

RESEARCH

Open Access



Deep image prior inpainting of ancient frescoes in the Mediterranean Alpine arc

Fabio Merizzi^{1,2*}, Perrine Saillard⁴, Oceane Acquier⁴, Elena Morotti³, Elena Loli Piccolomini¹, Luca Calatroni^{2*} and Rosa Maria Dessì⁴

Abstract

The unprecedented success of image reconstruction approaches based on deep neural networks has revolutionised both the processing and the analysis paradigms in several applied disciplines. In the field of digital humanities, the task of digital reconstruction of ancient frescoes is particularly challenging due to the scarce amount of available training data caused by ageing, wear, tear and retouching over time. To overcome these difficulties, we consider the Deep Image Prior (DIP) inpainting approach which computes appropriate reconstructions by relying on the progressive updating of an untrained convolutional neural network so as to match the reliable piece of information in the image at hand while promoting regularisation elsewhere. In comparison with state-of-the-art approaches (based on variational/PDEs and patch-based methods), DIP-based inpainting reduces artefacts and better adapts to contextual/non-local information, thus providing a valuable and effective tool for art historians. As a case study, we apply such approach to reconstruct missing image contents in a dataset of highly damaged digital images of medieval paintings located into several chapels in the Mediterranean Alpine Arc and provide a detailed description on how visible and invisible (e.g., infrared) information can be integrated for identifying and reconstructing damaged image regions.

Keywords Digital inpainting, Medieval paintings, Deep Image Prior

Introduction

The synergy between art history, mathematical image analysis and artificial intelligence (AI) is a stimulating meeting point between disciplines to favour the

development of new science and to complement historical studies in art and art history. These new tools and methods lead to an emerging approach in the comprehension of medieval images as living objects, see, e.g., [1]. In this work we focus on the digital reconstruction of wall paintings of medieval chapels located in the south of the Alpine arc. The wall paintings in this area were produced mainly between the second half of the 15th century and the early 16th century [2]. We are interested in particular in the wall paintings signed or attributed to the painters Giovanni Baleison and Tommaso and Matteo Biazaci. They were active in the last quarter of the 15th century in current France and Italy. Their peculiarity is the frequent use of texts in their painted images. As part of several restoration campaigns and/or more specific modifications linked to the shift of perception and reception of the images depicted in the murals, such paintings have been subject to modifications in later times. Furthermore, the

*Correspondence:

Fabio Merizzi
fabio.merizzi@unibo.it
Luca Calatroni
luca.calatroni@inria.fr

¹ Department of Computer Science and Engineering (DISI), University of Bologna, via Mura Anteo Zamboni 7, 40126 Bologna, Italy

² Laboratoire d'Informatique, Signaux et Systèmes de Sophia-Antipolis (I3S), CNRS, Inria, Université Côte d'Azur, Les Algorithmes - bât. Euclide B, 2000, route des Lucioles, 06903 Sophia-Antipolis, France

³ Department of Political and Social sciences, University of Bologna, Str. Maggiore 45, 40125 Bologna, Italy

⁴ Cultures et Environnements Préhistoire, Antiquité, Moyen Âge, Université Côte d'Azur, Pôle Universitaire Saint Jean d'Angély, 24, avenue des Diables Bleus, 06300 Nice, France



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

effect of the environment and/or the intentional erasure and vandalism caused the disappearance of several imaging data crucial for the understanding of some images and painted texts.

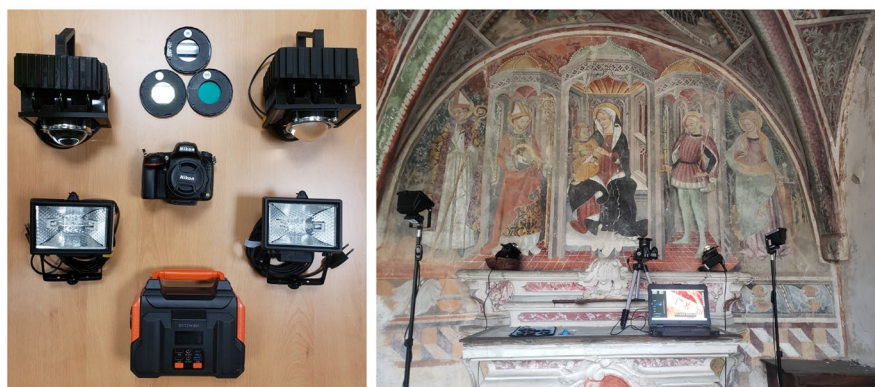
In order to digitally restore the missing/lost image elements made indecipherable by such processes, digital reconstruction approaches and among them, image inpainting [3], can be applied, see [4–6] for previous applications in digital humanities contexts. Given the lack of information, the restoration of the original version of the degraded image under consideration is impossible (inpainting is indeed an ill-posed problem lacking uniqueness) so the objectives of inpainting in this context are rather concerned to the reconstruction of a coherent visual experience to the observer, which may help the comprehension and interpretation of damaged images in historic studies. Moreover, a careful analysis of the output images may shed light on whether the observed corruptions are involuntary or intentional, thus generally favouring a better understanding of the overall artistic process. By combining inpainting with multi-spectral techniques, interesting piece of information can be unveiled, such as the stratification of murals and the evolution of images over time. A further aim of our digital reconstructions is to determine both the dates and the authors of each image layer which, compared to major artworks, are still debated. From a historical viewpoint, our objective is to grasp the causes at the roots of transformations that may be aesthetic, religious, or ideological. In this way, we think this interdisciplinary project between art history, mathematical image processing, and AI, can allow us to chronicle the life of the paintings and better understand their impact and evolution in past societies. The reconstruction of digital images of frescoes characterized by large occlusions with irregular shapes is a very challenging task. A large variety of the inpainting approaches proposed in the literature rely either on the expert choice of the reconstruction model by the user [7, 8] or on the use of large training sets of data [9], which both limit their practical use in the field of digital humanities. We consider an unsupervised neural approach for the digital inpainting of images of highly damaged frescoes. Our method belongs to the class of so-called *Deep Image Prior* algorithms [10]. Compared to supervised approaches relying on large data sets of examples, the proposed approach is fully unsupervised and performs reconstruction based only on the observation of the damaged image and on the detection of the region to be filled in. We detail in this work how such existing approach can be applied to the challenging task of digital reconstruction of highly damaged frescoes and highlight the modifications performed both in the neural architecture and in the DIP loss function to improve both performance

and stability. Our setting is proved to be effective in comparison to state of the art approaches and validated on both simulated and real data including, e.g., the restoration of textual characters and the use of infrared data for the study of the transformation/retouching process the artworks have been subject to. This manuscript is organized in the following manner: In Sect. [Dataset description and challenges](#) the image dataset used for our study is described and enriched with information on the artistic/historical context. In Sect. [State-of-art methods for image inpainting](#) a comprehensive discussion on state-of-art inpainting methods is given, covering both handcrafted and data-driven approaches. In Sect. [Deep Image Prior inpainting](#), we introduce the DIP approach and our proposal. In Sect. [Experimental setup](#), the overall pipeline of our approach is described, spanning from the initial treatment and analysis performed on the given image to inpaint till the final inpainted result. Several numerical results are reported in Sect. [Numerical results](#) where comparisons between inpainting approaches and combined techniques making use of both visible and invisible (infrared) data are combined, thus showing the potential of the proposed approach to the study of imaging data in digital humanities. At last, we draw our conclusions in Sect. [Discussion and outlook](#).

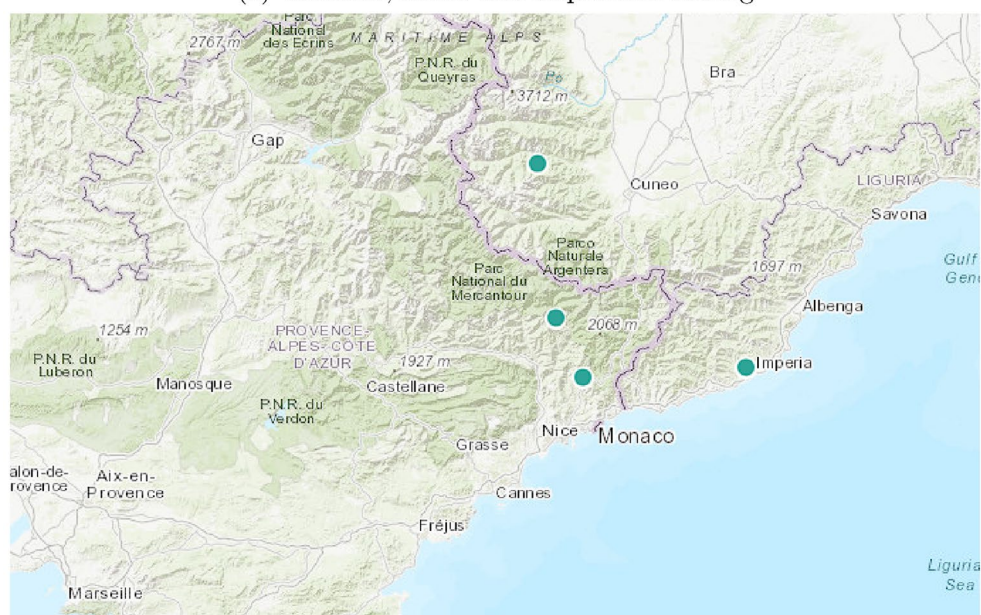
Dataset description and challenges

The image dataset used in this project has been collected in the online database PA'INT [2] (CEPAM, UCA, FR) which has been collected as part of the PhD thesis of O. Acquier [11]. The database is composed by a large collection of digital images of late medieval wall paintings representing visual scenes and epigraphic items in religious buildings of the south of the Alpine arc. In total, 269 painted monuments have been geolocated of which 75 have been the object of several image acquisition campaigns. As a result, 2600 pictures have been collected and indexed to various details such as the name of the painter(s) (when known), the date(s) of completion as well as a visual descriptions. A total number of 1172 inscriptions have been analysed in [11]. Note that currently PA'INT is in the process of being expanded with images in the infrared and ultraviolet spectral range, which will be analysed and integrated by means of AI tools in a later work. The images in the dataset have been acquired by a modified Nikon D610¹ [12], in which a filter that blocks ultraviolet and infrared (IR) has been removed, with the Nikon AF-S NIKKOR 50 mm f/1.8G lens. In order to limit the light reception to the desired spectral range, some light filters were used corresponding

¹ Our digital camera has been modified by EOS FOR ASRTO.



