 0.126	**34.55** ± 0.13

TABLE 4.10
Deblurring (Uniform): Mean PSNR and standard deviation over 5 runs

Algorithm	Butterfly	Leaves	Parrots	Starfish	Average
DIP	31.89 ± 0.064	30.21 ± 0.058	32.82 ± 0.071	31.88 ± **0.038**	31.70 ± 0.058
DIP- VBTV Euclidean	31.81 ± 0.073	29.66 ± 0.045	32.56 ± 0.083	31.70 ± 0.06	31.42 ± 0.065
DIP- VBTV Riemannian	31.78 ± 0.107	30.20 ± 0.072	33.24 ± 0.161	32.21 ± 0.046	31.86 ± 0.096
DIP-VBTV (3.13)	32.01 ± 0.083	30.29 ± 0.042	33.76 ± 0.146	**32.53** ± 0.07	32.15 ± 0.085
DIP-VBTV (3.18)	32.15 ± 0.089	**30.70** ± 0.03	33.58 ± 0.146	32.21 ± 0.067	32.16 ± 0.083
DIP-VBTV (3.16)	**32.19** ± **0.038**	30.34 ± **0.02**	**33.81** ± **0.07**	32.46 ± 0.071	**32.20** ± **0.05**
DIP-VBTV (3.20)	32.14 ± 0.071	30.68 ± 0.064	33.36 ± 0.183	32.21 ± 0.104	32.10 ± 0.101

TABLE 4.11
Deblurring (Gaussian) - PSNR of the mean image over 5 runs

Algorithm	Butterfly	Leaves	Parrots	Starfish	Average
DIP	34.28	33.14	36.47	36.83	35.18
DIP-VBTV Euclidean	33.98	32.55	36.33	36.78	34.91
DIP-VBTV Riemannian	33.38	32.89	36.47	36.74	34.87
DIP-VBTV (3.13)	34.00	32.59	36.67	**36.97**	35.06
DIP-VBTV (3.18)	**34.49**	**33.47**	**36.74**	36.85	**35.39**
DIP-VBTV (3.16)	34.18	32.95	**36.74**	**36.97**	35.21
DIP-VBTV (3.20)	34.45	33.39	36.71	36.90	35.36

