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

Thomas Batard, Gloria Haro, Coloma Ballester

DIP-VBTV: A COLOR IMAGE RESTORATION MODEL COMBINING A DEEP IMAGE PRIOR AND A VECTOR BUNDLE TOTAL VARIATION

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

Abstract. In this paper, we introduce a new variational model for color image restoration, called DIP-VBTV, which combines two priors: a deep image prior (DIP), which assumes that the restored image can be generated through a neural network and a Vector Bundle Total Variation (VBTV) which generalizes the Vectorial Total Variation (VTV) on vector bundles. VBTV is determined by a geometric triplet: a Riemannian metric on the base manifold, a covariant derivative and a metric on the vector bundle. Whereas VTV prior encourages the restored images to be piece-wise constant, VBTV prior encourages them to be piece-wise parallel with respect to a covariant derivative. For well-chosen geometric triplets, we show that the minimization of VBTV encourages the solutions of the restoration model to share some visual content with the clean image. Then, we show on experiments that DIP-VBTV benefits from this property by outperforming DIP-VTV and state-of-the-art unsupervised methods. It demonstrates the relevance of combining DIP and VBTV priors.

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],[3],[4],[6],[7],[9],[17],[25]). 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. This 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 level is not too high (this is the case in realistic situations). On the other hand, when the degradation comes from a blurring operator, local structures 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 colors, but also be able to adapt the invariance to a given degradation operator.

1.2. Related work.

1.2.1. VBTV priors to express perceptual invariance. Over the last 30 years, variational models have demonstrated their efficiency to tackle several tasks in image restoration, e.g. denoising, deblurring, inpainting, super-resolution, etc (see e.g. [28] and references therein). They are often expressed as a convex combination of

* 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)

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 for instance the Total Variation (TV) [13],[14],[24] whose minimization encourages the preservation of local structures by means of the L^1 norm of the Euclidean gradient. A second example is the Polyakov action [26] whose minimization encourages the preservation of local structures by means of the L^2 norm of a Riemannian gradient, the Riemannian metric being related to the structure tensor of the image [19]. These two penalty terms can be extended to color images in a 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) [8],[10].

A more perceptually-based color extension of TV is the Saturation-Value Total Variation (SVTV) [20], which takes into account that the spatial variations of the local structures of a natural image are mainly in its achromatic component. Then, SVTV penalizes the smoothing of the achromatic component of the image, and consequently of its local structures. This makes SVTV be a better prior for color image denoising than VTV.

In [2], a new geometric setting for imaging has been proposed, in which a color image is considered as a section of a vector bundle. In this context, the Vector Bundle Total Variation (VBTV) arises as the natural extension of VTV, and is defined by $\mathrm{VBTV}(u) = \|Du\|_{L^1(g^{-1}\otimes h)}$ for differentiable sections u. Here, g stands for a Riemannian metric on the base manifold, D is a covariant derivative determined by a connection 1-form ω, and h is a definite positive metric on the vector bundle (the explicit expression of $\mathrm{VBTV}(u)$ is given in Sect. 2.1.3). Hence, VBTV is determined by the geometric triplet g, h, ω. Then, the authors considered a particular geometric triplet which encodes some invariance of the local structures under a degradation by noise. Experiments showed that this VBTV is a better prior for denoising than the standard VTV in the sense that it provides better restored images (higher PSNR and SSIM). More recently, this approach has been coupled with SVTV, yielding a VBTV encoding that local structures are mainly in the achromatic component, and providing even better results [29].

Besides denoising, these priors/penalty terms have also been applied to various image restoration problems such as deblurring, inpainting, super-resolution [14]. Whereas they do encode some perceptual invariance with respect to a degradation by noise, they do not encode any perceptual invariance with respect to other degradations. Then, we claim that an image restoration model would benefit from the consideration of degradation-based penalty terms.

1.2.2. Deep Image Prior. In the variational models for image restoration aforementioned, the minimization is performed on a space of functions or sections having bounded variations. Recently, a new prior has been introduced for image restoration, called Deep Image Prior (DIP) [27]. In this framework, the minimization is performed on a set of functions generated by a well-chosen neural network. More precisely, the following minimization problem has been introduced

$$\begin{cases} \underline{\theta} = \arg\min_{\theta} \dfrac{1}{2}\|H(T_\theta(z)) - v\|_{L^2}^2 \\ \underline{u} = T_{\underline{\theta}}(z), \end{cases} \tag{1.1}$$

where T_θ is a neural network parametrized by θ whose input z is a random multi-channel image, v is the observed degraded image, H is a degradation operator, and $\underline{u}$ is the restored image. Experiments showed that model (1.1) outperforms standard VTV-based restoration models in a great extent on denoising and super-resolution.

More recently, DIP has been combined to an (anisotropic) TV in [22] yielding the so-called model DIP-TV, given by

$$\begin{cases} \underline{\theta} = \arg\min_{\theta} \dfrac{1}{2}\|H(T_\theta(z)) - v\|_{L^2}^2 + \lambda\,\mathrm{TV}(\mathrm{T}_\theta(\mathrm{z})) \qquad \lambda > 0 \\ \underline{u} = T_{\underline{\theta}}(z), \end{cases} \tag{1.2}$$

Experiments showed that DIP-TV outperforms DIP on denoising and deblurring.

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

1.3.1. Construction of an optimal geometric triplet. Given a color image $u = (u_1, u_2, u_3)\colon \Omega \subset \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ seen as a section of a vector bundle, we consider the differentiable energy

$$X(g, \omega, h) = \|Du\|^2_{L^2(g^{-1}\otimes h)}, \tag{1.3}$$

and determine some of its critical points in Sect. 2.

Optimal Riemannian metric. Fixing ω and h, we show that, for $\epsilon > 0$ small, the Riemannian metric g given by

$$g = \begin{pmatrix} \epsilon + h(D_{\partial_{x_1}} u, D_{\partial_{x_1}} u) & h(D_{\partial_{x_1}} u, D_{\partial_{x_2}} u) \\ h(D_{\partial_{x_1}} u, D_{\partial_{x_2}} u) & \epsilon + h(D_{\partial_{x_2}} u, D_{\partial_{x_2}} u) \end{pmatrix} \tag{1.4}$$

in the frame $(\partial_{x_1}, \partial_{x_2})$ induced by the Cartesian coordinates system (x_1, x_2) on Ω, approximates a critical point of the energy (1.3).

Optimal connection 1-form. Fixing g and h, and assuming that ω is $\mathbb{R} \times \mathfrak{so}(2)-$valued, we show that the energy (1.3) possesses a unique critical point, given by

$$\omega = \begin{pmatrix} -\dfrac{du_1}{u_1} & 0 & 0 \\ 0 & 0 & \dfrac{u_2 du_3 - u_3 du_2}{u_2^2 + u_3^2} \\ 0 & -\dfrac{u_2 du_3 - u_3 du_2}{u_2^2 + u_3^2} & 0 \end{pmatrix} \tag{1.5}$$

On the existence of optimal vector bundle metrics. Fixing g and ω, we show that the energy (1.3) does not possess critical points.

1.3.2. Perceptual invariance associated to the minimization of VBTV induced by well-chosen geometric triplets. A vector bundle metric can be used to assign different weights to the different image components. Then, through the minimization of VBTV, a vector bundle metric enables for instance to process some image components in a smaller extent than others. This can be a desirable property in the context of image denoising. Indeed, by smoothing the achromatic component of the noisy image less than its chromatic components, the minimization of VBTV

TABLE 1.1
Geometric triplets for denoising and deblurring

Degradation	Color space	ω	h	g
Noise	Achromatic-Chromatic space	0	$diag(\alpha\beta, \beta, \beta)$	(2.9)
Blur	Achromatic-Chromatic space	(1.5)	$diag(\beta, \beta, \beta)$	(2.9)
Blur	Achromatic-Chromatic space	Dual of (1.5)	$diag(\beta, \beta, \beta)$	(2.9)

encourages the local structures of the restored image to be similar to the ones of the degraded image, and consequently similar to the ones of the clean image. This is exactly the purpose of the SVTV prior aforementioned.

The minimization of VBTV encourages the generation of images which are piecewise parallel with respect to the corresponding covariant derivative. In Sect. 2.3.1, we show that, under the assumption that u is a blurred image expressed in a well-chosen achromatic-chromatic space (e.g. opponent space), the parallel sections of the covariant derivative induced by the optimal connection 1-form (1.5) share some perceptual content with both u and the clean original image. As a consequence, the minimization of VBTV encourages the restored image to share some perceptual content with the clean image. We also show that the dual of the connection 1-form (1.5), defined in Sect. 2.3.2, satisfies similar properties.

A Riemannian metric of the form (1.4) is a generalization to vector bundles of the Riemannian metric used in the Beltrami framework [26]. This latter approximates the structure tensor (at scale 0) of the image, which is known to provide some information about its local structures. We show in the experiments conducted in this paper that image restoration benefits from the use of a Riemannian metric (1.4).

Based on this analysis, we consider in this paper the geometric triplets described in Table 1.1, for well-chosen $\beta > 0$, $0 < \alpha < 1$.

1.3.3. A variational model for color image restoration combining DIP and VBTV priors. 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, we consider VBTV as a penalty term of a variational problem for image restoration, yielding the so-called model DIP-VBTV

$$\begin{cases} \underline{\theta} = \arg\min_{\theta} \dfrac{1}{2}\|H(T_\theta(z)) - v\|^2_{L^2(h)} + \lambda\,\mathrm{VBTV}(\mathrm{T}_\theta(\mathrm{z})) \qquad \lambda > 0 \\ \underline{u} = T_{\underline{\theta}}(z). \end{cases} \tag{1.6}$$

It generalizes both DIP (1.1) and DIP-TV (1.2).

In Sect. 3.2 and Sect. 3.4, we test DIP-VBTV for the restoration of color images corrupted with additive white Gaussian noise. We show that DIP-VBTV with the geometric triplet for denoising described in Table 1.1, outperforms other DIP-based models: DIP, DIP-VTV, DIP-SVTV. We also show that, when combined with a boosting technique introduced in [27], DIP-VBTV outperforms DeepRED [23] and gets similar results as C-BM3D [18],[21].

In Sect. 3.3 and Sect. 3.4, we test DIP-VBTV for the restoration of color images corrupted with Gaussian blur. We show that DIP-VBTV, for VBTV being induced by the geometric triplets for deblurring described in Table 1.1, outperforms other DIP-based models: DIP, DIP-VTV, DIP-SVTV.

The codes are available at `https://github.com/tombatard/dip_vbtv`.

2. Construction of geometric triplets for color image restoration.

2.1. The notion of geometric triplet on a G-associated bundle and the induced Vector Bundle Total Variation.

2.1.1. Color image as a section of a G-associated bundle. We first recall the correspondence between sections of G-associated bundles and G-equivariant functions on principal bundles.

DEFINITION 2.1. *A smooth* ***principal bundle*** *is a quadruplet (P, π, M, G) where M and P are two C^∞ manifolds, G is a Lie group, $\pi\colon P \longrightarrow M$ is a surjective map such that for all $x \in M$, the preimage $\pi^{-1}(x)$ is diffeomorphic to G and there is an action $\cdot$ of G on P satisfying:*
- $\pi(p \cdot g) = \pi(p)$ for $p \in \pi^{-1}(x)$ and $g \in G$.
- the restriction $\cdot\colon G \times \pi^{-1}(x) \longrightarrow \pi^{-1}(x)$ is free and transitive.