(a) Cameras, filters and acquisition setting



(b) Locations of the four chapels, along the Alpine arc between France and Italy.

Fig. 1 Locations, devices and experimental setup for data acquisition

to a wavelength of 380–780 nm for the visible spectrum and 780–1100 nm for the infrared spectrum. Flashes BOWENS GEMINI 1500 pro as well as lighter and less bulky halogen lamps from CHSOS [13] were used, see Fig. 1a. For the infrared emissions, halogen lamps are placed at approximately 45° of the studied painted surfaces, which were also captured in the visible range for comparisons/data-integration, see Fig. 1. The interest of IR acquisitions is that they can reveal retouches and underwritings if the overpainter layer is IR-transparent and the underpaintings are not. For some references on the use of scientific imaging in digital humanities, we refer to [14].

As a case study, we analysed incomplete and retouched images of wall paintings acquired in four chapels: the chapel *Sainte-Claire*² in Venanson, France, the sanctuary *Nostra Signora delle Grazie* in Imperia, Italy, the chapel *Notre Dame de Bon Coeur* in Lucéram, France and the chapel *San Sebastiano* in Celle di Macra, Italy. See Fig. 1b for their geolocalizations.

The decoration of the *Sainte Claire* chapel was painted by Giovanni Baleison in 1481. The Venanson community had this chapel constructed, and the decorations were commissioned by Guillaume Cobin, as indicated in the signature (Fig. 16). It is best known as the Saint Sébastien

² Also called chapel of Saint Sébastien because of the representation of the saint.

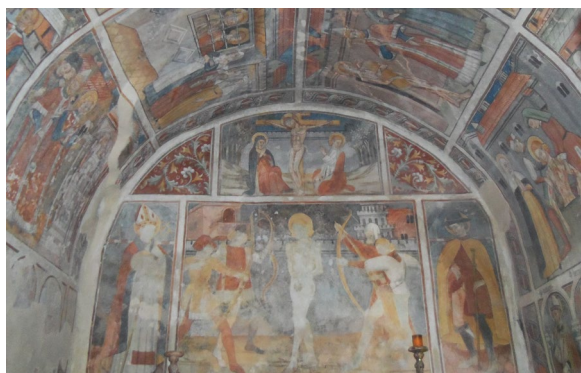


Fig. 2 Martyrdom of S. Sébastien in Venanson

chapel because a large portion of the wall paintings is dedicated to the life of saint Sebastian, and his martyrdom is depicted in the chevet of the chapel, see Fig. 2. Unlike the frescoes in Celle di Macra and Montegrazie, the chapel walls do not depict Hell. However, they still feature, like *Nostra Signora delle Grazie*, the theme of *cavalcade* of vices, a popular motif in the Alps during that period.

The sanctuary of *Nostra Signora delle Grazie* has undergone at least four decoration campaigns since the late 15th century. In this paper, we will focus on the

frescoes painted by the Biazaci brothers in 1483 (Fig. 17) and by Pietro Guido da Ranzo between 1524 and 1540 (Fig. 18). The decorations were overpainted during the 18th century and were rediscovered during restoration campaigns throughout the 20th century. The images presented in this paper illustrate the virtues of *charitas* and *sobrietas* as painted by Tommaso and Matteo Biazaci and details from Pietro Guido’s *Mocking of Christ*, respectively. The wall paintings from the chapel *Notre Dame de Bon Coeur* are attributed to either Giovanni Baleison or the Master of Lucéram. The decoration was executed between 1480 and 1485.

Figure 3 shows the chapel of San Sebastiano in Celle di Macra and the representation of Hell painted therein by Giovanni Baleison in 1484. The fresco is divided into eight parts, among which seven are dedicated to a particular capital sin, while the last one is Lucifer’s den. In this work, we will focus in particular on the images of *Lusuria* and *Invidia*, see Fig. 4. The scene represented in *Lusuria*, Fig. 4a, is ruled by the demon Asmodeus. Its circle welcomes souls prone to lust and carnal pleasures in their earth life. In this scene, green and yellow demons are torturing sinners: a demon is whipping a woman while pulling her hair. Three sinners are sitting on a grill fed by a demon, while a group of men and women are



0 1m

Fig. 3 The chapel of San Sebastiano in Cella di Macra, Italy

(a) *Lusuria*(b) *Invidia***Fig. 4** Two selected scenes from the chapel of San Sebastiano in Cella di Macra, from Fig. 3

burning inside a building. *Invidia*, see Fig. 4b, constitutes the fourth infernal pit, ruled by the blue demon Belzebub. The pit hosts sinners culpable of envy and malignancy. The demon is accompanied by four green and yellow dragons which are painted in the action of lacerating sinners. The damned souls are divided into two groups, each composed by three persons tied up to a spike. Due to the extensive deterioration of these paintings, responsible for making numerous painted texts present in the background not understandable and prone to possible misinterpretations. A digital reconstruction procedure is expected to facilitate the understanding of the written text and, overall, of the painted scene.

State-of-art methods for image inpainting

The problem of image inpainting consists of filling in missing or damaged parts of an image (representing, e.g., a fresco) using a source of prior information.

In mathematical terms, given a colour image \tilde{x} defined on an image domain $\Omega = \{(i, j) : i = 1, \dots, m, j = 1, \dots, n\}$ of size $m \times n$ having an occluded region $D \subset \Omega$, the problem is defined in terms of a masking operator $m \in \{0, 1\}^{m \times n}$ acting pointwise as follows:

$$m_{i,j} = \begin{cases} 1 & \text{if } \tilde{x}_{i,j} \in \Omega \setminus D \\ 0 & \text{if } \tilde{x}_{i,j} \in D. \end{cases} \quad (1)$$

By definition, the mask m is thus nothing but the characteristic function of the set $\Omega \setminus D$ and identifies the reliable (i.e., unoccluded) pixels in the observed image.

Most of the classical approaches employed over the last three decades rely on the use of mathematical approaches favouring the transfer of the available image content within the region to be filled in by means of diffusion/transport processes and/or by copy-paste procedures of appropriate patches.

Often, their design requires a certain modelling expertise aimed at choosing which type of diffusion (linear VS. non-linear, for instance) is preferred for the image at hand. We will refer to this class of approaches as *hand-crafted* approaches, meaning by that name the fact that they are designed by an expert user. As their numerical implementation often relies on the use of iterative algorithms, these approaches have been also called *sequential* algorithms in the recent literature [15]. We provide a review of these methods and of their main features in Sect. [Inpainting by hand-crafted approaches](#).

More recent techniques rely on the shared idea of filling in the incomplete image regions by novel image content generated by neural networks trained on large image datasets [9]. Due to the prominent role played by the data for this class of approaches, we will refer to them as *data-driven* approaches and describe their main features in Sect. [Inpainting by data-driven approaches](#).

In the following paragraphs we review the main available literature on both approaches, with a particular attention to their application to their use in the field of cultural heritage.

Inpainting by hand-crafted approaches

Hand-crafted methods for digital image inpainting have been actively proposed since the early 2000s. The most famous approaches are based on local diffusion techniques, which can fill the missing regions by diffusing image information locally, from the known image portions into the adjacent damaged ones, at the pixel level, see, e.g. [7, 8] for reviews. These approaches model the problem in a variational form where the inpainted image \hat{x} solves:

$$\hat{x} \in \operatorname{argmin}_x \lambda \|m \odot (x - \tilde{x})\|^2 + R(x), \quad (2)$$

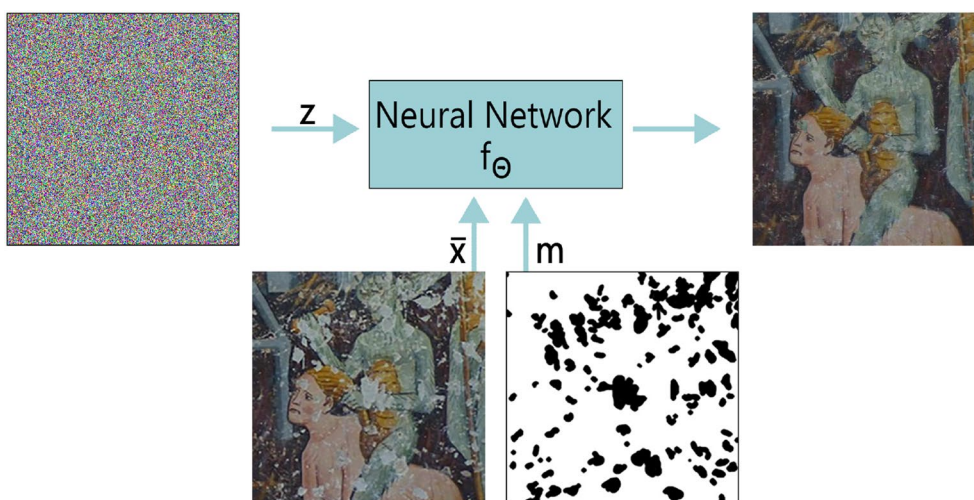


Fig. 5 DIP inpainting methodology. The network is fed random noise z , original image \bar{x} , and binary mask m , to produce as output the inpainted image

where the data term forces x to stay close to the data \bar{x} on $\Omega \setminus D$ and $R(\cdot)$ is a regularisation term favoring the propagation of contents within D . The effect of regularization against data fidelity is weighted by $\lambda > 0$. In the data term, the symbol \odot stands for the Hadamard element-wise product. Partial Differential Equation (PDE) approaches stem from (2) by considering the corresponding Euler-Lagrange equations, possibly embedded within an artificial evolution towards the minimizer(s) of the corresponding functional.