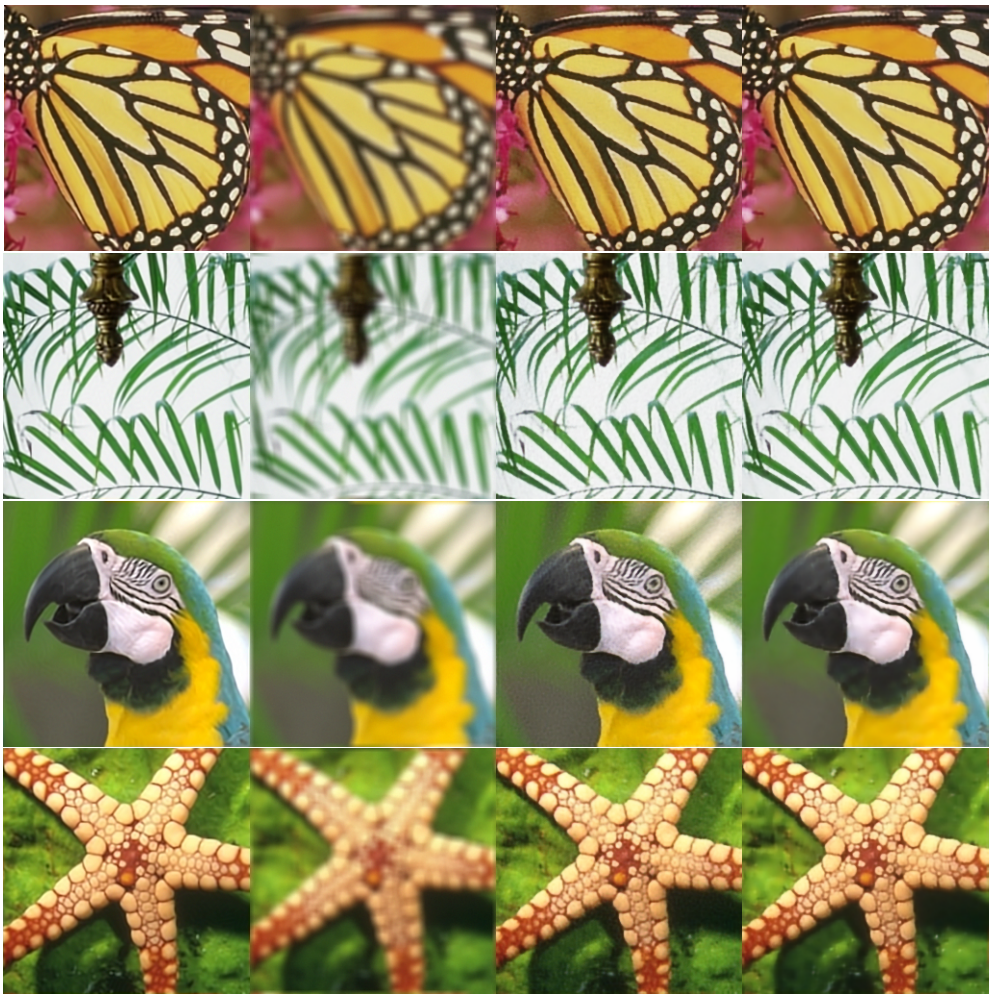

FIG. 4.7. *Comparison of DIP and DIP-VBTV (3.16) model for deblurring (uniform blur and best results, in terms of PSNR, over the 5 runs). From left to right: Clean ground truth image - Clean image corrupted by uniform blur - Result of DIP - Result of DIP-VBTV (3.16).*

TABLE 4.12
Deblurring (Uniform) - PSNR of the mean image over 5 runs

Algorithm	Butterfly	Leaves	Parrots	Starfish	Average
DIP	**32.91**	31.23	33.83	32.96	32.73
DIP-VBTV Euclidean	32.31	30.15	32.91	32.14	31.88
DIP-VBTV Riemannian	32.20	30.77	33.70	32.79	32.37
DIP-VBTV (3.13)	32.56	30.91	34.22	**33.10**	32.70
DIP-VBTV (3.18)	32.73	**31.48**	34.19	32.88	**32.82**
DIP-VBTV (3.16)	32.79	30.99	**34.31**	33.07	32.79
DIP-VBTV (3.20)	32.69	31.43	33.96	32.88	32.74

TABLE 4.13
Deblurring (Gaussian) - Comparison to DeepRED on the luminance channel by comparing the increase of PSNR with respect to the input blurred image

Algorithm	Butterfly	Leaves	Parrots	Starfish	Average
DeepRED	+ 9.38	+ 10.15	+ 5.88	+6.91	+8.08
DIP- VBTV (3.18)	+ 11.90	+11.22	+9.72	+10.34	+10.8

TABLE 4.14
Deblurring (Uniform) - Comparison to DeepRED on the luminance channel by comparing the increase of PSNR with respect to the input blurred image.

Algorithm	Butterfly	Leaves	Parrots	Starfish	Average
DeepRED	+12.37	+12.93	+8.16	+8.5	+ 10.49
DIP- VBTV (3.18)	+14.44	+13.80	+ 11.01	+10.42	+12.42

5. Conclusion. In this paper, we introduced a variational model for color image restoration which combines two priors: a Vector Bundle Total Variation (VBTV) prior determined by three geometric quantities, called geometric triplet, and a Deep Image prior (DIP) determined by a neural network. We showed that, for well-chosen geometric triplets, the minimization of VBTV can reveal some perceptual invariance of colors and local features with respect to a degradation operator. We tested our model with different geometric triplets on denoising and deblurring, and results showed that it outperforms other methods involving DIP. Finally, results show that the geometric triplet which provides the best result depends on both the image and the degradation operator, which makes us believe that our results can be improved by including some additional learning about the geometric triplet.

Appendix A. Proof of Proposition 2.1.