M is called the ***base manifold****, P the* ***total space*** *and G the* ***structure Lie group*** *of the principal bundle. The set $\pi^{-1}(x)$ is called the* ***fiber*** *over x, and is denoted by P_x.*

DEFINITION 2.2. *Let G be a Lie group and V a vector space. A* ***representation*** *ρ of G on V is a group morphism $\rho\colon G \longrightarrow GL(V)$, where $GL(V)$ denotes the group of invertible endomorphisms of V.*

DEFINITION 2.3. *Let (P, π, M, G) be a principal bundle and ρ a representation of G on a finite dimensional vector space V. A function $J\colon P \longrightarrow V$ is called* ***G-equivariant*** *if it satisfies*

$$J(p \cdot g) = \rho(g^{-1})J(p).$$

We denote by $C^\infty(P, V)^G$ the set of smooth G-equivariant functions on P.

DEFINITION 2.4. *Let (P, π, M, G) be a principal bundle, ρ a representation of G on V of dimension n, and $E = (P \times V)/G$, i.e. a point in E is of the form*

$$[p, f] := \{(p \cdot g, \rho(g^{-1})f)\,,\ g \in G\}$$

where $p \in P$ and $f \in V$. Let $\pi_E\colon E \longrightarrow M$ given by $\pi_E[p, f] = \pi(p)$. Then, the triplet (E, π_E, M) forms a vector bundle of rank n, called ***G-associated bundle*** *and denoted by $P \times_{(\rho,G)} V$.*

There is a correspondence between sections of associated bundles and G-equivariant functions on principal bundles. Indeed, given $f \in C^\infty(P, V)^G$, we put $S_f(x) = [p, f(p)]$ for any $p \in \pi^{-1}(x)$. By the G-equivariance property of f, the map S_f is independent of the choice of p. Hence, S_f is a section of $E = P \times_{(\rho,G)} V$. Conversely, for S being a section of E, we put $f_S(p) = v$ such that $S \circ \pi(p) = [p, v]$. We observe that f_S is G-equivariant.

In what follows, we denote by TM the tangent bundle of M, by T^*M the cotangent bundle of M, and by $\mathrm{G}(E)$ the bundle of linear maps acting on the fibers of E given by matrices in the group G. We denote by $\mathfrak{g}$ the Lie algebra of G and by $\mathfrak{g}(E)$ the bundle of linear maps acting on the fibers of E given by matrices in the set $\mathfrak{g}$. Given a bundle F, we denote by $\Gamma(F)$ the set of smooth sections of F.

Let $\Omega \subset \mathbb{R}^2$ and $u\colon \Omega \longrightarrow \mathbb{R}^3$ be a color image. Let G be a Lie group acting on $\mathbb{R}^3$ through a representation ρ. Let $P = \Omega \times G$ and (P, π, Ω, G) be a principal bundle. Let E be the G-associated bundle $P \times_{(\rho,G)} \mathbb{R}^3$. In this paper, we extend u to a section of E or equivalently to a G-equivariant function on P of the form

$$u(x, g) = \rho(g^{-1})u(x) \qquad \forall x \in \Omega, \forall g \in G.$$

2.1.2. Geometric triplet on a G-associated bundle. A **geometric triplet** on a G-associated bundle over a manifold M is a triplet (g, h, ω) where:
- g is a **Riemannian metric on the base manifold**: A Riemannian metric on a manifold is a positive definite metric on its tangent bundle.
- h is a **positive definite metric on the bundle**: A positive definite metric h on a vector bundle E is the assignment of a positive definite scalar product h_x on each fiber $\pi_E^{-1}(x)$.
- ω is a **connection 1-form on the bundle**: A connection 1-form is an element of the set $\Gamma(T^*M \otimes \mathfrak{g}(E))$ which satisfies a certain transformation law under a moving frame change. More precisely, let ω be the expression of a connection 1-form in a moving frame, and $\mathcal{G}$ be another moving frame. Then, the expression of ω in the frame $\mathcal{G}$ is given by

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

where d stands for the standard differential operator.

Note that by formula (2.1), a connection 1-form is completely determined by its value in a moving frame.

2.1.3. Covariant derivative and Vector Bundle Total Variation (VBTV) induced by a geometric triplet. A covariant derivative on a G-associated bundle E is a differential operator $D \coloneqq d + \omega$, where d stands for the standard differential operator, and ω is a connection 1-form.
The transformation law (2.1) makes D satisfy a G-equivariance property with respect to a moving frame change, i.e.

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

for $\psi \in \Gamma(E)$.
A Riemannian metric g on TM and a positive definite metric h on E determine a positive definite metric $g^{-1} \otimes h$ on $T^*M \otimes E$ and an L^p norm on $\Gamma(T^*M \otimes E)$. In particular, we have

$$\|D\psi\|_{L^p(g^{-1}\otimes h)} = \left(\int_M \sum_{i,j=1}^{m} (g^{ij} h(D_{\partial_{x_i}}\psi, D_{\partial_{x_j}}\psi))^{p/2} dM \right)^{1/p},$$

where $(\partial_{x_1}, \cdots, \partial_{x_m})$ is the basis of TM induced by a coordinates system $(x_1, x_2, \cdots, x_m)$ and g^{ij} denotes the coefficients of the inverse matrix of g in the frame $(\partial_{x_1}, \cdots, \partial_{x_m})$.

Definition 2.5. *The Vector Bundle Total Variation (VBTV) of $\psi \in \Gamma(E)$ is the quantity*

$$VBTV(\psi) = \|D\psi\|_{L^1(g^{-1}\otimes h)}. \tag{2.3}$$

We denote by $BV(E)$ the set of sections ψ such that $VBTV(\psi) < \infty$.

Remark: Unlike [2], we do not require the covariant derivative to be compatible with the vector bundle metric h in the definition of $VBTV$.

2.2. An optimal Riemannian metric generalizing the structure tensor. Let u be a color image seen as a section of a G-associated bundle equipped with a connection 1-form ω and a definite positive vector bundle metric h. We consider the following energy

$$X(g) = \|Du\|^2_{L^2(g^{-1}\otimes h)}. \tag{2.4}$$

We have the following result.

PROPOSITION 2.6. *If u satisfies that $D_{\partial_{x_1}}u(x) \neq \alpha D_{\partial_{x_2}}u(x)\ \forall \alpha \in \mathbb{R}, \forall x \in \Omega$, then the Riemannian metric*

$$\begin{pmatrix} h(D_{\partial_{x_1}}u, D_{\partial_{x_1}}u) & h(D_{\partial_{x_1}}u, D_{\partial_{x_2}}u) \\ h(D_{\partial_{x_1}}u, D_{\partial_{x_2}}u) & h(D_{\partial_{x_2}}u, D_{\partial_{x_2}}u) \end{pmatrix} \tag{2.5}$$

is a critical point of the energy (2.4).

Proof. Denoting by A the quantity $h(D_{\partial_{x_1}}u, D_{\partial_{x_1}}u)$, B the quantity $h(D_{\partial_{x_1}}u, D_{\partial_{x_2}}u)$, C the quantity $h(D_{\partial_{x_2}}u, D_{\partial_{x_2}}u)$, and g_{ij} the coefficients of the matrix g, the critical points of the functional (2.4) satisfy

$$\begin{cases} \dfrac{\partial X}{\partial g_{11}} = C(g_{11}g_{22} - g_{12}^2) - \dfrac{1}{2}g_{22}(g_{22}A - 2g_{12}B + g_{11}C) = 0 \\ \dfrac{\partial X}{\partial g_{22}} = A(g_{11}g_{22} - g_{12}^2) - \dfrac{1}{2}g_{11}(g_{22}A - 2g_{12}B + g_{11}C) = 0 \\ \dfrac{\partial X}{\partial g_{12}} = B(g_{11}g_{22} - g_{12}^2) - \dfrac{1}{2}g_{12}(g_{22}A - 2g_{12}B + g_{11}C) = 0. \end{cases} \tag{2.6}$$

Reordering the terms in the first equation gives

$$g_{11} = 2\frac{g_{12}^2}{g_{22}} - 2g_{12}\frac{B}{C} + g_{22}\frac{A}{C}. \tag{2.7}$$

Note that $g_{22} \neq 0$ and $C \neq 0$ by positive definiteness of g and h, respectively. Substituting g_{11} according to (2.7) in the second and third equations in (2.6) yields

$$\begin{cases} -2g_{12}A + 6\dfrac{g_{12}^2}{g_{22}}B - 4g_{12}\dfrac{B^2}{C} + 2g_{22}\dfrac{AB}{C} - 2\dfrac{g_{12}^3}{g_{22}^2}C = 0 \\ -g_{12}g_{22}A + 3g_{12}^2B - 2g_{12}g_{22}\dfrac{B^2}{C} + g_{22}^2\dfrac{AB}{C} - \dfrac{g_{12}^3}{g_{22}}C = 0. \end{cases} \tag{2.8}$$

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 (2.8) has an infinite number of solutions (g_{12}^*, g_{22}^*), and consequently the original (2.6) system does.

In particular, we observe that the triplet $g^*_{11} = A, g^*_{12} = B, g^*_{22} = C$ is a solution of (2.6). Finally, the assumption $D_{\partial_{x_1}} u(x) \neq \alpha D_{\partial_{x_2}} u(x)\ \forall \alpha \in \mathbb{R}, \forall x \in \Omega$ guarantees that the matrix field (2.5) is positive definite on Ω. □

For h represented by $\mathbb{I}_3$, the 3x3 Identity matrix, and D given by the connection 1-form $\omega \equiv 0$ in the RGB color space, the optimal metric (2.5) corresponds to the structure tensor of the image u at scale 0.

In practice, it is likely that u satisfies $D_{\partial_{x_1}} u(x) = \alpha D_{\partial_{x_2}} u(x)$ for some $\alpha \in \mathbb{R}$ at some points x of the domain Ω, which makes the metric (2.5) be singular. In order to overcome this issue, we consider the following approximation of the metric (2.5)

$$g = \begin{pmatrix} \epsilon + h(D_{\partial_{x_1}} u, D_{\partial_{x_1}} u) & h(D_{\partial_{x_1}} u, D_{\partial_{x_2}} u) \\ h(D_{\partial_{x_1}} u, D_{\partial_{x_2}} u) & \epsilon + h(D_{\partial_{x_2}} u, D_{\partial_{x_2}} u) \end{pmatrix} \tag{2.9}$$

for $\epsilon > 0$ small, which endows (Ω, g) with a Riemannian manifold structure for any u.

2.3. An optimal connection 1-form on a $\mathbb{R}^{+*}\times$ SO(2)-associated bundle and its interpretation in color imaging.

2.3.1. The optimal connection 1-form and the parallel sections of the corresponding covariant derivative. Let u be a color image seen as a section of a G-associated bundle equipped with a Riemannian metric g and a positive definite vector bundle metric h. Without loss of generality, we assume that u is expressed in a moving frame in which h is the Euclidean metric $\| \|_2$. We consider the energy

$$X(\omega) = \|Du\|^2_{L^2(g^{-1} \otimes \| \|_2)}. \tag{2.10}$$

In this Section, we assume that the Lie group representation (ρ, G) is $\mathbb{R}^{+*}\times$ SO(2) acting on $\mathbb{R}^3$ through the representation

$$\rho(k, \theta) \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} k u_1 \\ \cos\theta u_2 + \sin\theta u_3 \\ -\sin\theta u_3 + \cos\theta u_2 \end{pmatrix}, \tag{2.11}$$

where $u = (u_1, u_2, u_3)$.

Critical points of the energy (2.10) with respect to the group representation (2.11). We have the following result.

