Deep image prior regularized by coupled total variation for image colorization

Gaetano Agazzotti[1], Fabien Pierre[2], and Frédéric Sur[2]

[1] Mines Nancy and LORIA UMR 7503, Université de Lorraine, CNRS, INRIA, Nancy, France
[2] LORIA UMR 7503, Université de Lorraine, CNRS, INRIA, Nancy, France

Abstract. Automatic image colorization is an old problem in image processing that has regained interest in the recent years with the emergence of deep-learning approaches with dramatic results. A careful examination shows that these methods often suffer from the so-called "color halos" or "color bleeding" effect: some colors are not well localized and may cross shape edges. This phenomenon is caused by the non-alignment of edges in the luminance and chrominance maps. We address this problem by regularizing the output of an efficient image colorization method with deep image prior and coupled total variation.

Keywords: Image colorization · Color halos · Deep image prior · Coupled total variation

1 Introduction

Image and video colorization is the process of adding colors to monochromatic (black-and-white or sepia) pictures and movies, for example legacy documents. This is an ill-posed problem since many solutions may match the human perception. Industrial applications require automatic methods to reduce costs. Historically, the first semi-automatic methods made use of a reference color image or of user-defined color scribbles, in particular through variational approaches. We refer the interested reader to the recent survey paper [14]. Since large datasets made of pairs of monochromatic and color images are easy to build (by simply desaturating a color image), it is possible to train neural networks to infer a color image from a monochromatic input image. Effective colorization algorithms based on deep learning have therefore emerged in the past few years. The present contribution focuses on automatic deep-learning-based approaches without user interaction (contrary to [20] or to the hybrid [6] for instance).

Most deep-learning models infer at each pixel either a color component [7, 9, 13, 18, 19] or a probability distribution over the color space [4, 5, 10, 21]. To limit the computing time, these models are primarily able to infer a low-resolution color image, or are based on a super-pixel representation, or on an auto-encoder. This requires upsampling in a further stage. Upsampling the low-resolution chrominance information or erroneously picking a color through the learnt probability distribution are, however, likely to give colors which are poorly localized

Fig. 1. Typical color halos in three outputs of Colorful Image Colorization [21]. The blue halo on the lion's rump, the orange shade on the building facade or in the sky on the right of the lighthouse, the yellow spots on the jellyfish image should be eliminated from colorization results. Input monochromatic images are shown in Fig. 8.

and go across shape edges, resulting in unpleasant halos (also called color bleeding in [8]) across edges in some situations. Fig. 1 shows typical outputs of a state-of-the-art method. Other examples are described in [16], a careful analysis of DeOldify [1] which is a recent colorization software with impressive results. DeOldify indeed renders color images at a lower resolution than the original monochromatic image, which may give in turn erroneous color halos. This halo effect is often quite subtle and cannot be seen in small-size images as shown in most papers, which mainly aim at getting images with vivid colors. A careful examination suggests, however, that color halos prevent the raw output of these models from being used in real industrial applications.

Several contributions try to circumvent this effect. For instance, the authors of [12] propose to adapt the variational method of [15] in order to restore the output of a deep network [21] to reduce these unwanted defects. They make use of the so-called coupled total variation whose minimization tends to align luminance and chrominance edges, so that color halos are reduced. In the context of color transfer, the authors of [2] use edge-aware texture descriptors with bilateral filtering to reduce halos. More recently, a deep-learning method based on user-defined scribbles was proposed [8]. However, these approaches do not explicitly consider the role of upsampling.

In this paper, we propose to investigate an automated restoration using deep image prior (DIP) [17] together with coupled total variation [15]. The goal is to obtain a colorization free from color halos, based on the hypothesized probability distribution given by [21]. DIP has shown good performances in solving super-resolution from a single image [17]. The expected benefit is thus a correct

upsampling of color information to the resolution of the input monochromatic image. We think that this restoration process is of interest in most colorization methods.

Section 2 introduces deep image prior regularized by coupled total variation and its use in image colorization. Numerical experiments are presented in Section 3. We conclude with Section 4.