Proof. Denoting by A the quantity $h(D_{\partial_{x_1}}u, D_{\partial_{x_1}}u)$, B the quantity $h(D_{\partial_{x_1}}u, D_{\partial_{x_2}}u)$ and C the quantity $h(D_{\partial_{x_2}}u, D_{\partial_{x_2}}u)$, the critical points of the functional (2.7) satisfy

$$\begin{cases} \dfrac{\partial X}{\partial g_{11}} = C(g_{11}g_{22} - g_{12}^2) - \dfrac{1}{2}g_{22}(g_{22}A - 2g_{12}B + g_{11}C) = 0 \\ \dfrac{\partial X}{\partial g_{22}} = A(g_{11}g_{22} - g_{12}^2) - \dfrac{1}{2}g_{11}(g_{22}A - 2g_{12}B + g_{11}C) = 0 \\ \dfrac{\partial X}{\partial g_{12}} = B(g_{11}g_{22} - g_{12}^2) - \dfrac{1}{2}g_{12}(g_{22}A - 2g_{12}B + g_{11}C) = 0. \end{cases} \tag{A.1}$$

Reordering the terms in the first equation gives

$$g_{11} = 2\frac{g_{12}^2}{g_{22}} - 2g_{12}\frac{B}{C} + g_{22}\frac{A}{C}. \tag{A.2}$$

Substituting g_{11} according to (A.2) in the second and third equations in (A.1) yields the system

$$\begin{cases} -2g_{12}A + 6\dfrac{g_{12}^2}{g_{22}}B - 4g_{12}\dfrac{B^2}{C} + 2g_{22}\dfrac{AB}{C} - 2\dfrac{g_{12}^3}{g_{22}^2}C = 0 \\ \\ -g_{12}g_{22}A + 3g_{12}^2B - 2g_{12}g_{22}\dfrac{B^2}{C} + g_{22}^2\dfrac{AB}{C} - \dfrac{g_{12}^3}{g_{22}}C = 0. \end{cases} \tag{A.3}$$

These two equations are linearly dependent. Fixing g_{22}, we obtain an equation of the form $p(g_{12}) = 0$ where p is a polynomial of order 3, which guarantees the existence of at least one solution of this equation. Hence, the system (A.3) has an infinite number of solutions (g_{12}^*, g_{22}^*), and consequently the original (A.1) system does.

In particular, we observe that the triplet $g_{11}^* = A, g_{12}^* = B, g_{22}^* = C$ is a solution of (A.1). □

Appendix B. Proof of Proposition 3.1.

Proof. We have

$$D^v u = 0 \Longleftrightarrow \begin{cases} du_1 + u_2\left(\dfrac{v_1dv_2 - v_2dv_1}{\|v\|^2}\right) = 0 \\ \\ du_2 - u_1\left(\dfrac{v_1dv_2 - v_2dv_1}{\|v\|^2}\right) = 0 \end{cases}$$

$\Longleftrightarrow$

$$\begin{cases} u_1du_1 + u_1u_2\left(\dfrac{v_1dv_2 - v_2dv_1}{\|v\|^2}\right) = 0 \\ u_2du_1 + (u_2)^2\left(\dfrac{v_1dv_2 - v_2dv_1}{\|v\|^2}\right) = 0 \\ u_1du_2 - (u_1)^2\left(\dfrac{v_1dv_2 - v_2dv_1}{\|v\|^2}\right) = 0 \\ u_2du_2 - u_2u_1\left(\dfrac{v_1dv_2 - v_2dv_1}{\|v\|^2}\right) = 0. \end{cases} \tag{B.1}$$

Summing the first and fourth equations in (B.1) yields

$$u_1du_1 + u_2du_2 = 0,$$

i.e. $d\|u\|^2 = 0$, which implies that $d\|u\| = 0$.
Subtracting the third equation from the second equation in (B.1) gives

$$u_2du_1 - u_1du_2 + \|u\|^2\left(\frac{v_1dv_2 - v_2dv_1}{\|v\|^2}\right) = 0,$$

i.e.

$$\frac{u_1du_2 - u_2du_1}{\|u\|^2} = \frac{v_1dv_2 - v_2dv_1}{\|v\|^2},$$

which proves that

$$D^v u = 0 \Longrightarrow \begin{cases} d\|u\| = 0 \\ \dfrac{u_1du_2 - u_2du_1}{\|u\|^2} = \dfrac{v_1dv_2 - v_2dv_1}{\|v\|^2}. \end{cases}$$

On the other hand, assuming that

$$\frac{u_1du_2 - u_2du_1}{\|u\|^2} = \frac{v_1dv_2 - v_2dv_1}{\|v\|^2},$$

i.e. $\omega^v = \omega^u$, we have $D^v u = D^u u$. Then,

$$d\|u\| = 0 \Longrightarrow D^u u = 0$$

according to (3.7), and it leads to $D^v u = 0$ as $D^v u = D^u u$. □

Appendix C. Proof of Proposition 3.2.