Proposition 2.7. *The unique critical point of the energy (2.10) with respect to the group representation (2.11) is*

$$\omega^u = \begin{pmatrix} -\dfrac{du_1}{u_1} & 0 & 0 \\ 0 & 0 & \dfrac{u_2 du_3 - u_3 du_2}{\|u_{2,3}\|^2_2} \\ 0 & -\dfrac{u_2 du_3 - u_3 du_2}{\|u_{2,3}\|^2_2} & 0 \end{pmatrix}, \tag{2.12}$$

where $u_{2,3}$ denotes the $\mathbb{R}^2$-valued image (u_2, u_3) and $\|u_{2,3}\|_2$ its Euclidean norm.

Note that we denote the connection 1-form (2.12) by ω^u in order to emphasize its dependence with respect to u.

Proof. By the group representation (2.11), ω is of the form

$$\omega = \begin{pmatrix} \omega_{11} & 0 & 0 \\ 0 & 0 & \omega_{23} \\ 0 & -\omega_{23} & 0 \end{pmatrix}, \tag{2.13}$$

and we have

$$\|Du\|^2 = \|du_1 + \omega_{11}u_1\|^2 + \|du_2 + \omega_{23}u_3\|^2 + \|du_3 - \omega_{23}u_2\|^2.$$

As a consequence, the critical points of the energy (2.10) are the connection 1-forms (2.13) satisfying

$$\begin{aligned} \frac{\partial X}{\partial \omega_{11}} =& 2(u_1 du_1 + \omega_{11}(u_1)^2) = 0 \\ \frac{\partial X}{\partial \omega_{23}} =& 2(u_3 du_2 - u_2 du_3 + \omega_{23}(u_2^2 + u_3^2)) = 0. \end{aligned}$$

We deduce the existence of a unique critical point given by (2.12). □

Let $D^u = d + \omega^u$. We have

$$D^u \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = d\log(\|u_{2,3}\|_2) \begin{pmatrix} 0 \\ u_2 \\ u_3 \end{pmatrix}, \tag{2.14}$$

which proves the $\mathbb{R}^{+*}\times$ SO(2)-equivariance of D^u with respect to u.

Parallel sections of the covariant derivative induced by (2.12). Let v be a section and ω^v the optimal connection 1-form (2.12) induced by v. We have the following result.

Proposition 2.8. *The parallel sections of D^v are the sections u satisfying*

$$D^v u = 0 \iff \begin{cases} \dfrac{d\,u_1}{u_1} = \dfrac{d\,v_1}{v_1} \\ d\|u_{2,3}\|_2 = 0 \\ \dfrac{u_2 du_3 - u_3 du_2}{\|u_{2,3}\|_2^2} = \dfrac{v_2 dv_3 - v_3 dv_2}{\|v_{2,3}\|_2^2}. \end{cases} \tag{2.15}$$

Proof. See Appendix A. □

The existence of parallel sections of a covariant derivative is a direct consequence of the flatness of the corresponding connection 1-form, and it has been shown in [3] that the connection 1-form (2.12) is flat.

Let us assume that u, v represent two color images expressed in an opponent color space, in which u_1, v_1 is the achromatic component and $(u_2, u_3), (v_2, v_3)$ are the chromatic components.
The first equality in (2.15) encodes the equality of the perceptual gradients of the achromatic components of u and v according to Weber's law in visual psychophysics. In polar coordinates $u_{2,3} = (r(u_{2,3}), \varphi(u_{2,3}))$ and $v_{2,3} = (r(v_{2,3}), \varphi(v_{2,3}))$, the coordinate r corresponds to the saturation component (up to the multiplication by a constant) and the coordinate φ to the hue. Hence, the second equality in (2.15) expresses some regularity of the saturation of u. Finally, the third equality in (2.15) reads

$$d\varphi(u) = d\varphi(v), \tag{2.16}$$

which encodes the equality of the gradients of the hue components of u and v.

2.3.2. Dual of the connection 1-form (2.12) and parallel sections of the corresponding covariant derivative. To any connection 1-form ω on a vector bundle E over a manifold M is associated a connection 1-form ω^* by means of an involution on $\Gamma(T^*M \otimes \mathrm{End}(E))$, where $\mathrm{End}(E)$ denotes the bundle of endomorphisms of E. ω and ω^* are said to be dual to each other.
Let us consider the Cartan involution on $\mathfrak{gl}(\mathbb{R}^n)$, the Lie algebra of $\mathrm{GL}(\mathbb{R}^n)$, given by

$$\theta(X) = -X^T. \tag{2.17}$$

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

In this context, the dual of the connection 1-form ω^u in (2.12) is

$$\omega^{u*} = \begin{pmatrix} \dfrac{du_1}{u_1} & 0 & 0 \\ 0 & 0 & \dfrac{u_2 du_3 - u_3 du_2}{\|u_{2,3}\|_2^2} \\ 0 & -\dfrac{u_2 du_3 - u_3 du_2}{\|u_{2,3}\|_2^2} & 0 \end{pmatrix} \tag{2.18}$$

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

$$D^u u = \left(2du_1, d\log\left(\|u_{2,3}\|_2\right) u_2, d\log\left(\|u_{2,3}\|_2\right) u_3\right)^T.$$

It shows that, unlike (2.12), the covariant derivative induced by the connection 1-form (2.18) does not satisfy a $\mathbb{R}^{+*}\times$ SO(2)-equivariance.

Parallel sections of the covariant derivative induced by (2.18). Let v be a section and ω^v the optimal connection 1-form (2.18) induced by v. It follows from Sect. 2.3.1

that

$$D^v u = 0 \Longleftrightarrow \begin{cases} \dfrac{d\,u_1}{u_1} = -\dfrac{d\,v_1}{v_1} \\ d\|u_{2,3}\|_2 = 0 \\ \dfrac{u_2 du_3 - u_3 du_2}{\|u_{2,3}\|_2^2} = \dfrac{v_2 dv_3 - v_3 dv_2}{\|v_{2,3}\|_2^2}. \end{cases} \tag{2.19}$$

Let us assume that u, v represent two color images 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.
The only difference with the parallel sections of the covariant derivative induced by (2.12) is that here, the parallel sections u are such that the perceptual gradient of their achromatic component is opposite to the one of v.

2.4. On optimal vector bundle metrics. Let u be a color image seen as a section of a G-associated bundle equipped with a connection 1-form ω and a Riemannian metric g. We consider the following energy.

$$X(h) = \|Du\|^2_{L^2(g^{-1}\otimes h)}. \tag{2.20}$$

We have the following result.

PROPOSITION 2.9. *The energy (2.20) does not possess critical points.*

Proof. The energy (2.20) is linear with respect to h. □

2.5. Geometric triplets for denoising and deblurring.

2.5.1. VBTV as regularizing term in variational models for color image restoration. Let us consider an image degradation model of the form

$$v = Hu_{clean} + n \tag{2.21}$$

where H is a degradation operator, u_{clean} is the original clean image, v is the observed degraded image, and n some noise. Then, a typical variational model for recovering u_{clean} is

$$\underset{u\in X}{\arg\min} \quad \mathcal{E}(Hu, v) + \lambda R(u), \qquad \lambda > 0 \tag{2.22}$$

for some functional space X, where $\mathcal{E}(Hu, v)$ represents an attachment of Hu to v, and R some regularizer.

The choice of R greatly impacts the nature of the solutions of the models (2.22). For R being the Vectorial Total Variation (VTV), defined by $\mathrm{VTV}(u)=\|\nabla u\|_{L^1}$ (here, ∇ denotes the Jacobian operator), the models tend to provide piece-wise constant solutions, the piece-wise property arising from the use of the L^1 norm and the constancy property arising from ∇u.

As the metrics g and h are positive definite in a geometric triplet, we have

$$\|Du\|_{L^1(g^{-1}\otimes h)} = 0 \Longleftrightarrow Du = 0. \tag{2.23}$$

As a consequence, under the assumption that D possesses parallel sections (which is guaranteed if the curvature of the connection 1-form vanishes identically), models of the form (2.22) with R =VBTV tend to provide piece-wise parallel solutions.

Whereas the existing restoration models of the form (2.22) mainly differ on the choice of the regularizer R, most of them fix the regularizer independently of the operator H. As an example, the Total Variation (TV) has been employed in denoising, deblurring, super-resolution, etc (see e.g. [14]).

We claim that the models could benefit from the selection of a regularizer R that takes into account the specificity of the degradation.

In what follows, we show that R =VBTV can satisfy this property for well-chosen geometric triplets which depend on the observed degradation. More precisely, we show that VBTV can encourage the solutions of the models (2.22) to share some visual content with the original clean image.

2.5.2. A geometric triplet for color image denoising. For $H \equiv Id$, the observed image v is a noisy version of u_{clean}. It is well-known that the perception of contours and textures is less affected by noise than the image intensity values (see e.g. [5]). Hence, a good regularizer for denoising should intend to preserve, or at most smooth in a small extent, the contours and textures of the observed noisy image. Indeed, it would make the solutions of a model of the form (2.22) to have their contours and textures similar to the ones of the original clean image.

VTV is an efficient regularizer for denoising color images as the piece-wise constancy property aforementioned makes the models (2.22) tend to preserve the contours and textures of v. However, one of the drawback of the regularizer $\|\nabla u\|_{L^1}$ is its isotropic property, as it encodes the local structures of the image without considering their orientations. On the other hand, the experiments conducted in the Beltrami framework [26] consider as regularizer the square of the L^2 Riemannian norm of the image gradient: $\|\nabla u\|^2_{L^2(g)}$, g being of the form (2.9) with $\omega \equiv 0$ and $h = diag(\beta, \beta, \beta)$, for $\beta > 0$, in the RGB space. This regularizer is anisotropic due to the particular form of the Riemannian metric. However, by the use of the square of the L^2 norm, this approach tends to oversmooth the contours and textures of the image. As a consequence, we claim that both approaches can be improved by combining them and considering the L^1 Riemannian norm.

In VTV-based models [8],[10] and in the experiments conducted in [26], the different color components are treated in a similar extent by assigning the Euclidean metric or the metric $\mathrm{diag}(\beta, \beta, \beta)$ to the RGB color space. However, the contours and textures of an image are mainly in its achromatic component. Hence, it is desirable to denoise the achromatic component in a smaller extent than the chromatic components. In order to address this problem, we propose to follow the approach in [20] which assigns different weights to the different components of the image expressed in the opponent space given by the basis

$$P = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \end{pmatrix} \tag{2.24}$$

in the RGB frame. These different weights can then be encoded into a vector bundle metric given by the matrix $h = diag(\alpha\beta, \beta, \beta)$ for $\alpha < 1$ in the space (2.24).

Based on the analysis performed in the two previous paragraphs, we consider the VBTV induced by the geometric triplet described in Table 2.1 as regularizer for

TABLE 2.1
Proposed geometric triplet for denoising

Color space	ω	h	g
Opponent space (2.24)	0	$diag(\alpha\beta, \beta, \beta),\ \alpha < 1$	(2.9)

denoising tasks.

Finally, let us mention that it has been shown in a very recent paper [29] that the regularizer SVTV in [20] is a VBTV induced by a particular geometric triplet. Actually, SVTV can be derived from the proposed VBTV described in Table 2.1 by setting $\beta = 1$ in h and considering g as the Euclidean metric on Ω. It turns SVTV into an Euclidean restriction of the proposed VBTV.

2.5.3. A geometric triplet for color image deblurring. For H being a convolution with a normalized kernel of positive coefficients and n small, the observed image v is a blurred version of u_{clean} according to the degradation model (2.21).