2 Deep image prior and image colorization

Deep image prior [17] has been recently introduced to solve ill-posed inverse problems in image processing. It is based on the observation that convolutional neural networks are good at producing images in many applications. The solution of an inverse problem is thus sought as the output of a given neural network f_θ, with a fixed random input z, whose weights θ are learnt to minimise some function based on data misfits between the observation x and the output $f_\theta(z)$. Minimization is performed with the standard optimization machinery based on the back-propagation of errors. DIP falls within the scope of unsupervised learning methods.

Although several neural architectures are possible, we use an encoder-decoder with skip connections all along this paper. This architecture is used in [17] for super-resolution and some other problems.

2.1 What information is captured by deep image prior?

Such a model encodes short and long range correlations between pixels, as illustrated by Fig. 2. This figure shows an experiment in which the preceding network is trained so that θ minimizes $\|f_\theta(z) - x\|$ where x is some image, the optimum weights being denoted by $\theta(x)$. New images are then generated as $f_{\theta(x)}(z')$ where the input z' is randomly drawn. We can see that the distribution of the colors is kept, fine textures and larger structures as well. These properties probably explain why DIP performs well as a Bayesian prior: contrary to the classic TV prior which tends to produce cartoon images, DIP is likely to produce textured images with geometrical shapes.

2.2 Deep image prior and image colorization

Of course, it is hopeless to try to solve the colorization problem by simply minimizing with respect to θ

$$\|L(f_\theta(z)) - y_0\|^2, \tag{1}$$

where $L(f_\theta(z))$ is the luminance channel of $f_\theta(z)$, y_0 is the monochromatic image to colorize, and $\|\cdot\|$ is the quadratic norm. The resulting $f_\theta(z)$ would just show random colors. Additional prior information is required.

From a monochromatic image, the "colorful image colorization" (CIC) model of [21] gives, at each pixel of a low-resolution 64×64 grid $\mathcal{G}$, the probability distribution over 313 samples covering the bidimensional chrominance space (see

Fig. 2. Experimenting with deep image prior. First column: an image x. Second, third, and fourth columns: $f_{\theta(x)}(z')$ with different random inputs z'.

Fig. 3). In [21], a post-processing stage ensures a vivid output image whose resolution is the same as the input resolution. Although very good results are obtained, the probability distribution may not be able to clearly pick a consistent color, especially around the edges of the objects which are not accurately localized over the low-resolution grid. This is illustrated by Fig. 4. Besides, upsampling the chrominance channels is required to build the full-resolution output. This step is likely to produce interpolation artifacts: in the context of colorization, the high-resolution luminance is available but upsampling the chrominance independently from the luminance gives color halos around some objects of the image.

We propose to incorporate the probability distribution given by CIC over the low-resolution pixel grid in the framework of deep image prior to produce color images at the same resolution as the input monochromatic image.

2.3 From probability distributions over a low-resolution grid to the color image

While an RGB output is chosen in the original DIP [17], in the present paper the output of the neural network is made of the three channels of the CIE

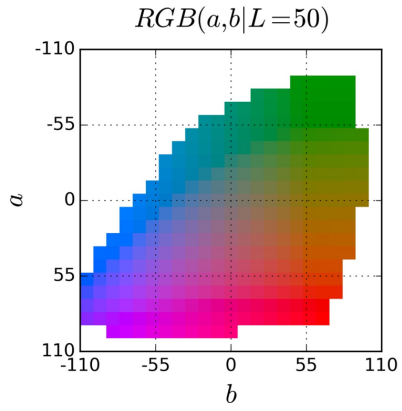

Fig. 3. The 313 color samples in (a, b) chrominance channels used in CIC [21]. Colors correspond to the luminance value $L = 50$.