Proof. We have $D^v u \Longleftrightarrow$

$$\begin{cases} du_1 + u_2\left(\dfrac{v_1dv_2 - v_2dv_1}{\|v\|^2}\right) + u_3\left(\dfrac{v_1dv_3 - v_3dv_1}{\|v\|^2}\right) = 0 \\ du_2 - u_1\left(\dfrac{v_1dv_2 - v_2dv_1}{\|v\|^2}\right) + u_3\left(\dfrac{v_2dv_3 - v_3dv_2}{\|v\|^2}\right) = 0 \\ du_3 - u_1\left(\dfrac{v_1dv_3 - v_3dv_1}{\|v\|^2}\right) - u_2\left(\dfrac{v_2dv_3 - v_3dv_2}{\|v\|^2}\right) = 0 \end{cases} \tag{C.1}$$

Denoting by (a),(b),(c) the three equalities from top to bottom in (C.1), we have that $u_1(a) + u_2(b) + u_3(c)$ yields

$$u_1du_1 + u_2du_2 + u_3du_3 = 0, \tag{C.2}$$

i.e. $d\|u\|^2 = 0$ and consequently $d\|u\| = 0$.

Then, $u_2(u_2(a) - u_1(b)) + u_3(u_3(a) - u_1(c))$ yields

$$u_2\frac{(u_1du_2 - u_2du_1)}{\|u\|^2} + u_3\frac{(u_1du_3 - u_3du_1)}{\|u\|^2} = u_2\frac{(v_1dv_2 - v_2dv_1)}{\|v\|^2} + u_3\frac{(v_1dv_3 - v_3dv_1)}{\|v\|^2}. \tag{C.3}$$

Moreover, $u_1(u_2(a) - u_1(b)) - u_3(u_3(b) - u_2(c))$ yields

$$u_1\frac{(u_1du_2 - u_2du_1)}{\|u\|^2} + u_3\frac{(u_3du_2 - u_2du_3)}{\|u\|^2} = u_1\frac{(v_1dv_2 - v_2dv_1)}{\|v\|^2} + u_3\frac{(v_3dv_2 - v_2dv_3)}{\|v\|^2}. \tag{C.4}$$

Finally, $u_1(u_3(a) - u_1(c)) + u_2(u_3(b) - u_2(c))$ yields

$$u_1\frac{(u_1du_3 - u_3du_1)}{\|u\|^2} + u_2\frac{(u_2du_3 - u_3du_2)}{\|u\|^2} = u_1\frac{(v_1dv_3 - v_3dv_1)}{\|v\|^2} + u_2\frac{(v_2dv_3 - v_3dv_2)}{\|v\|^2}. \tag{C.5}$$

Then, (C.2) together with (C.3),(C.4),(C.5) give

$$D^v u = 0 \Longrightarrow \begin{cases} d\|u\| = 0 \\ \omega^u u = \omega^v u. \end{cases}$$

On the other hand, assuming that $\omega^u u = \omega^v u$, we have $D^u u = D^v u$. Then, by (3.11), we have $d\|u\| = 0 \Longrightarrow D^u u = 0$ and consequently $D^v u = 0$. □

Appendix D. Proof of Proposition 3.3.

Proof. Let G be a Lie group and $\mathfrak{g}$ its Lie algebra. Recall that a connection 1-form on a G-associated bundle is $\mathfrak{g}$-valued, and the curvature $F(\omega)$ of a connection 1-form ω is given by $d\omega + \frac{1}{2}[\omega, \omega]$, where d stands for the exterior derivative and $[\cdot, \cdot]$ stands for the Lie algebra wedge product on $\mathfrak{g}$.
The connection 1-form ω in formula (3.16) is $\mathfrak{gl}(3, \mathbb{R})$-valued. This Lie algebra is generated by the single entry matrices $E_{ij}, i, j = 1, 2, 3$, which are equipped with the following Lie bracket

$$[E_{ij}, E_{kl}] = \delta_{jk} E_{il} - \delta_{il} E_{kj}, \tag{D.1}$$

where δ is the Dirac delta function. It gives

$$[\omega, \omega] = \sum_{i,j,k,l=1}^{3} \omega_{ij} \wedge \omega_{kl} \otimes [E_{ij}, E_{kl}].$$