As mentioned above, most approaches for image restoration consider the same regularizer for denoising and deblurring tasks. Unlike degradation by noise, the perception of colors is less affected than the perception of local structures under a degradation by blur. As a consequence, VBTV induced by the geometric triplet described in Table 2.1 is not optimal for deblurring. On the other hand, a solution of a variational model of the form (2.22), for R=VBTV induced by the geometric triplet described in Table 2.2 (first row) with $u = v$ in (2.12), tends to be piece-wise parallel with respect to $D^v = d + \omega^v$. According to formula (2.15), it implies that, locally, we have:
1. The perceived gradient (according to Weber's law) of its achromatic component is equal to the one of the achromatic component of v. Moreover, assuming that the noise is negligible, we have the following equality

$$\frac{dv_1}{v_1} := \frac{d(Hu_{clean1})}{Hu_{clean1}} = d\log(Hu_{clean1}).$$

Due to the particular shape of the function log, we have $\log(Hu_{clean1}) \simeq \log(u_{clean1})$ in the regions where the values of u_{clean1} are sufficiently high, which implies that

$$\frac{dv_1}{v_1} \simeq \frac{du_{clean1}}{u_{clean1}}.$$

We deduce that, in bright regions, the model (2.22) encourages the perceived gradients of the achromatic components of the solution and the clean image to be similar.
2. Its saturation component is constant, which can be viewed as a regularity property.
3. Its hue component is identical to the one of v up to an additive constant.
Moreover, a blurring operator (with a reasonable blur level) affects the hue component in a small extent. It implies that:
- the hues of v and u_{clean} are similar (provided that the noise is negligible)
- the data term $\mathcal{E}(Hu, v)$ encourages the solution to have a hue similar to the one of v.
All together, these properties encourage the solution of the model (2.22) to have its hue similar to the one of u_{clean}.

Let us now consider the VBTV induced by the geometric triplet in Table 2.2 (second row). We deduce from the analysis of the parallel sections of the covariant derivative

TABLE 2.2
Proposed geometric triplets for deblurring

Color space	ω	h	g
Opponent space (2.24)	(2.12)	$diag(\beta, \beta, \beta)$	(2.9)
Opponent space (2.24)	(2.18)	$diag(\beta, \beta, \beta)$	(2.9)

induced by the dual connection 1-form (2.18) that the properties 2. and 3. of the solutions of models (2.22) induced by the geometric triplet in Table 2.2 (first row) still hold, but not the first one. Nonetheless, we will see in the experiments conducted in Sect. 3.3 that image restoration benefits from the use of this VBTV in some cases.

3. DIP-VBTV for image restoration. In this Section, we test the model DIP-VBTV

$$\begin{cases} \underline{\theta} = \arg\min_{\theta} \frac{1}{2}\|H(T_\theta(z)) - v\|^2_{L^2(h)} + \lambda\, VBTV(T_\theta(z)), \quad \lambda > 0 \\ \underline{u} = T_{\underline{\theta}}(z), \end{cases} \tag{3.1}$$

with the geometric triplet (g, h, ω) described in Table 2.1 for denoising ($H \equiv Id$) and with the geometric triplets (g, h, ω) described in Table 2.2 for deblurring (H is a blur operator). Here, v denotes the input degraded image and $\underline{u}$ is the output of the model. We show that DIP-VBTV provides state-of-the-arts results.

We use the same network T_θ in all the experiments conducted in this Section. It is an encoder-decoder with skip connections between the down and up layers. It corresponds to the default network in [27], to which we refer for details about the architecture. In particular, for an input image v of size $M \times N \times 3$ (3 represents the number of color channels), the input z of the network is a random image of size $M \times N \times 32$.

3.1. On the numerical scheme to solve the optimization problem.

3.1.1. A boosting numerical scheme. Following the approach in [27], and denoting by $E(\theta; z)$ the energy in (3.1), we consider the following numerical scheme in order to approximate a solution of the model DIP-VBTV

$$\begin{cases} n_{k+1} \sim \mathcal{N}(0, \sigma) \\ z_{k+1} = z_0 + n_{k+1} \\ \theta_{k+1} = \theta_k - lr \nabla E(\theta_k; z_{k+1}) \\ u_{k+1} = \gamma\, u_k + (1-\gamma)\, T_{\theta_{k+1}}(z_{k+1}), \end{cases} \tag{3.2}$$

where $u_0 = v$, z_0 is a fixed random image, lr denotes the learning rate, $\nabla E(\theta_k; z_{k+1})$ stands for the gradient of E with respect to θ_k, and $0 < \gamma < 1$. The iterative scheme is stopped after a certain number of iterations $\underline{k}$ and the output image is $\underline{u} = u_{\underline{k}}$.
We can observe from (3.2) that the input z_k of the network differs at each iteration by perturbing the initial 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.

Last line in (3.2) reveals another boosting technique employed in the numerical scheme, which consists of using an exponential sliding window for some weight γ.

Finally, a last boosting technique employed in [27] consists of averaging the output images of the numerical scheme (3.2) over two different runs.

3.1.2. Stopping criteria. It has been observed in [27] that the numerical scheme (3.2) applied to DIP generates first the low frequencies, then the high frequencies of the image. In particular, it generates noise when the number of iterations is too large.

The numerical scheme can also suffer from destabilization. It means that a significant increase of the energy $E(\theta_k; z_k)$ can occur during the iterative procedure, generating blur in the image $T_{\theta_k}(z_k)$. Then, from such destabilization point, the energy goes down again till destabilized one more time. In order to prevent destabilization, the strategy adopted in [27] consists of tracking the optimization loss and return to parameters from the previous checkpoint iteration 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 $\underline{k}$ should be early enough so that the image $u_{\underline{k}}$ does not possess noise, but it should also stop late enough so that the image $u_{\underline{k}}$ possesses details.

In [30], 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], the need of early stopping is avoided by conducting posterior inference using stochastic gradient Langevin dynamics.

In the experiments conducted in this paper, we follow the strategy in [27].

3.1.3. Parameters of the model and the numerical scheme. We split the parameters in two categories:

1. The parameters of the model DIP-VBTV (3.1):
- The geometric triplet (g, h, ω).
- The trade-off parameter λ between the data term and the penalty term. As in TV-based standard restoration models (see e.g. [14]), the output image in DIP-VBTV can be over-smoothed if λ is too high.

2. The parameters of the numerical scheme (3.2):
- The variance σ of the noise-based regularization, the learning rate lr, the weight γ of the exponential sliding window, whose values are discussed below.
- The number of iterations, which has to be carefully chosen according to Sect. 3.1.2.

3.2. DIP-VBTV for denoising. In this Section, we test DIP-VBTV for denoising, i.e. we consider the model (3.1) with $H \equiv Id$, on a dataset of 9 color images `http://www.cs.tut.fi/~foi/GCF-BM3D/index.html#ref_results` corrupted with additive white Gaussian noise of variance 25. The parameters of the model are described in Table 3.1, where the trade-off parameter λ has be manually tuned with the aim of providing the best average PSNR over the dataset. The parameters of the numerical scheme (3.2) are the default parameters of DIP for denoising [27]: $\sigma = 1/30, lr = 0.01, \gamma = 0.99$.

We compare DIP-VBTV to state-of-the-art unsupervised denoising methods. The results show that, for a well-chosen stopping criteria, DIP-VBTV combined with a boosting technique provides the best average PSNR.

3.2.1. Evolution of the PSNR of DIP and DIP-VBTV with respect to the number of iterations on the whole dataset. In the denoising experiments performed in [27] with DIP and [23] with DeepRED on this dataset, the numeri-

TABLE 3.1
Model DIP-VBTV tested for denoising

Model	Color space	ω	h	g	λ
DIP-VBTV	Opponent space (2.24)	0	$diag(900, 3000, 3000)$	(2.9)	0.05

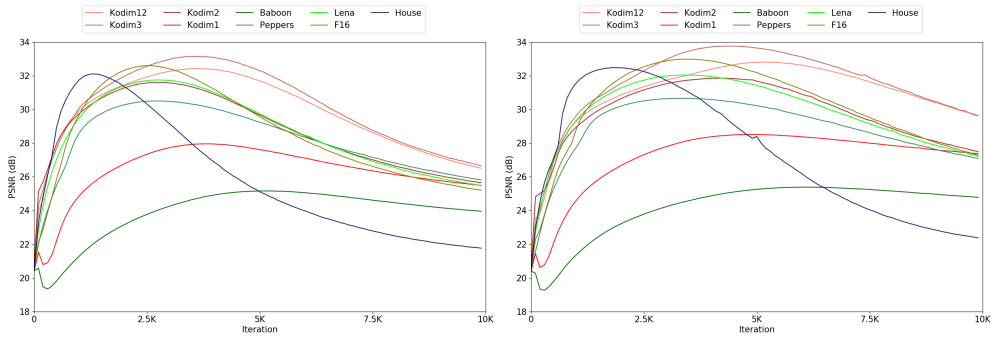

FIG. 3.1. *Denoising: Evolution of the PSNR with respect to the number of iterations for two different models: DIP (left plot) and DIP-VBTV (right plot).*

TABLE 3.2
Denoising: Highest PSNR value over one run (in parenthesis, iteration at which it is reached).

Model	House	Peppers	Lena	Baboon	F16	Kod.1	Kod.2	Kod.3	Kod.12	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	**32.48** (1840)	**30.66** (3330)	**32.04** (3383)	**25.39** (5916)	**32.98** (3440)	**28.52** (4896)	**31.86** (4286)	**33.76** (4399)	**32.81** (5144)	**31.17** (4070)

cal schemes are stopped after different numbers of iterations (1.8K for DIP and 6K for DeepRED). It shows that the number of iterations at which a DIP-based model reaches its maximum PSNR value greatly varies with the model itself. In this Section, we show that the optimal stopping criteria for DIP and DIP-VBTV greatly varies with the image as well.

We run the numerical scheme (3.2) on each of the 9 images of the dataset for both models, and stop it after 10K iterations. Fig. 3.1 shows the evolution of the PNSR with respect to the clean image for each model. The results show that DIP-VBTV is more stable than DIP. Indeed, for each image, the curve is more flat around the peak.

Table 3.2 shows the highest PSNR value reached by each model. The corresponding iteration is indicated in parenthesis. We observe that DIP reaches its highest PSNR at earlier iterations but DIP-VBTV gives better results (mean improvement of 0.37 dB). In this table, images are ordered according to their size from left to right: 256x256 ("House"), 512x512 ("Peppers", "Lena", "Baboon", "F16"), 768x512 ("Kodim1", "Kodim2", "Kodim3", "Kodim12").

Fig. 3.2 compares intermediate results at the iterations 1K, 2.5K, 5K, 7.5K. They confirm that DIP generates noise when the number of iterations is large enough. By adding a VBTV regularizer, we observe that noise is still generated, but in a smaller extent. Indeed, as in any variational model containing a data fidelity term and a regularizer, the trade-off parameter λ in DIP-VBTV determines the amount of the contribution of each term. Hence, if the trade-off parameter is small enough (which

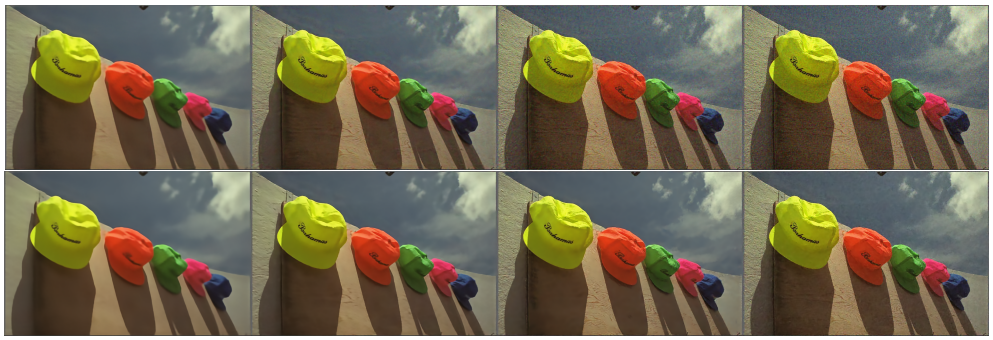

FIG. 3.2. *Denoising: Intermediate results of two different models tested on the image "Kodim3". Top row: DIP - Bottom row: DIP-VBTV. From left to right: 1K, 2.5K, 5K, 7.5K.*

is the case here), then the data term contributes more than the regularizer in the minimization problem. As a consequence, when the number of iterations is large enough, i.e. when we get closer to the solution of the variational problem, the effect of the data fidelity term DIP (generation of noise) has more impact than the effect of the regularizer (smoothing of the image).