A popular instance of (2) proposed in [16] consists in choosing a regularization term $R(\cdot)$ favouring piecewise constant reconstructions via non-linear diffusion. This can be done by choosing $R(x) = TV(x)$, the Total Variation (TV) regularization functional which acts on images as:

$$TV(x) = \sum_{c \in \{R,G,B\}} \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \sqrt{(x_{i+1,j}^c - x_{i,j}^c)^2 + (x_{i,j+1}^c - x_{i,j}^c)^2}, \tag{3}$$

where $x_{i,j}^c$ denotes the intensity value of the $c \in \{R, G, B\}$ channel of the image at pixel $(i, j) \in \Omega$.

More complex choices can be made at a variational level such as, e.g., higher-order regularization (see, e.g., [17]). On the other hand, from a PDE viewpoint, advanced approaches making use of Navier–Stokes models propagating colour information by means of complex diffusive fluid dynamics laws have been considered in [3, 18–21]. Other approaches involved the use of transport and curvature-driven approaches [22–24].

Being based on the discretization of differential operators, the hand-crafted approaches described above favour *local* regularization. As a consequence, they are particularly suited to reconstruct only small occluded regions such as scratches, text, or similar. In the context of heritage science, they have been employed for restoring ancient frescoes in works such as [3–5] showing effective performance.

On the other hand, such techniques fail in reconstructing large occluded regions and in the retrieval of more complex image content such as texture. To overcome such limitation, *non-local* inpainting approaches have been proposed in a variety of papers (see, e.g. [25–27]) to propagate image information using patches. In more detail, the main idea consists of comparing patches from the known image regions in terms of a

suitable similarity metric which can further take into account rigid transformations and/or patch rescaling. The popularised PatchMatch approach [28] is based on this principle, with the further advantage of computing correspondence probabilities for each patch and thus weighting the contribution coming from different locations appropriately. Improved versions of PatchMatch have been proposed, e.g., in [29, 30] where such averaging is performed in a non-local manner. Compared to local approaches, patch-based inpainting methods show remarkable performance and, where properly

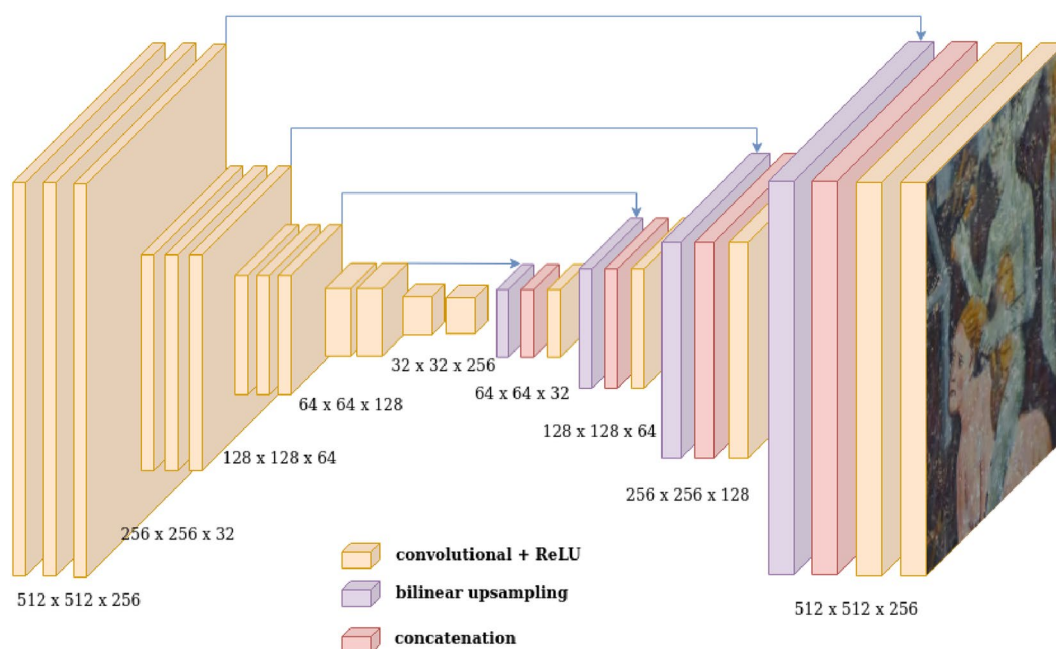


Fig. 6 The architecture of the DIP network: “hourglass” architecture, downsampling via convolution and upsampling via bilinear upsampling and skip connections

tuned, good reconstruction of both geometric and textured contents. Nonetheless, due to their intrinsic non-convexity, they are often initialization dependent and are sensitive to the choice of hyperparameters such as, e.g., the patch size. In the context of art restoration, in [6] a combination of a local (as initialization) and non-local (as the main inpainting process) procedure was used for the digital restoration of severely damaged illuminated manuscripts.

An interesting comparison between local/non-local sequential approaches for the inpainting of digital images of artworks has been conducted in [31]. Interestingly, the authors therein noted that while manual restoration still seems to lead to the best results, reconstructions obtained by model-based approaches appear often misleading for expert evaluation, while as good as a manual reconstruction for naïve eyes.

The choice of the most appropriate hand-crafted model (in particular, of the most appropriate term $R(\cdot)$ favouring inpainting within D) often requires some technical modelling expertise. This limits the use of this class of approaches in practice, as an optimal choice of such term typically requires the understanding of advanced concepts in linear/non-linear diffusion and smooth/non-smooth optimisation which are highly non-standard for practitioners.

Inpainting by data-driven approaches

Data-driven approaches for image inpainting offer an alternative strategy to the conventional methods of modeling image regularity through predefined energy functionals. Instead, these methods leverage an extensive array of training data and employ neural techniques to estimate mappings from occluded input images to inpainted images. Due to their better deep encoding capabilities, neural approaches are indeed not limited to the modeling of the sole geometric/texture regularities in an image, but they further capture the presence of local/non-local patterns and the semantic meaning of image contents.

An exhaustive review of learning-based approaches for image inpainting is presented in [9]. Upon prior knowledge of the inpainting region, i.e. of the mask operator in (2), data-driven inpainting approaches based on convolutional networks have been designed in [32, 33] and improved in some recent works such as [34, 35], with the intent to adapt the convolutional operations only to those points providing relevant information.

The performance of data-driven inpainting dramatically improved after the introduction of the generative adversarial network (GAN) architectures in [36]. GANs aim to minimize the distance between ground truth images and reconstructed images not in a point-wise manner, but, rather, in a distributional sense, through the use of two competing networks, the former able to

discriminate between ground truth data and samples generated by the latter. Whenever a large number of examples is available, GANs and, more in general, generative neural approaches, are very effective for inpainting, see, e.g. [33, 37–41]. Improved approaches perform inpainting by working, rather than at an image level, at the level of feature space, by first reconstructing the geometric content and finally adding finer textures, see for instance [42, 43].

More recently, Denoising Diffusion Probabilistic Models (DDPM) [44] have emerged with comparable and possibly overall greater inpainting performance than GANs. DDPMs can achieve optimal results in generative tasks without the impairment typical of GAN models, such as adversarial learning instabilities and high computational cost [45]. A recent effort in inpainting with diffusion models reported impressive results [46] by conditioning the reverse diffusion process with mask information. Other recent examples of neural data-driven inpainting techniques based, e.g., on diffusion models include [47–50].

Despite their excellent performance, data-driven approaches have scarcely been used to perform digital inpainting tasks. Some examples are, e.g., [51–53] where (generative) learning approaches are employed. In order to generate suitable image contents, these approaches require the availability (or the synthetic generation) of large datasets of relevant and high-quality data and occlusion type for training. This constitutes indeed a major limitation in the reconstruction of highly-damaged frescoes painted by local authors for which, therefore, very little training data is available.

Generally speaking, the use of data-driven approaches to solve the problem of digital inpainting is often limited due, essentially, to:

- The scarce availability of reference data to be used for training;
- The bias induced by non relevant data during inpainting.

Deep image prior inpainting

To overcome the limitations of the approaches described before, we will consider in the following a tailored approach, popularised under the name of *Deep Image Prior* (DIP) in [10]. This approach combines the interpretability of hand-crafted regularisation models with the power of data-driven methods. It employs a neural procedure to inpaint the image and, in comparison to classical learning schemes, makes use of the sole observed image as a training example.

This technique pioneers the use of low-level image statistics extracted from an image by the network structure itself, hence DIP allows to obtain an accurate inpainted image without a training set, exploiting an expressive untrained architecture on just one degraded image. In other words, DIP enables the use of a neural technique in our specific inpainting application.