In order to show that $F(\omega) \neq 0$, we show that $F(\omega)_{12} \neq 0$ in what follows.
Let us denote by $\omega_{\mathrm{SO}(3)}$ the connection 1-form described in formula (3.10). From the fact that $\omega_{12} = \omega_{\mathrm{SO}(3)_{12}}$, we have

$$(d\omega)_{12} = (d\omega_{\mathrm{SO}(3)})_{12}. \tag{D.2}$$

Let us now consider the term $[\omega, \omega]_{12}$. According to (D.1), it is given by

$$\frac{1}{2}[\omega, \omega]_{12} = \omega_{11} \wedge \omega_{12} + \omega_{12} \wedge \omega_{22} + \omega_{13} \wedge \omega_{32}. \tag{D.3}$$

On the other hand, we have

$$\frac{1}{2}[\omega_{\mathrm{SO}(3)}, \omega_{\mathrm{SO}(3)}]_{12} = \omega_{13} \wedge \omega_{32}$$

by property of $\mathfrak{so}(3)$, the set of 3×3 skew-symmetric matrices.
Hence, $(d\omega_{\mathrm{SO}(3)})_{12} = -\omega_{13} \wedge \omega_{32}$ since $F(\omega_{\mathrm{SO}(3)})_{12} = 0$ (it has been shown in [5] that $F(\omega_{\mathrm{SO}(3)}) \equiv 0$), and we deduce from (D.2) that

$$(d\omega)_{12} = -\omega_{13} \wedge \omega_{32}. \tag{D.4}$$

(D.4) together with (D.3) give that

$$F(\omega)_{12} = \omega_{11} \wedge \omega_{12} + \omega_{12} \wedge \omega_{22}.$$

Finally, a straightforward computation yields

$$\omega_{11} \wedge \omega_{12} + \omega_{12} \wedge \omega_{22} \neq 0$$

□

REFERENCES

[1] D. Barbieri, G. Citti, G. Cocci and A. Sarti, *A cortical-inspired geometry for contour perception and motion integration*, J. Math. Imag. Vis., 49(3) (2014), pp. 511–529.

[2] T. Batard and M. Berthier, *Spinor Fourier transform for image processing*, IEEE J. Selected Topics in Sign. Proces., 7(4) (2013), pp. 605–613.

[3] T. Batard and N. Sochen, *A class of generalized Laplacians devoted to multi-channels image processing*, J. Math. Imag. Vis., 48(3) (2014), pp. 517–543.

[4] T. Batard and M. Bertalmío, *On covariant derivatives and their applications to image regularization*, SIAM J. Imag. Sci., 7(4) (2014), pp. 2393–2422.

[5] T. Batard and M. Bertalmío, *A geometric model of brightness perception and its application to color images correction*, J. Math. Imag. Vis., 60(6) (2018), pp. 849–881.

[6] T. Batard, J. Hertrich and G. Steidl, *Variational models for color image correction inspired by visual perception and neuroscience*, J. Math. Imag. Vis., 62(9) (2020), pp. 1173–1194.

[7] M. Bertalmío, *Vision Models for High Dynamic Range and Wide Colour Gamut Imaging: Techniques and Applications*, Academic Press, 2019.

[8] M. Bertalmio, L. Calatroni, V. Franceschi, B. Franceschiello and D. Prandi, *Cortical-inspired Wilson–Cowan-type equations for orientation-dependent contrast perception modelling*, J. Math. Imag. Vis., (2020), pp. 1–19.

[9] P. Blomgren and T.F. Chan, *Total variation methods for restoration of vector-valued images*, IEEE Trans. Im. Proces., 7(3) (1998), pp. 304–309.