A closer look at Fig. 3.1 and Table 3.2 shows a correlation between the iteration corresponding to the highest PSNR and the image size. Note that the synthetic image "Baboon" is an exception as its highest PSNR is reached after a very large number of iterations. We argue that the behavior of the two models comes from:
1) The particular form of the network $T_\theta(z)$, which makes the numerical scheme reconstruct the image from its lowest to highest frequencies throughout the iterative process (see Fig. 3.2).
2) For natural images, the bigger the image size the finer the details (in general).
3) The synthetic image "Baboon" possesses a lot of fine details.

3.2.2. Results of DIP and DIP-VBTV for an image size-based automatic stopping criteria. Following the results in Sect. 3.2.1, we propose a stopping criteria for each model based on image size, and which is detailed in Table 3.3.

Table 3.4 reports the mean PSNR over 5 runs and its standard deviation for DIP and DIP-VBTV tested with the automated stopping criteria described in Table 3.3. We observe that the improvement of DIP-VBTV over DIP (+0.37 dB) reported in Table 3.2 has increased (+0.52 dB). We argue that the main reason for this improvement (+0.15 dB) is that, as mentioned previously, the curve describing the evolution of the PSNR of DIP-VBTV is more flat around its peak than the one of DIP. As a consequence, an automatic stopping criteria is more likely close to the optimal stopping criteria in the case of DIP-VBTV. It explains why the drop of PSNR from Table 3.2 to Table 3.4 is 0.29 dB for DIP and only 0.14 dB for DIP-VBTV.

Finally, note that the standard deviation is rather small in both cases (0.04 for DIP and 0.05 for DIP-VBTV). It shows the stability of the numerical scheme (3.2) with the chosen parameters.

3.2.3. Boosting the results by averaging over several runs. We consider the boosting technique mentioned in Sect. 3.1.1, and which consists of averaging the output images over several runs. Whereas an averaging over 2 runs has been used in [27], we noticed that an averaging over 5 runs provides a bigger improvement. Results are reported in Table 3.5, and they show that DIP benefits slightly more from the

TABLE 3.3
Denoising: Iteration at which the numerical scheme (3.2) is stopped.

Model	256×256	512×512	768×512
DIP	1500	2500	4000
DIP-VBTV	2000	3500	5000

TABLE 3.4
Denoising: Average PSNR and standard deviation over 5 runs

Model	House	Peppers	Lena	Baboon	F16	Kod.1	Kod.2	Kod.3	Kod.12	Average
DIP	32.05 ± **0.02**	30.45 ± **0.02**	31.73 ± 0.04	23.80 ± **0.06**	32.52 ± **0.02**	27.88 ± **0.02**	30.81 ± 0.13	32.97 ± **0.04**	32.37 ± **0.03**	30.51 ± **0.04**
DIP-VBTV	**32.45** ± 0.03	**30.63** ± **0.02**	**31.98** ± **0.02**	**24.60** ± 0.07	**32.97** ± **0.02**	**28.52** ± 0.03	**31.67** ± **0.11**	**33.67** ± 0.08	**32.74** ± 0.04	**31.03** ± 0.05

TABLE 3.5
Denoising: PSNR of the mean image over 5 runs

Algorithm	House	Peppers	Lena	Baboon	F16	Kod.1	Kod.2	Kod.3	Kod.12	Average
DIP	32.58	30.73	32.09	23.98	33.02	28.50	31.46	33.49	32.87	30.97
DIP-VBTV	**32.83**	**30.91**	**32.31**	**24.86**	**33.42**	**29.10**	**32.24**	**34.11**	**33.16**	**31.44**

TABLE 3.6
Denoising: Comparison between DIP-VBTV and unsupervised methods.

Algorithm	PNSR (in dB)
NL-Means	30.26
Bayesian DIP	30.81
DeepRED	31.24
C-BM3D	31.42
DIP-VBTV	**31.44**

boosting technique than DIP-VBTV does (+0.46 dB for DIP and +0.41dB for DIP-VBTV with respect to the results in Table 3.4).

Fig. 3.3 compares the results of DIP and DIP-VBTV boosted by averaging over 5 runs applied to the image "Kodim12". It shows that DIP-VBTV (bottom-right) provides an image which is perceptually closer to the clean image (top-left) than DIP (bottom-left image). Indeed, we can observe that the noise has been completely removed and the sharpest details of the clean image are better recovered (compare for instance the textures in the sand area).

3.2.4. Comparison to other unsupervised methods. We compare DIP-VBTV (boosted by averaging over 5 runs) to other unsupervised methods tested on this dataset for additive white Gaussian noise of variance 25: DeepRED [23], Bayesian DIP [16], NL-Means [11],[12] and C-BM3D [18],[21]. Table 3.6 shows the average PSNR of each of these methods over the dataset. For DeepRED and Bayesian DIP, the results are the ones reported in [23] and [16] respectively. The results of NL-Means and C-BM3D are the ones reported in [27]. The table shows that the best method among these four ones is C-BM3D (PSNR 31.42 dB), which is slightly worse than the results of DIP-VBTV (PSNR 31.44 dB).

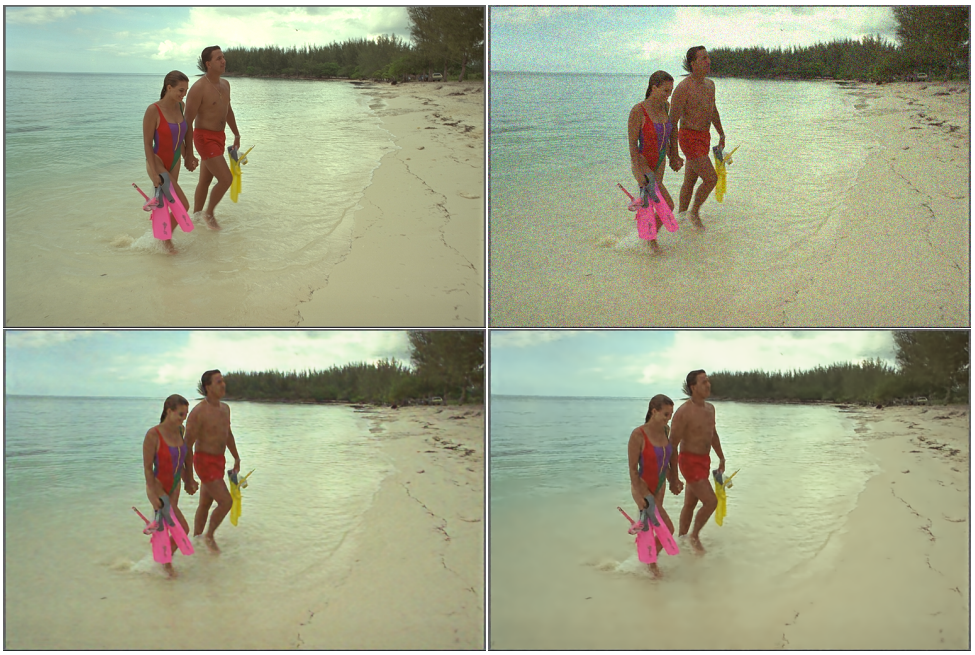

FIG. 3.3. *Comparison between DIP and DIP-VBTV. Clockwise from top-left to bottom-left: Clean ground truth image "Kodak12" - Input noisy image - Result of DIP-VBTV with the boosting technique (PSNR 33.16 dB) - Result of DIP with the boosting technique (PSNR 32.87 dB).*

TABLE 3.7
DIP-VBTV models tested for deblurring

Name	Color space	ω	h	g	λ
DIP-VTV	RGB	0	$\mathbb{I}_3$	$\mathbb{I}_2$	0.0001
DIP-VBTV Optimal	Opponent space (2.24)	(2.12)	$diag(3000, 3000, 3000)$	(2.9)	0.0005
DIP-VBTV Dual	Opponent space (2.24)	(2.18)	$diag(3000, 3000, 3000)$	(2.9)	0.001

3.3. DIP-VBTV model for deblurring. In this Section, we test DIP-VBTV for deblurring, i.e. we consider the model (3.1) for H being a blur operator, on a dataset of 4 color images of size 256x256 corrupted first with a 25x25 Gaussian blur of variance 1.6 and then with additive white Gaussian noise of variance $\sqrt{2}$ (we reproduce the experiments conducted in [23]). The parameters of the models are described in Table 3.7. Note that the trade-off parameter λ for each model has been manually tuned with the aim of providing the best average PSNR over the dataset. The parameters of the numerical scheme (3.2) for each model are the ones used in [23] for DIP: $\sigma = 0.01$, $lr = 0.001$, $\gamma = 0.99$.

3.3.1. Evolution of the PSNR of DIP and DIP-VBTV with respect to the number of iterations on the whole dataset. We run the numerical scheme (3.2) for DIP, and the three DIP-VBTV models described in Table 3.7 for a very large number of iterations (30K). We report the highest PSNR and the iteration at which it is reached for each image in Table 3.8. The results confirm the ones obtained for denoising: the model DIP-VBTV, for a well-chosen geometric triplet, provides higher

TABLE 3.8
Deblurring: Highest PSNR over one run (in parenthesis, iteration at which it 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-VTV	33.35 (8650)	31.92 (10565)	35.88 (12731)	35.76 (12544)	34.23 (11122)
DIP-VBTV Optimal	33.46 (12168)	**32.54** (15538)	**36.30** (22949)	**36.07** (12259)	**34.59** (15728)
DIP-VBTV Dual	**33.88** (12727)	32.51 (20789)	36.01 (13146)	35.84 (11097)	34.56 (14440)

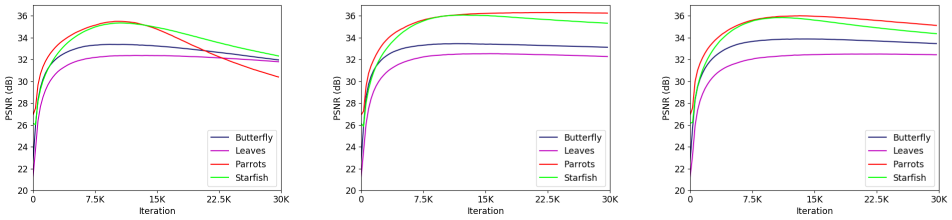

FIG. 3.4. *Deblurring: Evolution of the PSNR with respect to the number of iterations for three different models: DIP (left), DIP-VBTV Optimal (center), DIP-VBTV Dual (right).*

PSNR values than DIP. We also observe that the model DIP-VBTV Optimal provides the best average result and the best results in three of the four images.

Fig. 3.4 compares the evolution of the PSNR throughout the iterative process for DIP (left), DIP-VBTV Optimal (center) and DIP-VBTV Dual (right). It shows that the models DIP-VBTV are more stable than DIP in the sense that the PSNR decreases slower after the peak in all the four cases.