In Fig. 5, we graphically represent how DIP works for the inpainting problem at hand. In particular, we show that the neural network takes as input an image z , randomly sampled from a uniform distribution with a variable number of channels, and it also considers the damaged image \bar{x} and its corresponding mask m , then it gives as output the restored image. Formally, the DIP approach computes the vector of neural network parameters $\hat{\Theta}$ by solving the minimisation problem:

$$\hat{\Theta} \in \operatorname{argmin}_{\Theta} \|m \odot (f_{\Theta}(z) - \bar{x})\|^2, \quad (4)$$

where $f_{\Theta}(\cdot)$ is a neural network with parameters Θ . By solving (4), the parameters $\hat{\Theta}$ generate an output image $\hat{x} = f_{\hat{\Theta}}(z)$ matching at best \bar{x} outside D and filling contents in $\Omega \setminus D$. Numerically, this problem can be solved by standard iterative optimisation algorithms such as gradient descent with back-propagation. Being (4) a non-convex optimisation problem, different initialisations for Θ may lead to different results. Note that DIP implicitly enforces regularisation through the network structure, unlike traditional methods, but the early stopping of iterations is necessary to avoid overfitting.

Clearly, the training procedure (4) depends on the given image \bar{x} to be inpainted. In case several images are to be restored, the weights must be recomputed for each degraded image, independently. As a consequence, the DIP computational cost is more similar to the one of model-based methods than to data-driven approaches, where the parameters are computed only once using large exemplar sets with a very expensive training phase.

DIP architecture and regularisation

The DIP reconstruction procedure depicted in Fig. 5 makes use of the network architecture represented in Fig. 6. The "hourglass" structure consists of convolutional downsampling and bilinear upsampling with a filter stride equal to 2, whereas the non-linearity considered is a LeakyReLU. In more detail, downsampling is achieved via strides and convolution or via max pooling and downsampling with Lanczos kernel. For the upsampling, the two most common approaches are bilinear upsampling and nearest neighbours upsampling. Regarding convolutional filters, we tested both filters with the same size and a progressively increasing number for both the encoder

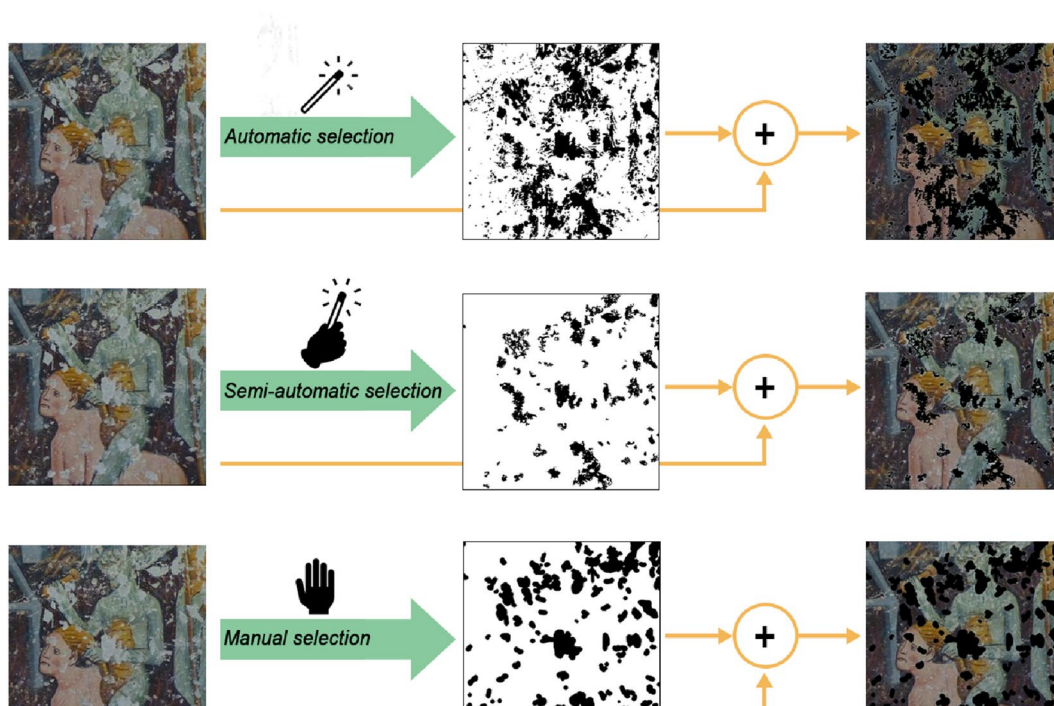


Fig. 7 Comparison of mask-making methods, for our application the manual method proved to be the most practical

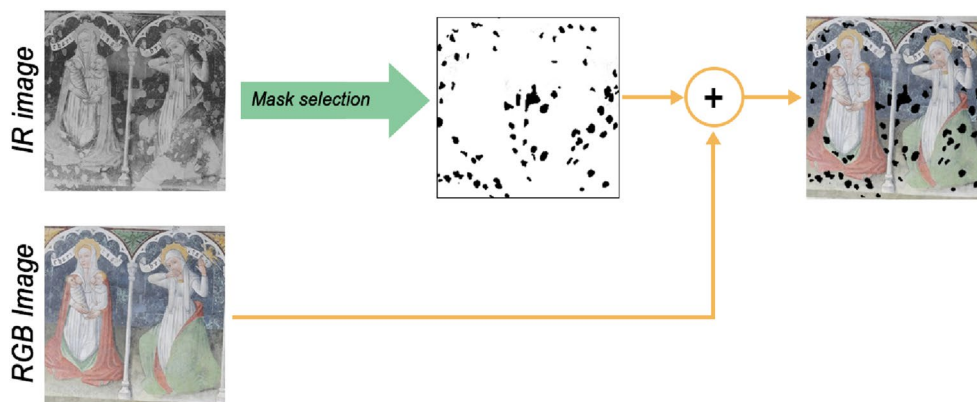


Fig. 8 Mask making via an IR version of the RGB image, exploiting IR-enhanced contrasts to effectively select damaged areas

and decoder. The size of the filters defines the sensitivity of the convoluted network to different scales of features. In our experiments, we kept the filter size at 3x3 for all the convolutional layers and we finally chose the reflection padding for more local coherent results in the corner areas.

Input and output images are of the same size, i.e. 512×512 pixels. The input image is generally drawn from a multi-variate uniform noise distribution with values in $[0, 1]$. The performance of the model is significantly impacted by the selection of the optimiser. After

evaluating various options, we ultimately decided to use RMSProp (Root Mean Square Propagation) by PyTorch, which exhibited robustness against artefacts. Optimisation was run for 3000 iterations with a learning rate of size 0.01.

Figure 6 shows the DIP architecture employed. We make use of skip connections, which are direct links between different parts of the convoluted network. They make information flow not only within the architectural structure but also outside of it, which allows an alternative gradient back-propagation path. This technique

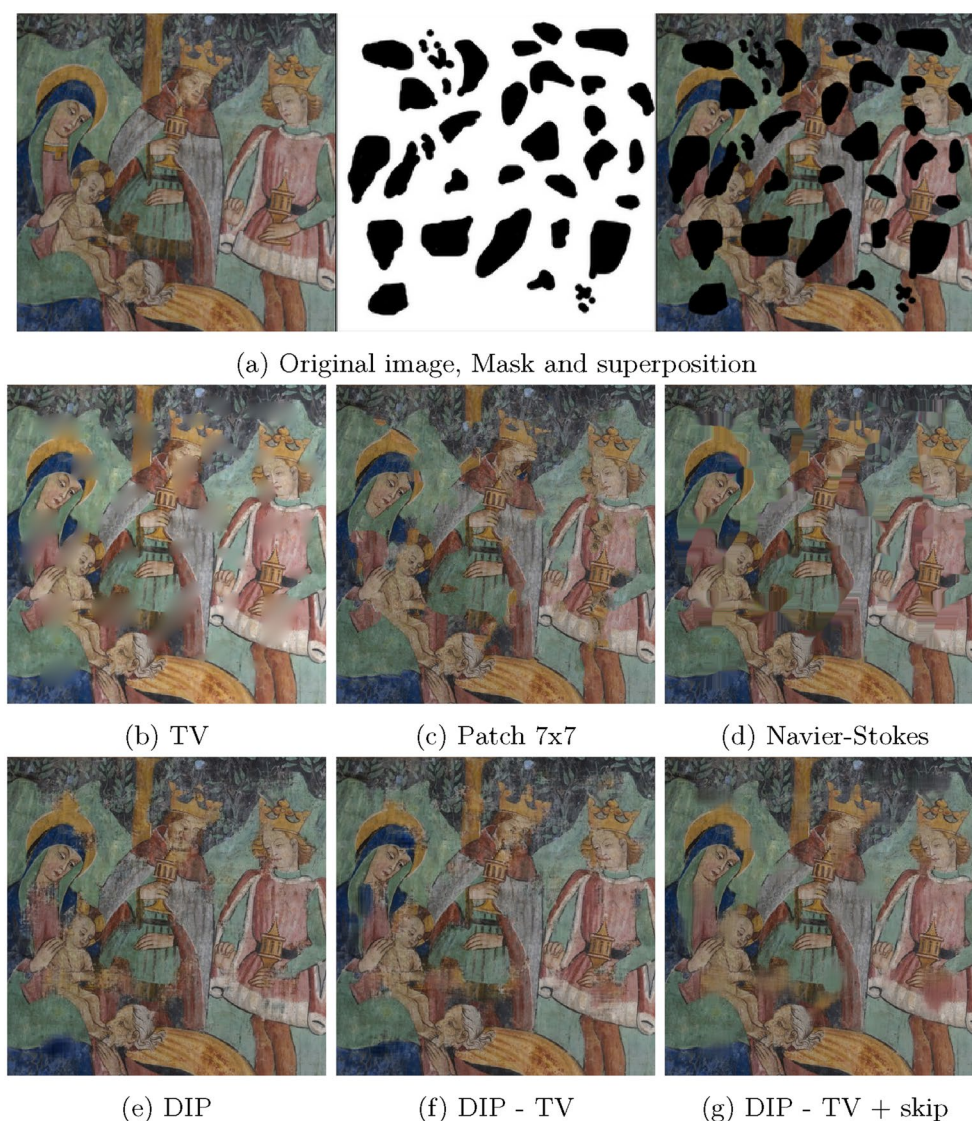


Fig. 9 Numerical study simulating the inpainting of an ancient fresco. On the top, the simulation setting with a hand-crafted mask. In the second and third rows, the images inpainted by different techniques, for a visual comparison

proved to be one of the most effective tools in improving the performance of convoluted networks, see, e.g., [54–56]. However, skip connections are typically viewed as disadvantageous in DIP, because they tend to allow structures to bypass the network’s architecture and it may lead to inconsistencies and smoothing effects, as outlined in [10]. In our specific scenario, on the other hand, such smoothing effect contributed positively to the overall consistency of the inpainted image. In Sect. [Numerical results](#), the usage benefits of skip connections will be discussed.