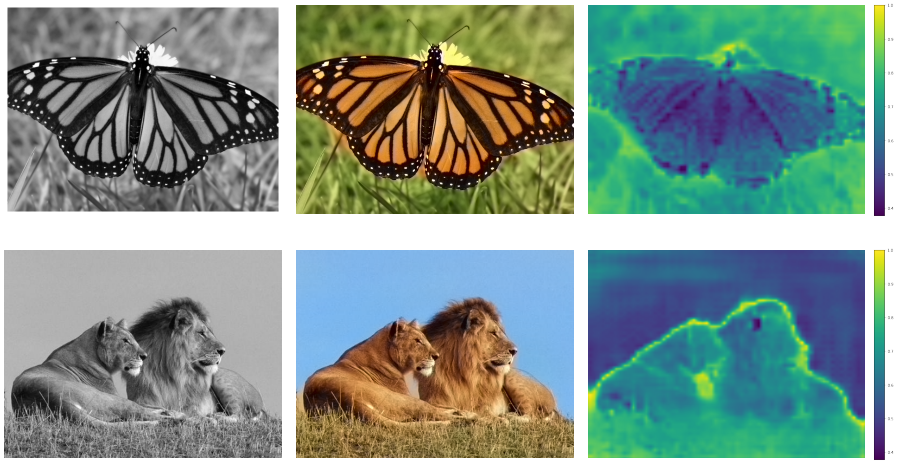

Fig. 4. From left to right: a monochromatic image, the output of "colorful image colorization" [21] (default parameters), and entropy of the hypothesized probability distribution at each pixel of the 64×64 grid $\mathcal{G}$ given by CIC. The larger the entropy, the less sharp the distribution. Large entropy values can be noticed, especially around edges. Color halos can be seen, for example the orange halo on the grass at the bottom left of the butterfly, or the blue halo on the right lion's rump.

Lab color representation, which is the color space used in [21]. Let us denote by L and C the mappings from an Lab image to the luminance and the (a, b) chrominance channels in CIE Lab, respectively. The original monochromatic image is denoted by y_0, and the neural network of DIP is f_θ. This network takes some fixed random z as input and produces an Lab image $f_\theta(z)$ with the same

resolution as y_0. Here, the 313 samples in the bidimensional chrominance space are denoted by $(c_i)_{1\leq i\leq 313}$, and $(w_i(x))_{1\leq i\leq 313}$ is the probability distribution over the c_i's at any pixel x of the low-resolution grid $\mathcal{G}$, as given by CIC. Let d be the subsampling operator which maps images of the same resolution as y_0 and $f_\theta(z)$ to the grid $\mathcal{G}$.

Plugging information from the CIC model into the deep image prior of Eq. (1) can be achieved by minimizing with respect to θ the following quantity:

$$\sum_{x\in\mathcal{G}}\sum_{i=1}^{313} w_i(x)\left\|C\big(d(f_\theta(z))(x)\big)-c_i\right\|^2 \;+\; \alpha\,\|L(f_\theta(z))-y_0\|^2 \tag{2}$$

where $\alpha>0$ is a hyperparameter.

If the neural network f_θ is complex enough, the optimum is attained for f_θ whose luminance channel is close to y_0 and chrominance channels are, at any pixel x of the low-resolution grid, the average of the (c_i) samples weighted by the probability distribution $(w_i(x))$. The DIP permits thus to upsample information from the low-resolution grid $\mathcal{G}$. The problem with this formulation is twofold. First, averaging chrominances gives dull colors, especially at pixels where the probability distribution is not sharp. Second, luminance and chrominance channels are coupled only through the hidden layers of f_θ, which still gives color halos.