Fig. 3.5 compares intermediate results at the iterations 2.5K, 5K, 10K, 30K of DIP (top row) and DIP-VBTV Optimal (bottom row) tested on the image "Parrots". The images confirm the evolution of the PSNR for this image in Fig. 3.4 (compare the red curves in the left and center plots). Indeed, the images at 10K (which corresponds more or less to the peak of the PSNR for both models) and at 30K are much more different in the case of DIP. 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. 3.4).

3.3.2. Results of DIP and DIP-VBTV for an automatic stopping criteria. Based on the results reported in Table 3.8, we propose an automatic stopping criteria for DIP and the three DIP-VBTV models (see Table 3.9).

Table 3.10 reports the mean PSNR over 5 runs and its standard deviation for the four models tested on the whole dataset. The corresponding stopping criteria are described in Table 3.9. We observe that the results of the previous experiment reported in Table 3.8 are preserved. Indeed, the ranking of the average PSNR among the four models is the same, with DIP-VBTV Optimal giving the best results (34.50 dB) and DIP giving the worse results (34.06 dB). Moreover, we observe that the differences between the average PSNR of the four models is almost maintained (compare the columns "Average" in both Tables), which demonstrates the accuracy of the pro-

FIG. 3.5. *Deblurring: Intermediate results of two models tested on the image "Parrots": Top row: DIP - Bottom row: DIP-VBTV Optimal. From left to right: 2.5K, 5K, 10K, 30K.*

TABLE 3.9
Deblurring: Iteration at which the numerical scheme (3.2) is stopped.

Model	Stopping criteria (number of iterations)
DIP	10K
DIP-VTV	11K
DIP-VBTV Optimal	12K
DIP-VBTV Dual	12K

posed stopping criteria. Finally, let us point out that DIP provides the most stable results in the sense that the mean standard deviation ($\pm$ 0.111) is the lowest among the four models.

In Fig. 3.6, we compare the images providing the best results among the 5 runs (in terms of PSNR) of DIP and DIP-VBTV Optimal tested on the four images. Whereas the best PSNR for DIP is 33.38 dB for "Butterfly", 32.23 dB for "Leaves", 35.69 dB for "Parrots" and 35.49 dB for "Starfish", DIP-VBTV Optimal reaches 33.71 dB for "Butterfly", 32.36 dB for "Leaves", 36.43 dB for "Parrots" and 36.16 dB for "Starfish". We observe that DIP (third column) provides noisier images than DIP-VBTV Optimal (last column), which might explain the difference in PSNR.

3.3.3. Boosting the results by averaging over several runs. We apply the boosting technique aforementioned by considering the mean of the output images of each model over the 5 runs. Table 3.11 reports the PSNR of the mean images. By comparing the "Average" columns of Table 3.10 and Table 3.11, we observe that DIP is the model which benefits the most of the boosting technique (+1.12 dB) and DIP-VTV is the one which benefits the less (+0.73 dB). We also observe that DIP-VBTV Dual benefits more than DIP-VBTV Optimal (+0.92 dB for DIP-VBTV Dual and +0.79 dB for DIP-VBTV Optimal).

In Fig. 3.7, we compare the results of the boosting technique applied to DIP and DIP-VBTV Optimal on the image "Parrots". We observe that, while the boosting technique does reduce the noise of the model DIP (compare the image in the third

TABLE 3.10
Deblurring: Average PSNR and standard deviation over 5 runs

Model	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-VTV	33.36 ± **0.11**	31.82 ± 0.197	35.74 ± 0.156	35.85 ± 0.081	34.19 ± 0.136
DIP-VBTV Optimal	33.57 ± 0.114	32.14 ± 0.192	**36.26** ± 0.162	**36.04** ± 0.114	**34.50** ± 0.146
DIP-VBTV Dual	**33.74** ±0.238	**32.57** ± 0.117	35.88 ± **0.05**	35.70 ± 0.107	34.47 ± 0.128

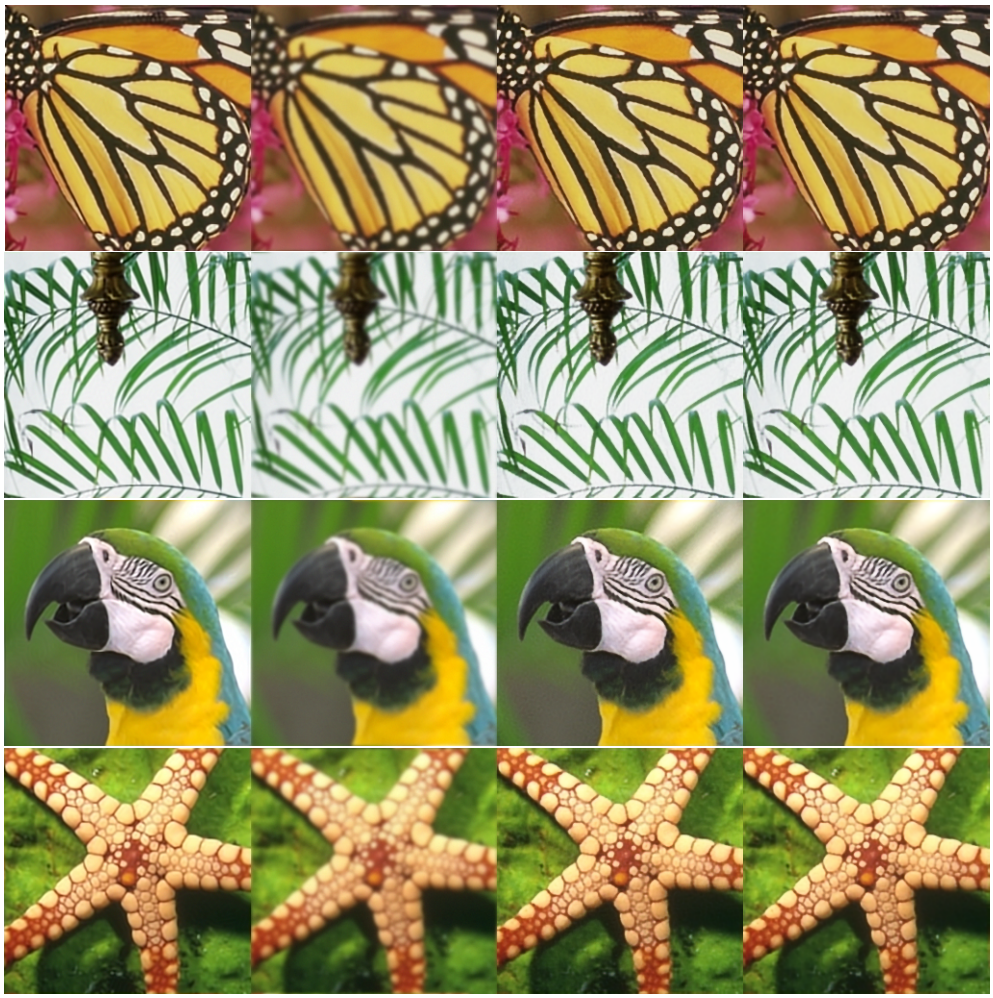

FIG. 3.6. *Deblurring: Comparison of DIP and DIP-VBTV Optimal (best results among the 5 runs in terms of PSNR). From left to right: Clean ground truth image - Corrupted image - Result of DIP - Result of DIP-VBTV Optimal.*

TABLE 3.11
Deblurring: PSNR of the mean image over 5 runs

Model	Butterfly	Leaves	Parrots	Starfish	Average
DIP	34.28	33.14	36.47	36.83	35.18
DIP-VTV	33.98	32.55	36.33	36.78	34.91
DIP-VBTV Optimal	34.24	32.98	**36.86**	**37.08**	35.29
DIP-VBTV Dual	**34.49**	**33.47**	36.74	36.85	**35.39**

FIG. 3.7. *Comparison of DIP and DIP-VBTV Optimal (averaging over 5 runs). From left to right: Clean ground truth image - Corrupted image - Result of DIP (PSNR 36.47 dB)- Result of DIP-VBTV Optimal (PSNR 36.86 dB).*

column in Fig. 3.7 to the image in third row and third column in Fig. 3.6), the average of DIP still possesses more noise than the average of DIP-VBTV.

3.3.4. On the comparison to DeepRED. Since DeepRED [23], even when applied to color images, is evaluated on their luminance channel, it would be fair to evaluate DIP-VBTV on the luminance channel as well in order to compare the two methods. However, for some reasons we ignore, we do not obtain the results reported in [23] when computing the PSNR of the luminance channel of the input corrupted images with respect to the luminance channel of the ground truth images. As a consequence, it makes unfair a direct comparison of the PSNR of the results of both methods. Note also that DeepRED reports an improvement of +0.89dB with respect to DIP, but the authors do not mention the number of iterations used to evaluate DIP.

3.4. Comparison between the models DIP-VBTV, DIP, and L2-VBTV for denoising and deblurring on the Kodak dataset. In this Section, we test different models on the Kodak dataset (`http://r0k.us/graphics/kodak/`), which contains 24 color images. For denoising, we consider the images at their original sizes (768x512 or 512x768) corrupted with additive white Gaussian noise of variance 25. For deblurring, we reduce the image size (384x256 or 256x384) and corrupt them first with a 25x25 Gaussian blur of variance 1.6 and then with additive white Gaussian noise of variance $\sqrt{2}$. Note that the aim of reducing image size is to limit the number of iterations required for DIP-based models.

3.4.1. Coupling DIP with VTV respectively SVTV improves both DIP and L2-VTV respectively L2-SVTV. In this Section, we compare the model DIP-

Algorithm 1 Primal-Dual Algorithm

1: **Initialization:** Choose $\tau, \nu > 0$ with $\nu\tau \leq 1/\|\nabla\|_2^2$ and $(u^0, \eta^0) \in L^2(\Omega;\mathbb{R}^3) \times C^\infty(\Omega;\mathbb{R}^6))$, $\theta \in (0,1]$

2: **Iterations:** For $n = 0, 1, \ldots$ until a stopping criterion is reached

$$\begin{aligned} u^{n+1} &= \mathcal{F}^{-1}\left(\frac{(\tau/\lambda)\mathcal{F}(v)\mathcal{F}(K) + \mathcal{F}(u^n - \tau\nabla^*\eta^n)}{(\tau/\lambda)\mathcal{F}(K)^2 + 1}\right) \\ \overline{u}^n &= 2u^{n+1} - u^n \\ \eta^{n+1} &= \frac{\eta^n + \nu\nabla\,\overline{u}^n}{\max(1, \|\eta^n + \nu\nabla\,\overline{u}^n\|_2)} \end{aligned}$$

TABLE 3.12
Geometric triplets of the different models tested

Name	Model	Color space	ω	h	g
L2-VTV	(3.5)	RGB	0	$\mathbb{I}_3$	$\mathbb{I}_2$
DIP RGB	(3.4)	RGB	-	$\mathbb{I}_3$	-
DIP-VTV	(3.3)	RGB	0	$\mathbb{I}_3$	$\mathbb{I}_2$
L2- SVTV	(3.5)	Opponent space (2.24)	0	diag(0.3,1,1)	$\mathbb{I}_2$
DIP Opp	(3.4)	Opponent space (2.24)	-	diag(0.3,1,1)	-
DIP-SVTV	(3.3)	Opponent space (2.24)	0	diag(0.3,1,1)	$\mathbb{I}_2$

VBTV

$$\begin{cases} \underline{\theta} = \arg\min_{\theta} \dfrac{1}{2}\|H(T_\theta(z)) - v\|^2_{L^2(h)} + \lambda\, VBTV(T_\theta(z)) \\ \underline{u} = T_{\underline{\theta}}(z) \end{cases} \tag{3.3}$$

to the model DIP

$$\begin{cases} \underline{\theta} = \arg\min_{\theta} \dfrac{1}{2}\|H(T_\theta(z)) - v\|^2_{L^2(h)} \\ \underline{u} = T_{\underline{\theta}}(z), \end{cases} \tag{3.4}$$

and to the model L2-VBTV

$$\arg\min_{u \in BV \cap L^2(E)} \frac{1}{2}\|H(u) - v\|^2_{L^2(h)} + \lambda\, VBTV(u), \tag{3.5}$$

where H is a convolution with a kernel K (K is a Gaussian blur for deblurring and K is the Dirac delta function for denoising) with the geometric triplets described in Table 3.12.