Inspired by previous work [38, 57, 58], we stabilised the training procedure (4) by further adding to the loss functional a TV regularisation term, thus considering:

$$\hat{\Theta} \in \operatorname{argmin}_{\Theta} \lambda \|m \odot (f_{\Theta}(z) - \bar{x})\|^2 + TV(f_{\Theta}(z)). \tag{5}$$

In comparison to (4), training under (5) reduces the sensitivity to the stopping time as the presence of TV (suitably balanced with the data term by λ) prevents noise overfitting.

Table 1 Quantitative assessment of inpainting methods applied to Fig. 9a

Quantitative inpainting assessment on synthetic data				
Model	SSIM	NRMSE	MSE	PSNR
Original image	1	0	0	∞
TV	0.82	2.44e-01	6.24e02	20.2
Navier-Stokes	0.83	1.11e-01	1.31e02	26.9
Patch 3x3	0.75	1.65e-01	2.89e02	23.5
Patch 5x5	0.76	1.46e-01	2.28e02	24.5
Patch 7x7	0.77	1.37e-01	2.00e02	25.1
DIP	0.81	1.17e-01	1.45e02	26.5
DIP - TV	0.81	1.21e-01	1.55e02	26.2
DIP - TV + skip	0.84	1.04e-01	1.15e02	27.5

Bold fonts denote best results

Experimental setup

The proposed inpainting workflow consists of three distinct steps. First, given an RGB to inpaint, we perform a basic pre-processing (i.e., resizing) to give it as an input to the DIP model, see Sect. [Image pre-processing](#). Next, a masking operator identifying the region to inpaint has to be defined, see Sect. [Mask detection](#). Lastly, both the input and the mask images are given as an input to the the DIP network whose weights are then optimised to produce the desired inpainting result.

Image pre-processing

The RGB images in the available dataset have different resolutions and have different quality. Some of them were taken for documentation purposes and are, generally, low quality. On the other hand, some were taken with high-resolution cameras for the visualisation of fine details. This makes the image dataset not homogeneous, which could be indeed a complication as the architecture neural networks for image reconstruction is typically fine-tuned typically for inputs of specific size and quality.

As discussed below in Sect. [DIP architecture and regularisation](#), the neural network considered in this work runs on square images, for which reason we chose a common image size of 512×512 pixels and used these rescaled data for inpainting. Note that the DIP approach considered requires indeed the whole occluded image as an input. The use of the proposed approach on (overlapping) image patches was therefore not considered in this work but could represent indeed an interesting direction of future research.

Mask detection

Computing the pixels in the input image that have to be inpainted is nothing but a binary image segmentation problem which can be handled separately by means of any available segmentation routine. Such procedure can be approached in different ways, depending on both how

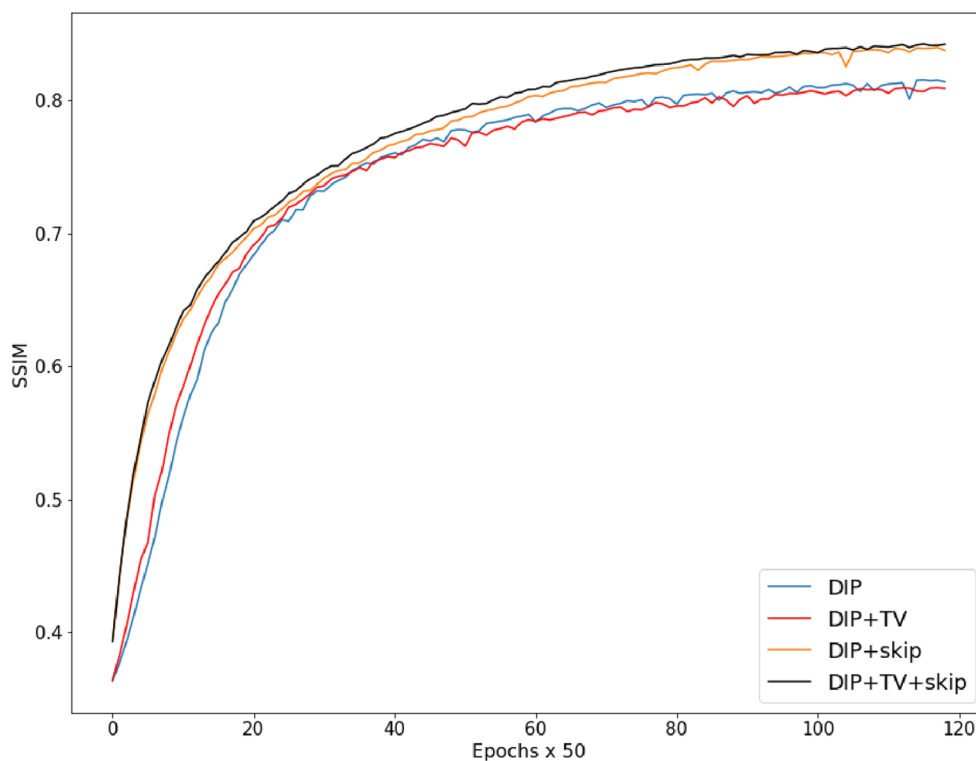


Fig. 10 Values of the SSIM metric over the training epochs, for four different configurations of the DIP approach

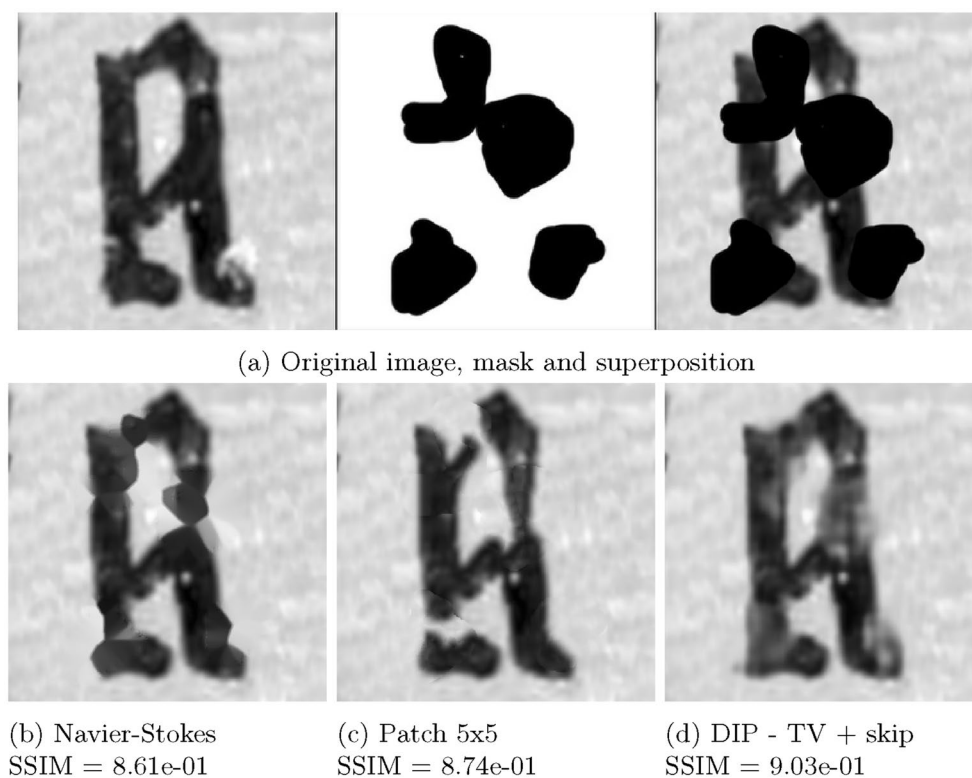


Fig. 11 Inpainting of "a" character with artificial mask

much automation one aims to implement and on how relevant the intervention of the restoration professional is. We describe in the following sections three techniques for mask detection falling into the category of automatic, semi-automatic and manual approaches. We stress that other approaches (based, e.g., on the use of deep learning based routines) could alternatively be used.

For several RGB images in the PAINT dataset under consideration, an effective segmentation was not possible due to difficulties in detecting the damaged areas. A valid tool to overcome this issue is the use of infrared (IR) imaging data, which is able to uncover *overpaints*, damages and previous restorations. The inpainting procedure can then be implemented either on the RGB image itself or possibly on the IR image, as schematically reported in Fig. 8 and discussed in the following section.