We therefore adapt Eq. (2) in two ways:

1. We add a regularization term which explicitly enforces the coupling between luminance and chrominance channels. To this end, we use the coupled total variation (TV) introduced in [15]. If I is a color image of domain Ω, with luminance L and chrominance channels a and b, the coupled TV writes:

$$\mathrm{TV}_\gamma(I)=\int_\Omega \sqrt{\gamma\|\nabla L\|^2+\|\nabla a\|^2+\|\nabla b\|^2} \tag{3}$$

   where ∇ denotes the gradient and $\gamma>0$ is a hyperparameter of the model. Minimizing the coupled TV makes the luminance and chrominance gradients to have small values at the same pixels [15]. This consequently aligns the edges in the chrominance and luminance channels, and reduces the color halo effect.
2. We allow the probability distribution $(w_i(x))$ to vary, in the same spirit as in [12]. Minimizing $\sum_{i=1}^{313} w_i(x)\|C\big(d(f_\theta(z))\big)(x)-c_i\|^2$ with respect to the non-negative $w_i(x)$ amounts to solving a simple linear program, the constraint being $\sum_{i=1}^{313} w_i(x)=1$. The minimum is therefore obtained at a vertex of the polytope defined by the constraints: the optimum $w_i(x)$ is such that there exists i^* satisfying $w_{i^*}(x)=1$ and for any $i\neq i^*, w_i(x)=0$. For any x, the index i^* is thus simply $\mathrm{argmin}_i\|C\big(d(f_\theta(z))(x)\big)-c_i\|^2$. Such a sharp probability distribution has the advantage of giving vivid colors.

Input: a monochromatic image y_0, and probability distributions $(w_i(x))_{i=1\ldots313}$ at every pixels x of a 64×64 grid $\mathcal{G}$ over samples $(c_i)_{i=1\ldots313}$ spanning the bidimensional chromatic space (from [21]).

Repeat until convergence:

1. Minimize the loss (Eq. (5)) with respect to θ.
2. For any $x \in \mathcal{G}$, $i^*(x) = \text{argmin}_i \|C\big(d(f_\theta(z))\big)(x) - c_i\|$ and $w_i(x) = 1$ if $i = i^*(x)$, $w_i(x) = 0$ otherwise.

Output: the RGB image obtained from the luminance channel y_0 and the chrominance channels of $f_\theta(z)$.

Fig. 5. Algorithm for colorization from the low-resolution probability distributions of CIC [21] with deep image prior and coupled total variation.

As a consequence, we seek θ and w such that $\forall x \in \mathcal{G}, \sum_{i=1}^{313} w_i(x) = 1$, minimizing:

$$\sum_{x\in\mathcal{G}} \sum_{i=1}^{313} w_i(x) \left\|C\big(d(f_\theta(z))\big)(x) - c_i\right\|^2 + \alpha \left\|L(f_\theta(z)) - y_0\right\|^2 + \beta\, \mathrm{TV}_\gamma(f_\theta(z)) \tag{4}$$

The optimal $f_\theta(z)$ is such that all three terms are small, that is, the luminance of f_θ is close to y_0, halo effects are reduced (small coupled TV), and the chrominances over the low-resolution grid are closed to ones picked from a 0-1 probability distribution.

Compared to the original DIP [17], it should be noted that the additional TV_γ regularisation term permits to get rid of overfitting. DIP is indeed known to overfit and requires an early-stopping strategy. The gradient descent optimization must be stopped after a certain number of iterations, which may critically depend on the inverse problem to solve. A classic way to prevent overfitting in neural network learning in general and in DIP in particular is to add a regularization term in the loss function, see [3, 11] for total-variation based regularization of DIP. In the experiments of Section 3, we indeed simply minimize the loss function without any early-stopping process.

2.4 Optimization algorithm

We use block coordinate descent to minimize Eq. (4): we alternate minimization with respect to θ, the parameters of the DIP neural network, and to w, the probability distributions. Minimizing with respect to θ is achieved with the classical back-propagation, and minimizing with respect to w amounts to solving a linear program as explained in Section 2.3.