Solutions of models L2-VTV and L2-SVTV can be computed through the primal-dual algorithm described in Algorithm 1 [15], where $\mathcal{F}, \mathcal{F}^{-1}$ denote, respectively, the Fourier transform and its inverse, and ∇, ∇^* the Jacobian operator and its adjoint.

In order to solve the model L2-SVTV with Algorithm 1, we first have to express u in an orthonormal basis with respect to the metric given by the matrix $diag(0.3, 1, 1)$ in the opponent space (2.24). One possible basis is

$$P = \begin{pmatrix} 1/\sqrt{0.9} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{0.9} & -1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{0.9} & 0 & -2/\sqrt{6} \end{pmatrix}$$

TABLE 3.13

Denoising: Trade-off parameter, stopping criteria and average PSNR over the Kodak database for each model.

Name	Trade-off parameter	Stopping criteria	PNSR (in dB)
L2-VTV	30	MSE $(u^{n+1}, u^n) < 0.001$ in Algorithm 1	28.70
DIP RGB	-	4K iterations	30.23
DIP-VTV	0.01	4.5K iterations	30.29
L2- SVTV	75	MSE $(u^{n+1}, u^n) < 0.001$ in Algorithm 1	29.20
DIP Opp	-	4.5K iterations	30.42
DIP-SVTV	0.01	5K iterations	**30.65**

in the RGB frame.

Table 3.13 and Table 3.14 report the average PSNR for each model on denoising and deblurring respectively. The trade-off parameter and the stopping criteria for each model and degradation have been manually tuned in such a wey that they provide the best average PSNR over the dataset.

According to the results, we have the following inequalities for denoising

$$\begin{array}{ccccc} \mathrm{PSNR(DIP-VTV)} & > & \mathrm{PSNR(DIP\ \ RGB)} & > & \mathrm{PSNR(L2-VTV)} \\ \wedge & & \wedge & & \wedge \\ \mathrm{PSNR(DIP-SVTV)} & > & \mathrm{PSNR(DIP\ \ Opp)} & > & \mathrm{PSNR(L2-SVTV),} \end{array} \tag{3.6}$$

and the following inequalities for deblurring

$$\begin{array}{ccccc} \mathrm{PSNR(DIP-VTV)} & > & \mathrm{PSNR(DIP\ \ RGB)} & > & \mathrm{PSNR(L2-VTV)} \\ \wedge & & \vee & & \wedge \\ \mathrm{PSNR(DIP-SVTV)} & > & \mathrm{PSNR(DIP\ \ Opp)} & > & \mathrm{PSNR(L2-SVTV).} \end{array} \tag{3.7}$$

The inequalities in (3.6) and (3.7) corroborate the results in [20] in which it has been shown that L2-SVTV outperforms L2-VTV for denoising and deblurring. Note that the comparison in [20] has been done on another dataset and with a different model (the authors consider $\frac{1}{2}\|H(u)-v\|^2_{L^2}$ as data term whereas we consider $\frac{1}{2}\|H(u)-v\|^2_{L^2(h)}$). These inequalities also corroborate the results in [22] in which it has been shown that DIP combined with an (anisotropic) TV outperforms DIP RGB and L2-(anisotropic) TV on denoising and deblurring.

Finally, note that the results reported in Table 3.14 show that replacing VTV by SVTV gives only a minor improvement for deblurring.

3.4.2. DIP-VBTV gives better results than DIP-VTV and DIP-SVTV. In this Section, we test the model DIP-VBTV described in Table 3.1 for denoising and the model DIP-VBTV Optimal described in Table 3.7 for deblurring on the Kodak dataset. The stopping criteria for both degradations have been manually tuned in such a way that they provide the best average PSNR over the dataset. The results, which are reported in Table 3.15 and Table 3.16, show an improvement with respect to the ones reported in Table 3.13 and Table 3.14. In particular, there is an improvement of 0.1 dB on denoising and 0.13 dB on deblurring with respect to DIP-SVTV, and an improvement of 0.46 dB on denoising and 0.14 dB on deblurring with respect to DIP-VTV. Note that the improvement with respect to DIP-VTV on deblurring was

TABLE 3.14
Deblurring: Trade-off parameter, stopping criteria and average PSNR over the Kodak database for each model.

Name	Trade-off parameter	Stopping criteria	PNSR (in dB)
L2-VTV	0.075	MSE $(u^{n+1}, u^n) < 0.001$ in Algorithm 1	28.44
DIP RGB	-	13K iterations	28.84
DIP-VTV	0.0001	20K iterations	29.14
L2- SVTV	0.1	MSE $(u^{n+1}, u^n) < 0.001$ in Algorithm 1	28.49
DIP Opp	-	13K iterations	28.81
DIP-SVTV	0.0001	22K iterations	**29.15**

TABLE 3.15
Denoising: Stopping criteria and average PSNR of DIP-VBTV over the Kodak dataset

Name	Stopping criteria	PSNR (in dB)
DIP-VBTV	5K iterations	30.75

TABLE 3.16
Deblurring: Stopping criteria and average PSNR of DIP-VBTV Optimal over the Kodak dataset

Name	Stopping criteria	PSNR (in dB)
DIP-VBTV Optimal	25K iterations	29.28

higher on the dataset of 4 images tested in Sect. 3.3 (0.31 dB according to Table 3.10).

Fig. 3.8 compares the results of DIP-VBTV and DIP-SVTV on the image "Kodim3" for denoising (PSNR 33.76 dB for DIP-VBTV and PSNR 33.48 dB for DIP-SVTV). We observe in this case that image quality is correlated to PSNR as DIP-VBTV provides an image perceptually closer to the original clean image (for instance, it does not generate noise unlike DIP-SVTV). Nonetheless, there are some details (see for instance the shade under the yellow cap) which are still not reconstructed by DIP-VBTV after 5K (there are partially reconstructed after 7.5K iterations according to Fig. 3.2).

Fig. 3.9 compares the results of DIP-VBTV Optimal and DIP-SVTV on "Kodim23" for deblurring (PSNR 32.44 dB for DIP-VBTV and PSNR 32.05 dB for DIP-SVTV). We observe that the models provide similar results on the recovery of the contours of the original image (see close-up images in the third row). However, as in the denoising case in Fig. 3.8, DIP-VBTV provides an image perceptually closer to the original image in the sense that DIP-VBTV does note generate noise unlike DIP-SVTV (see close-up images in the fourth row).

Finally, let us mention that we did not test the model L2-VBTV for the non Euclidean geometric triplets described in Table 3.1 and Table 3.7 because the non Euclidean metric (2.9), which depends on u, makes the model be non convex and the computation of its solutions not straightforward.

3.5. On the computational time of DIP-based models and the ways to reduce it. DIP-VBTV provides better results than DIP, but at the cost of higher computational time. Indeed, in the previous experiments, we showed that DIP-VBTV requires more iterations than DIP in order to reach its optimal result. Moreover, by computing the time taken by each model to execute one iteration, we observe that DIP-VBTV is slightly slower than DIP. This is actually coherent as DIP-VBTV

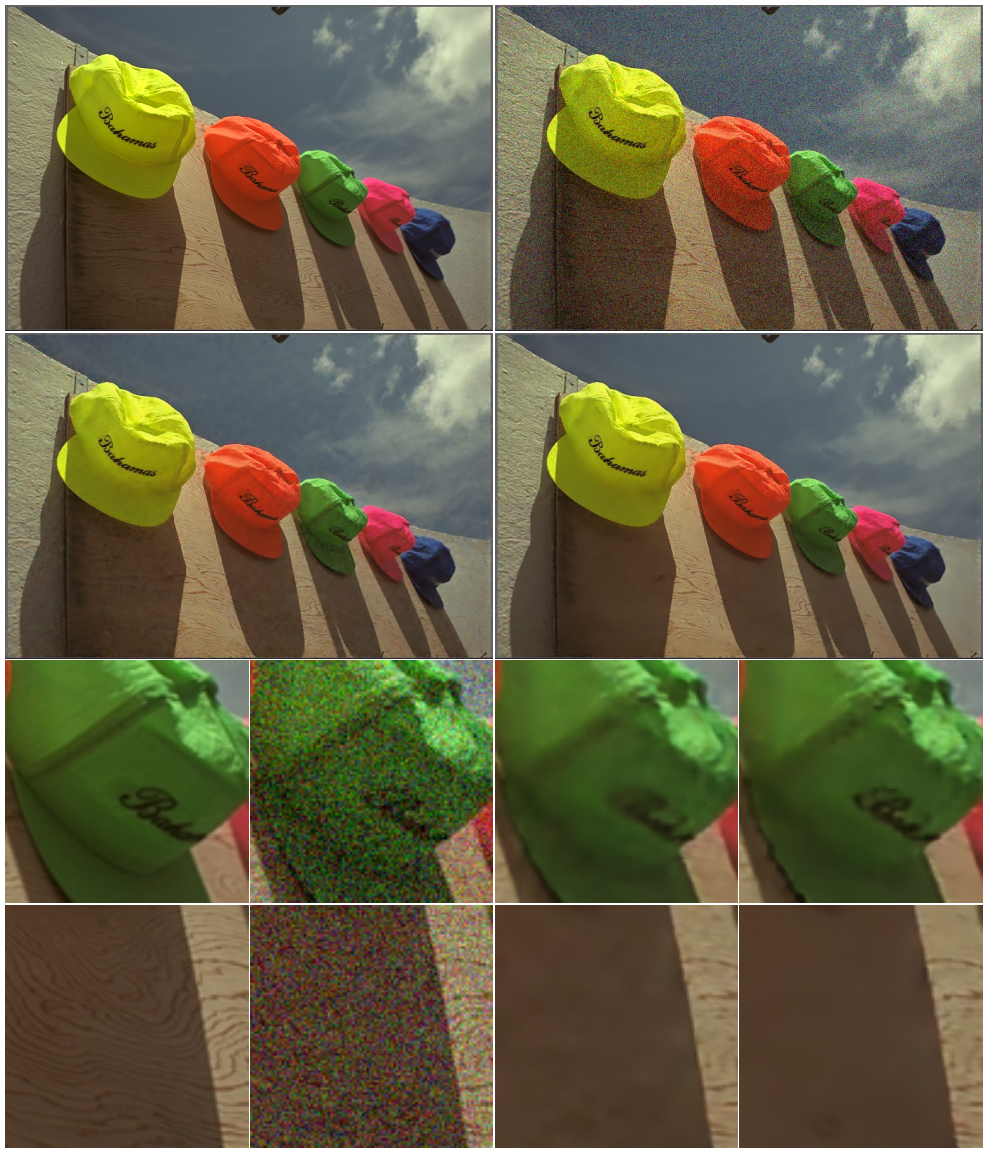

FIG. 3.8. *Denoising: Comparison between DIP-VBTV and DIP-SVTV on "Kodim3". First row from left to right: original image, corrupted image. Second row from left to right: DIP-SVTV (PSNR 33.48 dB), DIP-VBTV (PSNR 33.76 dB). Third and fourth rows: Close-up on some image regions, from left to right: original image, noisy image, DIP-SVTV, DIP-VBTV.*

possesses an extra term (the regularizing term). In Table 3.17 and Table 3.18, we report the computational times of DIP and DIP-VBTV for denoising and deblurring, where the GPU environment (in its free version) provided by Google Colab has been used for the experiments. These results show that the main flaw of DIP-based models is their low computational efficiency.