[10] U. Boscain, R. Chertovskih, J-P. Gauthier, D. Prandi and A. Remizov, *Cortical-inspired image reconstruction via sub-Riemannian geometry and hypoelliptic diffusion*, ESAIM: Proceedings and Surveys, 64 (2018), pp. 37–53.

[11] X. Bresson and T.F. Chan, *Fast dual minimization of the vectorial total variation norm and applications to color image processing*, Inverse problems and imaging, 2(4) (2008), pp. 455–484.

[12] A. Buades, B. Coll and J.-M. Morel, *A non-local algorithm for image denoising*, Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., 2 (2005), pp. 60–65.

[13] A. Buades, B. Coll and J.-M. Morel, *Non-Local means denoising*, Image Processing On Line, 1 (2011), pp. 208–212.

[14] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Im. Vis., 20(1-2) (2004), pp. 89–97.

[15] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imag. Vis., 40 (2011), pp. 120-145.

[16] Z. Cheng, M. Gadelha, S. Maji and D. Sheldon, *A Bayesian perspective on the deep image prior*, Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog., (2019).

[17] P. Chossat and O. Faugeras, *Hyperbolic planforms in relation to visual edges and textures perception*, PLoS Comput Biol., 5(12) (2009), e1000625.

[18] G. Citti, B. Franceschiello, G. Sanguinetti and A. Sarti, *Sub-Riemannian mean curvature flow for image processing*, SIAM J. Imag. Sci., 9(1) (2016), pp. 212–237.

[19] K. Dabov, A. Foi, V. Katkovnik and K. Egiazarian, *Image denoising by sparse 3D transform-domain collaborative filtering*, IEEE Trans. Image Process., 16(8) (2007), pp. 2080-2095.

[20] M.D. Fairchild and E. Pirrotta, *Predicting the lightness of chromatic object colors using CIELAB*, Color Research and Applications, 16(6) (1991), pp. 385–393.

[21] W. Förstner and E. Gülch, *A fast operator for detection and precise location of distinct points, corners and centres of circular features*, Proc. ISPRS Intercom-mission Conference on Fast Processing of Photogrammetric Data, (1987), pp. 281–305.

[22] T. Georgiev, *Relighting, Retinex theory, and perceived gradients*, Proceedings of MIRAGE 2005.

[23] T. Georgiev, *Image reconstruction invariant to relighting*, Proceedings of EUROGRAPHICS 2005.

[24] Z. Jia, M.K. Ng and W. Wang, *Color image restoration by saturation-value total variation*, SIAM J. Imag. Sci., 12(2) (2019), pp. 972–1000.

[25] M. Lebrun, *An analysis and implementation of the BM3D image denoising method*, Image Processing On Line, 2 (2012), pp. 175–213.

[26] J. Liu, Y. Sun, X. Xu and U.S. Kamilov, *Image restoration using total variation regularized deep image prior*, Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, (2019), pp. 7715–7719.

[27] G. Mataev, M. Elad and P. Milanfar, *DeepRED: Deep Image Prior Powered by RED*, Proc. ICCV 2019 workshop on Learning for Computational Imaging.

[28] L.I. Rudin, S. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60(1-4) (1992), pp. 259–268.
[29] J. Shen, *On the foundations of vision modeling: I. Weber's law and Weberized TV restoration*, Physica D: Non linear Phenomena, 175(3-4) (2003), pp. 241–251.
[30] N. Sochen, R. Kimmel and R. Malladi, *A general framework for low level vision*, IEEE Trans. Im. Proces., 7(3) (1998), pp. 310–318.
[31] D. Ulyanov, A. Vedaldi and V. Lempitsky, *Deep image prior*, Int. J. Comput. Vis., 128 (2020), pp. 1867-1888.
[32] L.A. Vese and C. Le Guyader, *Variational methods in image processing*, Chapman and Hall/CRC (2015).
[33] H.R. Wilson and J.D. Cowan, *Excitatory and inhibitory interactions in localized populations of model neurons*, Biophys. J., 12(1) (1972), pp. 1–24.
[34] H.R. Wilson and J.D. Cowan, *A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue*, Biol. Cybernet., 13(2) (1973), pp. 55-80.
[35] Q. Zhou, C. Zhou, H. Hu, Y. Chen, S. Chen and X. Li, *Towards the automation of Deep Image Prior*, *ArXiv: 1911.07185* (2019).