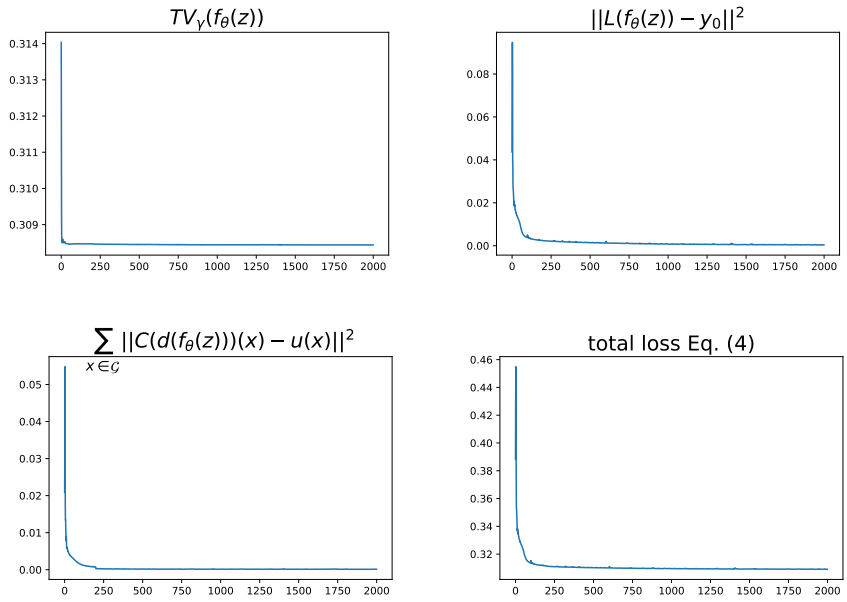

Fig. 6. Minimized loss (bottom right) and its components (related to coupled TV, luminance, and chrominance) against number of iteration in a typical case.

It can be noted that minimizing Eq. (4) with respect to θ by gradient descent is equivalent to minimizing:

$$\sum_{x \in \mathcal{G}} \left\| C\big(d(f_\theta(z))\big)(x) - u(x) \right\|^2 \; + \; \alpha \, \| L(f_\theta(z)) - y_0 \|^2 \; + \; \beta \, \mathrm{TV}_\gamma(f_\theta(z)) \tag{5}$$

with $u(x) = \sum_i w_i(x) c_i$. Indeed, with $v_\theta(x) = C\big(d(f_\theta(z))\big)(x)$,

$$\begin{aligned} \nabla_\theta \left(\sum_{x \in \mathcal{G}} \sum_{i=1}^{313} w_i(x) \, \| v_\theta(x) - c_i \|^2 \right) &= 2 \sum_{x \in \mathcal{G}} \sum_{i=1}^{313} w_i(x) \nabla_\theta v_\theta(x) \cdot (v_\theta(x) - c_i) \\ = 2 \sum_{x \in \mathcal{G}} \nabla_\theta v_\theta(x) \cdot \left(v_\theta(x) - \sum_i w_i(x) c_i \right) &= \nabla_\theta \left(\sum_{x \in \mathcal{G}} \| v_\theta(x) - u(x) \|^2 \right) \end{aligned} \tag{6}$$

where $\cdot$ denotes the dot product, since $\sum_i w_i(x) = 1$ for any $x \in \mathcal{G}$.

Our software implementation uses the equivalent Eq. (5) instead of Eq. (4) for the unsupervised learning step to reduce the computational burden of back-propagation. The resulting algorithm is given in Fig. 5. In practice, Step 2 is performed after repeating 200 gradient descent iterations of Step 1, which limits the number of Step 2 performed, this latter step being time-consuming in spite that only a few pixels are concerned by a change of their probability distribution.

Fig. 7. Chrominance channels. Left: output of CIC [21]. Right: proposed approach (minimization of Eq. (5)). The proposed approach with regularized deep image prior gives a better localization of the chrominance information, which explains reduced color halos. The corresponding color images can be seen in Fig. 2. Best seen on screen.

We can see the effect of this alternating scheme in Fig. 6. The curve of the total loss is globally smooth but contains some small jumps. The function becoming asymptotically constant after 1,000 iterations, stopping after 2,000 iterations ensures the convergence of the iterative algorithm. In all the experiments of this paper, convergence curves are similar to Fig. 6. Contrary to the standard DIP approach [17], there is no need for early stopping.