Different strategies can then be adopted in order to reduce the computational time of DIP-based models. The first approach consists of using a faster GPU. Then, one can reduce the computational time per iteration by reducing the number of parameters

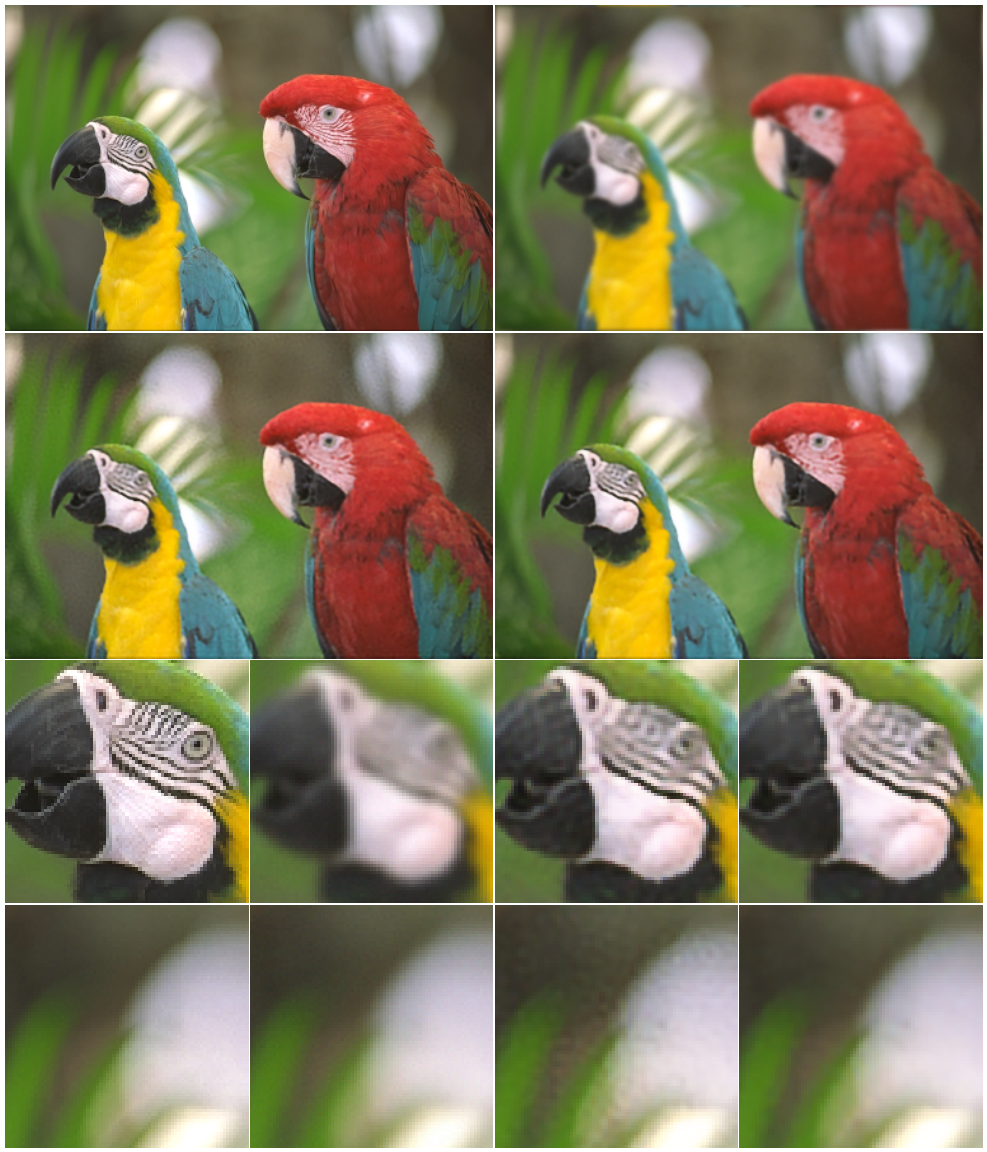

FIG. 3.9. *Deblurring: Comparison between DIP-SVTV and DIP-VBTV models on "Kodim23". First row from left to right: original image, corrupted image. Second row from left to right: DIP-SVTV (PSNR 32.05 dB) - Right: DIP-VBTV (PNR 32.44 dB). Third row: Close-up on some image region, from left to right: original image, blurred image, DIP-SVTV, DIP-VBTV.*

of the network. Finally, one can reduce the number of iterations required to reach the optimal result by improving the optimization process. For instance, we can tune the parameters of the numerical scheme (the learning rate lr and the variance σ in formula 3.2). A more challenging approach consists of developing a mathematical description of the optimization space (the space of parameters of the neural network), from which would derive a more accurate gradient descent. Note that improving the optimization process could also prevent from the destabilization problem mentioned in Sect. 3.1.2.

TABLE 3.17
Denoising: Computational times of DIP and DIP-VBTV in function of the image size

Name	Image size	Stopping criteria	Computational time (in min and sec)
DIP-VBTV	256x256	2K	6.37
DIP-VBTV	512x512	3.5K	11.26
DIP-VBTV	768x512	5K	23.48
DIP	256x256	1.5K	5.5
DIP	512x512	2.5K	8.19
DIP	768x512	4K	19.12

TABLE 3.18
Deblurring: Computational times of DIP and DIP-VBTV in function of the image size

Name	Image size	Stopping criteria	Computational time (in min and sec)
DIP-VBTV	256x256	12K	13.45
DIP-VBTV	384x256	25K	38.51
DIP	256x256	10K	11.7
DIP	384x256	13K	20.09

4. Conclusion. In this paper, we have introduced a variational model, called DIP-VBTV, for color image restoration which combines two priors: a Vector Bundle Total Variation (VBTV) determined by a geometric triplet, and a Deep Image prior (DIP) determined by a neural network. We showed that, for well-chosen geometric triplets arising as critical points of an energy, the minimization of VBTV encourages the solutions of DIP-VBTV to share some visual content with the clean image. Then, we showed on experiments that the restoration benefits from this property. Indeed, we tested DIP-VBTV with these geometric triplets on denoising and deblurring, and results showed that it outperforms other methods involving DIP. Results also showed that the geometric triplet which provides the best result depends on both the image and the degradation operator. Further work is devoted to investigate whether there exist geometric triplets providing better results on denoising and deblurring and to test DIP-VBTV on other image restoration problems.

Appendix A. Proof of Proposition 2.8. We have

$$D^v u = 0 \iff \begin{cases} du_1 - u_1 \dfrac{dv_1}{v_1} = 0 \\ du_2 + u_3 \left(\dfrac{v_2 dv_3 - v_3 dv_2}{v_2^2 + v_3^2} \right) = 0 \\ du_3 - u_2 \left(\dfrac{v_2 dv_3 - v_3 dv_2}{v_2^2 + v_3^2} \right) = 0 \end{cases}$$

$\Longleftrightarrow$

$$\begin{cases} \dfrac{du_1}{u_1} = \dfrac{dv_1}{v_1} \\ u_2 du_2 + u_2 u_3 \left(\dfrac{v_2 dv_3 - v_3 dv_2}{v_2^2 + v_3^2} \right) = 0 \\ u_3 du_2 + (u_3)^2 \left(\dfrac{v_2 dv_3 - v_3 dv_2}{v_2^2 + v_3^2} \right) = 0 \\ u_2 du_3 - (u_2)^2 \left(\dfrac{v_2 dv_3 - v_3 dv_2}{v_2^2 + v_3^2} \right) = 0 \\ u_3 du_3 - u_3 u_2 \left(\dfrac{v_2 dv_3 - v_3 dv_2}{v_2^2 + v_3^2} \right) = 0. \end{cases} \tag{A.1}$$

Hence, we have to show that

$$(D^v u)_{2,3} = 0 \Longleftrightarrow \begin{cases} d\|u_{2,3}\| = 0 \\ \dfrac{u_2 du_3 - u_3 du_2}{\|u_{2,3}\|_2^2} = \dfrac{v_2 dv_3 - v_3 dv_2}{\|v_{2,3}\|_2^2}. \end{cases} \tag{A.2}$$

Summing the second and fifth equations in (A.1) yields

$$u_2 du_2 + u_3 du_3 = 0,$$

i.e. $d\|u_{2,3}\|^2 = 0$, which implies that $d\|u_{2,3}\| = 0$.
Subtracting the fourth equation from the third equation in (A.1) gives

$$u_3 du_2 - u_2 du_3 + \|u_{2,3}\|^2 \left(\frac{v_2 dv_3 - v_3 dv_2}{\|v_{2,3}\|^2} \right) = 0,$$

i.e.

$$\frac{u_2 du_3 - u_3 du_2}{\|u_{2,3}\|^2} = \frac{v_2 dv_3 - v_3 dv_2}{\|v_{2,3}\|^2},$$

which proves that

$$(D^v u)_{2,3} = 0 \Longrightarrow \begin{cases} d\|u_{2,3}\| = 0 \\ \dfrac{u_2 du_3 - u_3 du_2}{u_2^2 + u_3^2} = \dfrac{v_2 dv_3 - v_3 dv_2}{v_2^2 + v_3^2}. \end{cases}$$

On the other hand, assuming that

$$\frac{u_2 du_3 - u_3 du_2}{u_2^2 + u_3^2} = \frac{v_2 dv_3 - v_3 dv_2}{v_2^2 + v_3^2},$$

i.e. $(\omega^v)_{2,3} = (\omega^u)_{2,3}$, we have $(D^v u)_{2,3} = (D^u u)_{2,3}$. Then,

$$d\|u_{2,3}\| = 0 \Longrightarrow (D^u u)_{2,3} = 0$$

according to (2.14), and it leads to $(D^v u)_{2,3} = 0$ as $(D^v u)_{2,3} = (D^u u)_{2,3}$.

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

[3] 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.

[4] 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.

[5] M. Bertalmío and S. Levine, *Denoising an image by denoising its curvature image*, SIAM J. Imag. Sci., 7(1) (2014), pp. 187–211.

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

[7] M. Bertalmío, 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

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

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

[14] 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.

[15] A. Chambolle and T. Pock, *An introduction to continuous optimization for imaging*, Acta Numer., 25 (2016), pp. 161-319.

[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] 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.

[18] 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.

[19] 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.

[20] 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.

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

[22] 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.

[23] G. Mataev, M. Elad and P. Milanfar, *DeepRED: Deep image prior powered by RED*, Proc. ICCV 2019 workshop on Learning for Computational Imaging.

[24] L.I. Rudin, S. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60(1-4) (1992), pp. 259–268.

[25] 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.

[26] N. Sochen, R. Kimmel and R. Malladi, *A general framework for low level vision*, IEEE Trans. Im. Proces., 7(3) (1998), pp. 310–318.

[27] D. Ulyanov, A. Vedaldi and V. Lempitsky, *Deep image prior*, Int. J. Comput. Vis., 128 (2020), pp. 1867-1888.

[28] L.A. Vese and C. Le Guyader, *Variational Methods in Image Processing*, Chapman and Hall/CRC (2015).
[29] W. Wang and M.K. Ng, *Color image restoration by saturation-value total variation regularization on vector bundles*, SIAM J. Imag. Sci., 14(1) (2021), pp. 178–197.
[30] Q. Zhou, C. Zhou, H. Hu, Y. Chen, S. Chen and X. Li, *Towards the automation of Deep Image Prior*, *ArXiv: 1911.07185* (2019).