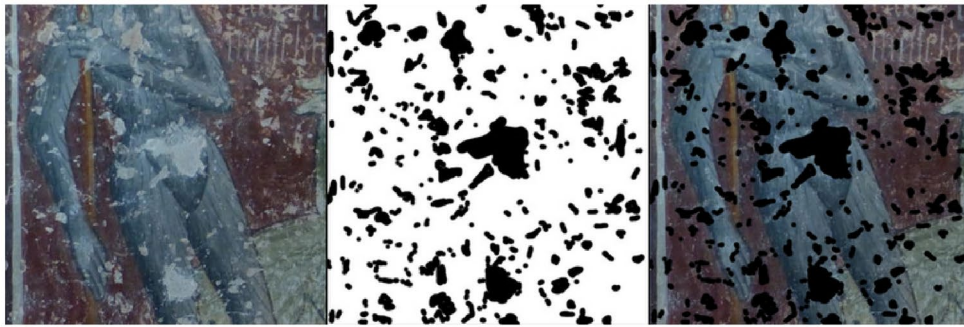
Automatic mask selection. For automatic mask selection we refer to a method where an algorithm takes as input a color, corresponding to the tone of the damaged areas, and automatically select all the pixels of that color (within a defined tolerance) in the entire image. For our results the threshold was defined on the composite of all three colour channels using GIMP [59]. Such procedure works effectively if the damaged areas have considerably distinguishable characteristics with respect to

the preserved content, and if this property is consistent throughout the image. If that is not the case and/or too much noise is present in the input data, precision may suffer.

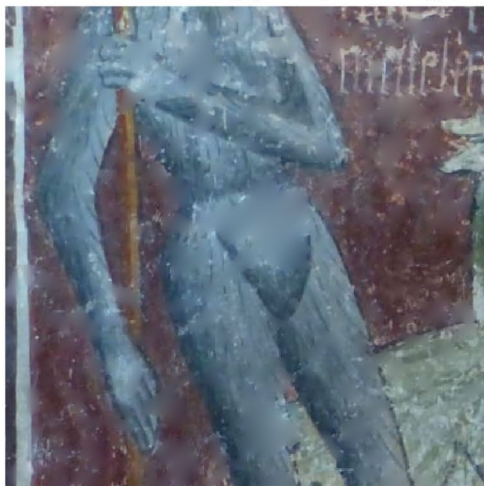
We found that this techniques was not precise enough for our purposes: additional pixels belonging to the undamaged areas were indeed wrongly detected, see, e.g., Fig. 7.

Semi-automatic mask selection. To prevent the mask from including pixels of the selected colour but not belonging to damages areas, we propose the semi-automatic mask creation. Unlike to the previous approach, it is done not only by providing a colour and a threshold, but also manually selecting one seed pixel for each connected region of the mask. Each region of the mask is then automatically detected by region growing from the selected pixel. Differently from the automatic technique, this approach allows for a better localization of large damages, but the seed selection may become challenging and potentially imprecise for small regions, as visible in Fig. 7.

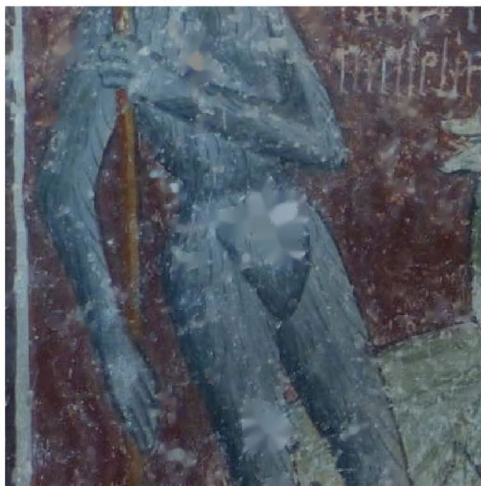
Manual mask selection. The manual mask selection process involves an expert user utilizing a paint tool to select the damaged areas. This technique is highly effective as it ensures complete coverage of the damage and



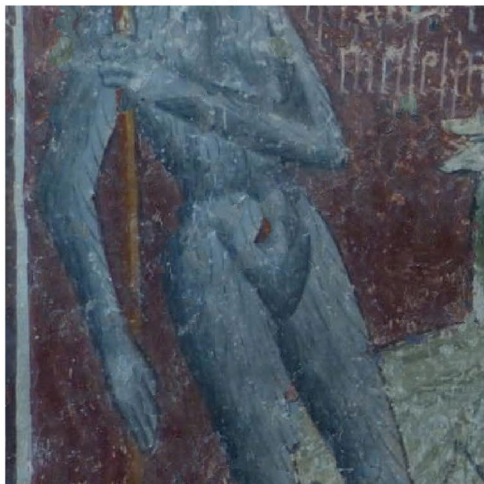
(a) Original image, Mask and superposition



(b) TV



(c) Navier-Stokes

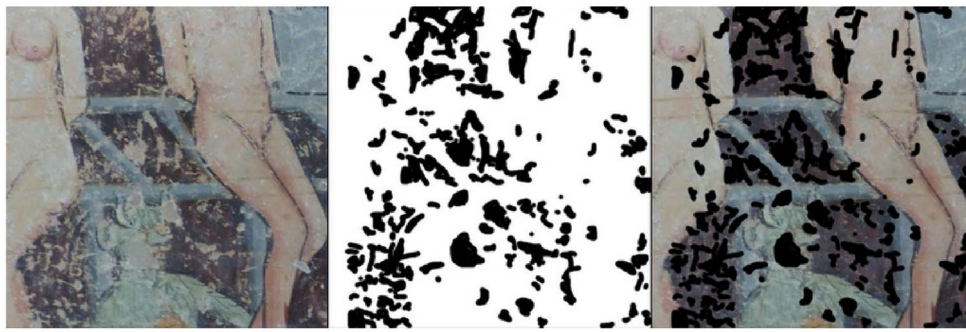


(d) Patch 5x5

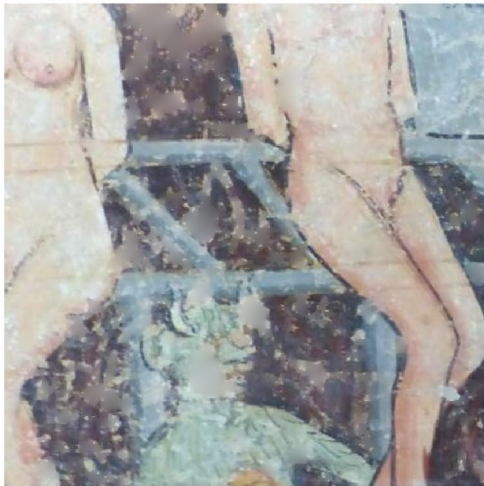


(e) DIP - TV + skip

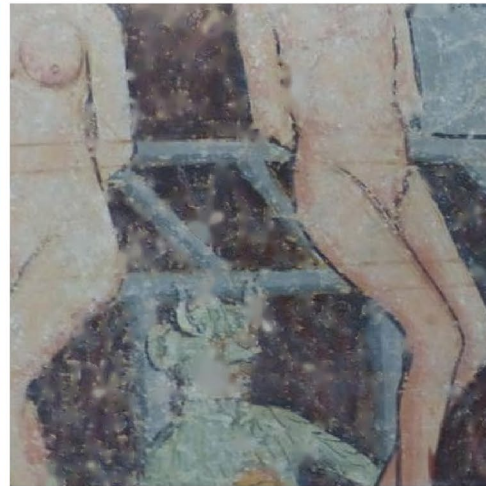
Fig. 12 Inpainting comparison on a detail from *Invidia*



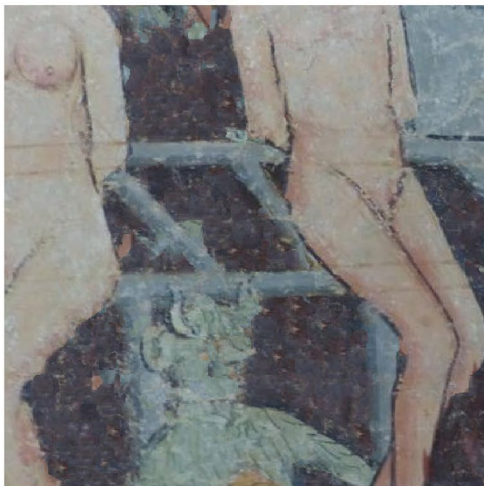
(a) Original image, Mask and superposition



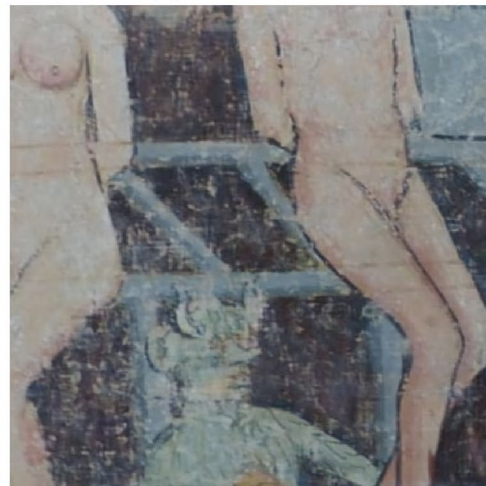
(b) TV



(c) Navier-Stokes



(d) Patch 5x5



(e) DIP - TV + skip

Fig. 13 Inpainting comparison on a detail from *Lusuria*

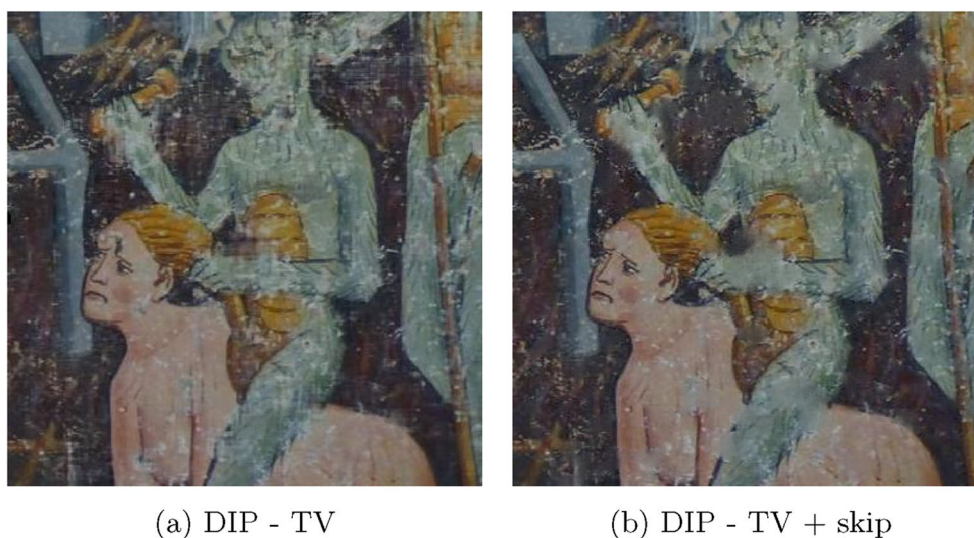


Fig. 14 Comparison of DIP based inpainting without and with skip connections, on a detail from *Lusura*