3 Numerical experiments

As mentioned in the introduction, colorization is an ill-posed problem as several different solutions may be consistent with a unique monochromatic image (for instance, the color of a car cannot be determined from its black-and-white image). Consequently, this section shows qualitative results: the goal is to assess whether the proposed approach gives plausible colorizations without color halos.

Our software programs are freely available at the following URL:
https://gitlab.univ-lorraine.fr/pierre26/diptv

3.1 Parameters

The parameters have been chosen experimentally once and for all. On all the tested images, the same parameters have been used: $\alpha = 1$, $\beta = 5.10^{-7}$, and $\gamma = 80$. The β parameter controls the regularity of the image and works together with the γ parameter. β encourages some flat areas and small variations of the image whereas γ controls the smoothness/sharpness of the contours of the chrominance channels and the coupling of these channels with the luminance one. The optimizer minimizing the loss is ADAM with a learning rate of 2.10^{-2}.

The optimization of Eq. (4) with respect to θ takes about 162 sec for 256×256 images on a GeForce GTX 1060 GPU with 6GB memory ($2,000$ iterations, as explained in the preceding section). In comparison, the computation of the initial distribution by CIC [21] takes less than 2 seconds.

3.2 Experiments

In Fig. 7, chrominance channels are shown by applying a constant luminance channel equal to a 50% value. This process helps to qualitatively validate the results of a colorization algorithm by neutralising the effect of luminance. Indeed, colorized images often seem to be visually satisfactory, in spite of poorly localized chrominances. This is caused by the sensitivity of the human visual system to luminance, which hides potential defects of the colorization process. We can see that contours are sharper with our method. Moreover, regularizing with coupled total variation removes some defects on flat areas (for instance the bottom part of the sky below the plane). Sharpness of the chrominance channels shows some benefits in terms of colorization results: for instance, the orange halo on the left of the butterfly (which is the result of a wrong color assignment because of a flat probability distribution given by CIC, as proven by Fig. 4) is removed by our optimization scheme.

Fig. 8 shows some comparisons between [12] and the results obtained with (4). It emphasizes the benefit of DIP with respect to a traditional variational method. It is important to note that the outputs of [12] are themselves significant improvements over the output of CIC [21]. It can be seen in the lion image that the contours are better respected. The blue halo over the lion's skin and the orange halo in the sky seen in the output of [12] are corrected by the proposed regularization step. The space between the two lions is not blue as one would

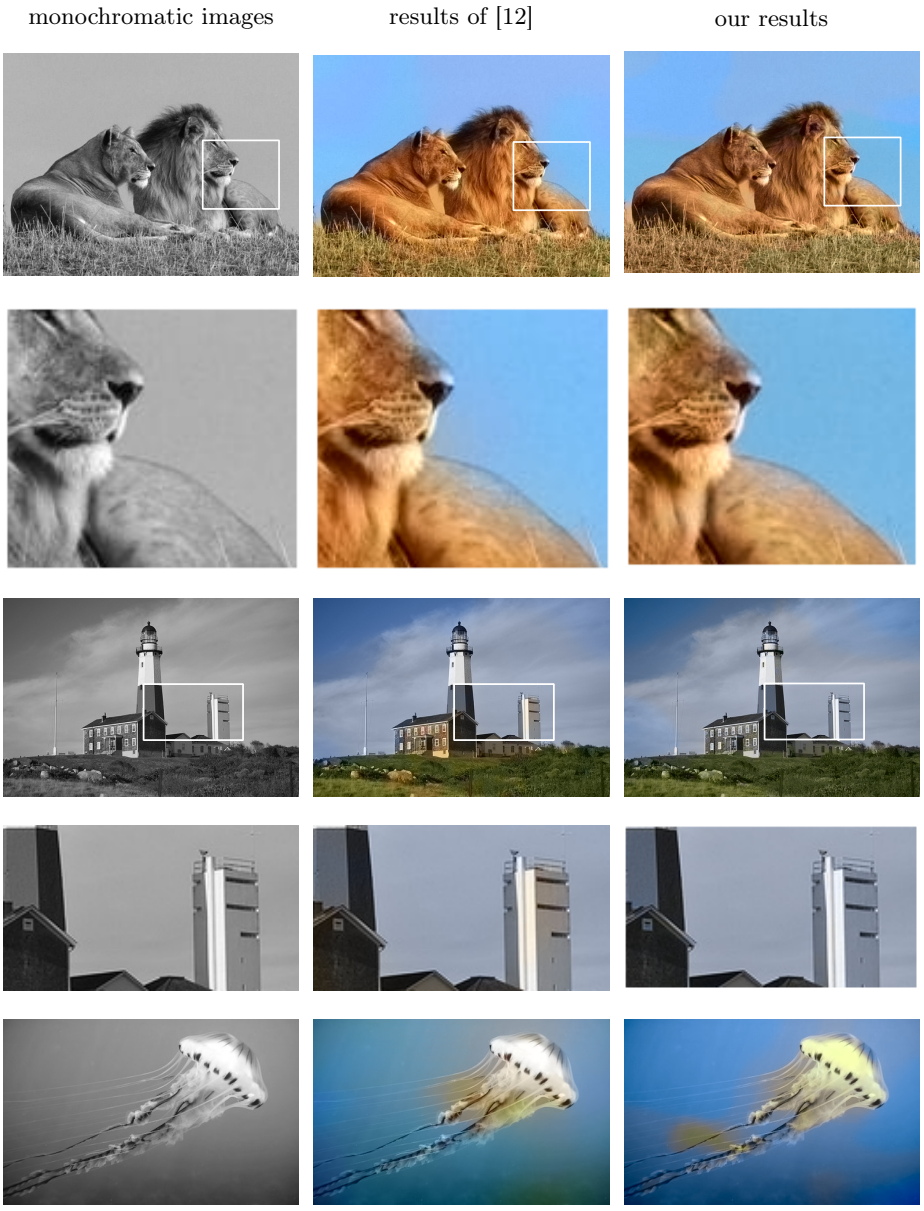

Fig. 8. Comparison with [12]. Left: original monochromatic images. Center: results of [12] (from Figs. 5, 6 and supplementary material of [12]); right: proposed approach, minimization of Eq. (5). From top to bottom: lion, close-up, lighthouse, close-up, jellyfish. The proposed approach shows significant improvements over [12], which itself improves over CIC [21] (see Fig. 1).

expect, but it turns out that, in this area, blue has a low probability in CIC on which our approach depends. In the lighthouse image, orange halo on the facade of the right-hand side building seen in the output of [12] disappears with the proposed approach. In the jellyfish image, the bottom right halo seen in [12] has been removed. Besides, the jellyfish is colorized with [12] through a brown blurry halo in the chrominance channels, whereas the proposed approach produces a structured colorization well-fitted to the body of the jellyfish. However, it can also be seen that the DIP approach produces a brown spot in the lower left of the image, not related to any structure of the luminance channel. It turns out that this phenomenon sometimes appears, depending on the random initialization z of DIP. It is quite rare: we show this particular output for the sake of completeness; most outputs are not affected by it.

4 Conclusion

This paper proposed a deep image prior approach to image colorization, based on the probability distribution given by a state-of-the-art neural network over a low resolution pixel grid. The proposed regularization scheme was able to produce full-resolution chrominance information through a deep image prior regularized by coupled total variation, which permitted to align the chrominance and luminance edges, the latter being available at full-resolution in the original monochromatic image. As a result, we have shown that the so-called color halos (or color bleeding) were reduced, which is a first step towards effective use of colorization.