allows for a customized selection. By employing this method, we can address the problem of not fully covering the border areas and at the same not extending the mask excessively into the preserved image, as it usually happened with the previous selection methods. Leaving portions of the edges of the damaged areas outside the mask, produces discontinuities in the restored images, with a detrimental impact on the quality of the inpainting process. In our experimental setting, it proved to be the most effective approach in generating the highest quality masks. However, manual mask selection may become impractical due to the considerable amount of manual work involved.

Numerical results

In this Section, we show the results of the proposed DIP inpainting technique on some images from the PAINT dataset described in Sect. [Dataset description and challenges](#).

We compare the performance of our DIP approach trained using (5) (DIP-TV), with the baseline approach in [10] (DIP). Whenever skip connections are considered we add “+skip” to the corresponding approach. When we use TV regularization, the parameter λ has been heuristically chosen by minimizing the error metrics of by visual inspection.

The DIP-TV+skip solver is compared to state-of-art hand-crafted inpainting models. In particular, we considered the TV-regularisation method [16], the diffusive Navier-Stokes approach [21], and the patch-based non-local approach [29, 30] with patches of different sizes. We remark that fully data-driven inpainting approaches cannot be applied here, as they rely on the use of training

data (from the same painter, chapel...) that could not be obtained for our case. We ran our experiments on a Ryzen 5600 G CPU in tandem with an RTX 3060 GPU. Hand-crafted solvers run on CPU, whereas DIP methods operate on the GPU. Execution times range from approximately 1 s for *Navier–Stokes* to 32 s for the patch-based non-local approach with a 5x5 patch size, and 81 s for size 7x7. For complete convergence, the DIP methods take around 11 min. The higher-computational costs are justified by a better reconstruction performance. The code is available on GitHub at https://github.com/fmerizzi/Deep_image_prior_inpainting_of_ancient_frescoes

Validation on synthetic data

We start our numerical discussion presenting some inpainting results obtained from simulated data where an artificially created mask is super-imposed to a representative image in the dataset so to simulate occlusions/damages. We compare the results obtained by hand-crafted approaches and the proposed DIP method and evaluate quantitatively their performance using some standard error measures assessing the quality of the computed reconstruction against the original image. The original image, the binary mask and the simulated occluded image are reported in Fig. 9a. The inpainting results computed using the different methods discussed are reported below. Generally, we observe that the greater the inpainting region, the harder the reconstruction with possibly some non coherent content.

We quantitatively assess the reconstruction in terms of the Structural Similarity index (SSIM), the Mean Square Error (MSE), the Normalized Root Mean Square Error (NRMSE) and Peak Signal to Noise Ratio (PSNR).



(a) Original image, mask and superposition



(b) TV



(c) Navier-Stokes



(d) Patch 5x5



(e) DIP - TV + skip

Fig. 15 Inpainting comparison with a detail of *Lusuria* with both text and figurative parts

For all the reconstructions performed, these metrics are presented in Table 1. The computed results consistently highlight that the DIP-TV+skip combination attains the top scores.

To highlight the improvement provided by the technical modifications of the DIP scheme detailed in Sect. [DIP architecture and regularisation](#), in Fig. 10 we report the behavior of the SSIM metric over the training epochs, for

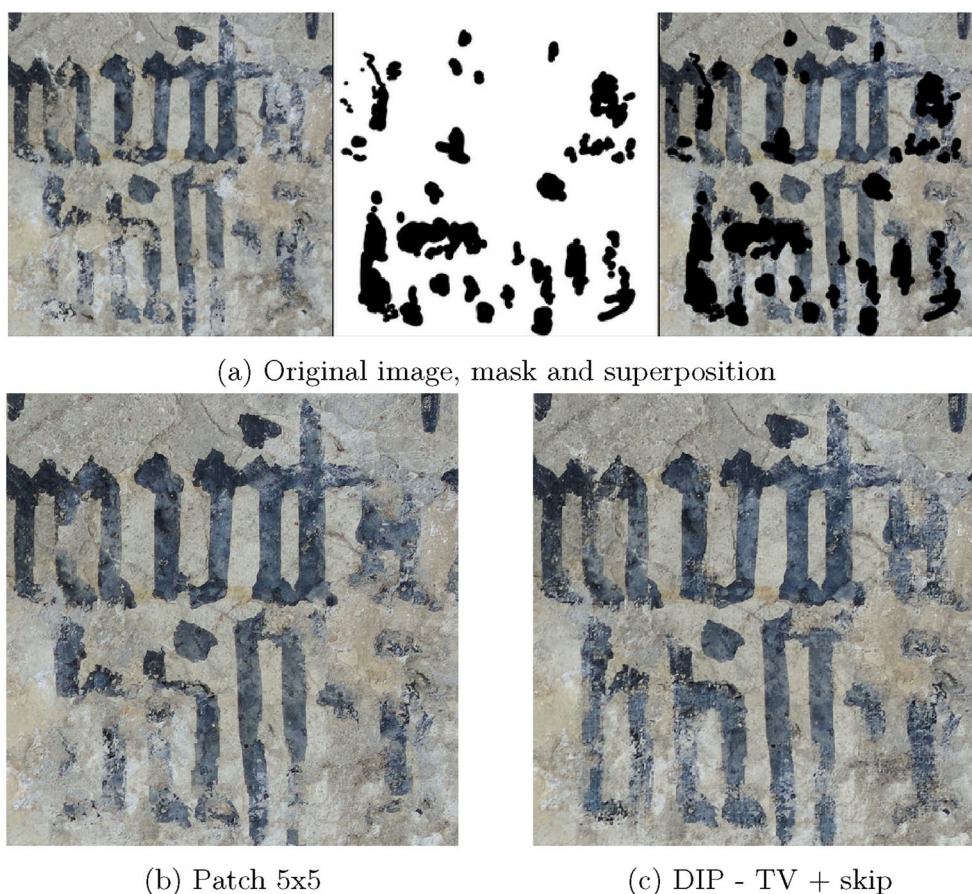


Fig. 16 Text inpainting comparison on a detail from the Venanson chapel

various DIP configurations. The naive DIP implementation shows lower SSIM values, in comparison to its versions including skip connections which improve the results throughout all epochs. We observe that the TV appears to enhance the quantitative results only marginally, although its presence stabilises the training process. For this reason we considered in the following the DIP-TV+skip combination to perform our tests.

We perform a similar simulation on a textual character of an “a” occluded with an artificially created large inpainting mask, see Fig. 11. We compare the solution obtained by DIP-TV+skip with the ones obtained by using the Navier-Stokes and Patch approaches. Both visually and in terms of SSIM we observe that the DIP approach better reconstructs the letter without spots or discontinuities (as in Fig. 11b-c), showing better visual coherence.

Comparison of inpainting techniques on digital pictures of degraded frescoes

In Figs. 12 and 13 we report a comparison between the reconstructions obtained by different inpainting methods

tested on the *Invidia* and *Lusuria* frescoes in Fig. 4, respectively.

We first consider a cropped image from *Invidia*, in Fig. 12. We note that the TV inpainted image is blurred in the larger damaged regions, whereas the Navier-Stokes image shows evident reconstruction artifacts and the image obtained by the non-local patch-based method is globally better, although a ghosting artifact appears in the largest inpainted area. The DIP-TV+skip inpainting result is the most visually satisfying reconstruction, with fewer artifacts and higher visual consistency. Similar considerations can be made when looking at the results reported in Figure 13.