Acknowledgments This research was funded, in whole or in part, by l'Agence Nationale de la Recherche (ANR), project ANR-21-0008-01. For the purpose of open access, the authors have applied a CC-BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

References

1. Antic, J., Howard, J., Manor, U.: DeCrappification, DeOldification, and super resolution. Fast.ai course (2019), https://www.fast.ai/posts/2019-05-03-decrappify.html
2. Arbelot, B., Vergne, R., Hurtut, T., Thollot, J.: Local texture-based color transfer and colorization. Computers & Graphics **62**, 15–27 (2017)
3. Batard, T., Haro, G., Ballester, C.: DIP-VBTV: A color image restoration model combining a deep image prior and a vector bundle total variation. SIAM Journal on Imaging Sciences **14**(4), 1816–1847 (2021)
4. Deshpande, A., Lu, J., Yeh, M.C., Chong, M., Forsyth, D.: Learning diverse image colorization. In: Proc. Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2877–2885 (2017)
5. Deshpande, A., Rock, J., Forsyth, D.: Learning large-scale automatic image colorization. In: Proc. International Conference on Computer Vision (ICCV). pp. 567–575 (2015)

6. Huang, Z., Zhao, N., Liao, J.: Unicolor: A unified framework for multi-modal colorization with transformer. ACM Transactions on Graphics (Proc. SIGGRAPH'22) **41**(6) (2022)
7. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Let there be Color!: Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. ACM Transactions on Graphics (Proc. SIGGRAPH'16) **35**(4), 110:1–110:11 (2016)
8. Kim, E., Lee, S., Park, J., Choi, S., Seo, C., Choo, J.: Deep edge-aware interactive colorization against color-bleeding effects. In: Proc. International Conference on Computer Vision (ICCV). pp. 14667–14676 (2021)
9. Kim, G., Kang, K., Kim, S., Lee, H., Kim, S., Kim, J., Baek, S.H., Cho, S.: Bigcolor: Colorization using a generative color prior for natural images. In: Proc. European Conference on Computer Vision (ECCV). pp. 350–366 (2022)
10. Larsson, G., Maire, M., Shakhnarovich, G.: Learning representations for automatic colorization. In: Proc. European Conference on Computer Vision (ECCV) (2016)
11. Liu, J., Sun, Y., Xu, X., Kamilov, U.: Image restoration using total variation regularized deep image prior. In: Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 7715–7719 (2019)
12. Mouzon, T., Pierre, F., Berger, M.O.: Joint CNN and variational model for fully-automatic image colorization. In: Proc. Scale Space and Variational Methods in Computer Vision (SSVM) Conference. Lecture Notes in Computer Science, vol. 11603, pp. 535–546 (2019)
13. Pan, X., Zhan, X., Dai, B., Lin, D., Loy, C., Luo, P.: Exploiting deep generative prior for versatile image restoration and manipulation. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(11), 7474–7489 (2022)
14. Pierre, F., Aujol, J.F.: Recent approaches for image colorization. In: Chen, K., Schönlieb, C.B., Tai, X.C., Younces, L. (eds.) Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision. Springer (2021)
15. Pierre, F., Aujol, J.F., Bugeau, A., Papadakis, N., Ta, V.T.: Luminance-chrominance model for image colorization. SIAM Journal on Imaging Sciences **8**(1), 536–563 (2015)
16. Salmona, A., Bouza, L., Delon, J.: DeOldify: A review and implementation of an automatic colorization method. Image Processing On Line **12**, 347–368 (2022)
17. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proc. Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
18. Vitoria, P., Raad, L., Ballester, C.: ChromaGAN: Adversarial picture colorization with semantic class distribution. In: Proc. Winter Conference on Applications of Computer Vision. pp. 2445–2454 (2020)
19. Xia, M., Hu, W., Wong, T.T., Wang, J.: Disentangled image colorization via global anchors. ACM Transactions on Graphics (Proc. SIGGRAPH'22) **41**(6) (2022)
20. Zhang, R., Zhu, J.Y., Isola, P., Geng, X., Lin, A., Yu, T., Efros, A.: Real-time user-guided image colorization with learned deep priors. ACM Transactions on Graphics (Proc. SIGGRAPH'17) **36**(4) (2017)
21. Zhang, R., Isola, P., Efros, A.A.: Colorful image colorization. In: Proc. European Conference on Computer Vision (ECCV) (2016)