We remark that the evaluation of results is here only qualitative due to the lack of ground truth images. Recalling reference works in imaging and vision such as [60, 61], the minimal property that should be guaranteed by any inpainting method is the so-called *good connection* property, i.e. the ability of connecting separated pieces of a curve (here, image level lines) in a coherent way. The approaches considered do satisfy this minimal property at least whenever the inpainting domain is sufficiently

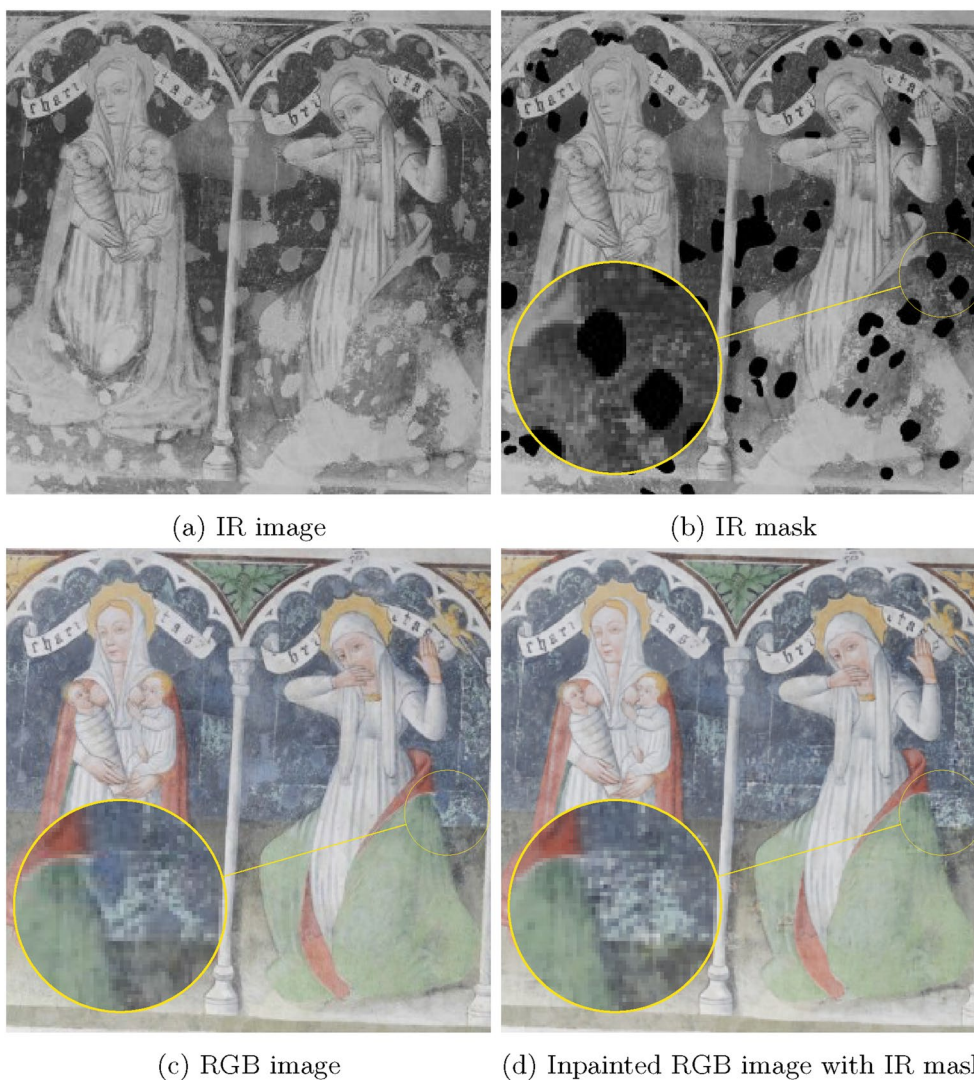


Fig. 17 DIP-TV + skip Inpainting on RGB image with IR mask

small. They are subject, however, to more variability in the reconstruction of oscillating content such as, e.g., texture.

In Fig. 14, we present a visual comparison of the inpainting process using DIP, both with and without skip connections. It is evident that incorporating skip connections results in smoother inpainted surfaces and fewer artifacts.

We now apply inpainting to restore textual images. The restoration of the textual detail in Fig. 15 is particularly interesting. Reliable inpainting approaches should indeed avoid any major modifications to image contents so as to guarantee a reliable, or even improved, interpretation of the artwork. In this respect, we observe that while local and non-local methods may alter the image

content, the DIP approach better preserves the desired text information with a higher level of precision.

Analogously, in Fig. 16 we provide a comparison of inpainting methods on a portion of damaged text from the Venanson chapel, where we observe that a more consistent text reconstruction is obtained by our DIP-TV+skip method.

Inpainting based on IR images

When an infrared image of a fresco is available, it may allow the discovery of under-drawings and under-writings not easily discernible within the visible spectrum, i.e. on the RGB image. In Fig. 17 we exploit such property by creating the mask of these regions using the IR image (Fig. 17a).

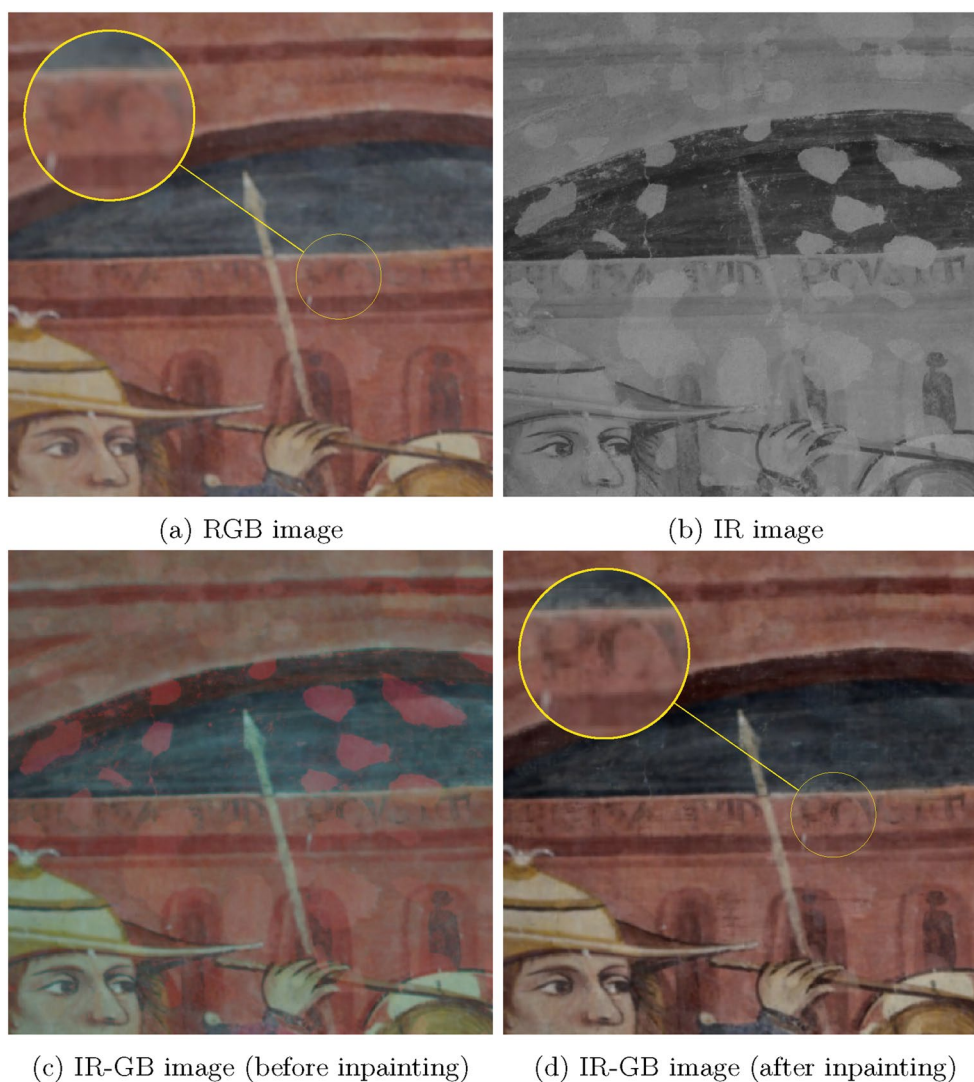


Fig. 18 Text enhancing by IR mask extraction. Inpainting is performed by DIP-TV + skip on the IR image. The inpainted IR image is then used as red channel for the original RGB image

Since the damaged areas are harder to detect (Fig. 17c), the mask has subsequently been super-imposed to the RGB picture of the fresco. DIP inpainting can there be applied so as to obtain the inpainted image shown in Fig. 17d. In such inpainting result the background looks very coherent to the remaining part of the fresco, thus providing probably a more faithful image of how the original fresco looked like before retouches.

Interestingly, in the “Mocking of Christ” painted by Pietro Guido, the IR data revealed ancient text appearing severely faded in the colour image (see Fig. 18a and b). The IR image can be embedded as the Red channel together with the original Green and Blue ones, so as to get the three channel image represented in 18c (denoted as IR-GB). In this case, the inpainting mask has been selected

on the IR picture and used to fill in the IR image directly, by our DIP-TV+skip method. We observe that, now, in the corresponding IR-GB image 18d the text appears more visible and interpretable than in the starting image 18a.

Discussion and outlook

In digital imaging, bringing back to light hidden and/or destroyed piece of information in ancient frescoes using techniques in the realm of variational methods and deep learning is often a very challenging task. The lack of reference data and the poor quality of both the fresco and of its digital representation often make hopeless the use of both standard approaches based on local reconstruction techniques and complex learning architectures relying on lots of training data.

In this paper, we consider the problem of image and text inpainting for images acquired in the Mediterranean Alpine arc (dataset PA'INT) and corrupted by severe degradations. The ultimate goal of this project is to ease the investigation of the actions taken by the authors toward painted images and their causes, which may emerge in a different context from the period of the artworks' creation. Intentional destruction and modifications are key aspects we seek to identify in this kind of study. For example, vandalism often targets images with negative connotations, such as devils and demons, leading to the loss of texts and visual representations. The retrieval of these elements is crucial for studying painted themes and patterns which are recurrent during the medieval period.

For such task, we applied the Deep Image Prior Inpainting procedure introduced in [10] stabilized as in [58] as a hybrid technique relying on the expressivity of (an untrained) neural network and on its interpretability as a non-convex variational approach based on iterative regularisation. By using as a training image the sole given data, improved reconstructions are obtained in the occluded/damaged areas. In comparison with classical approaches, the results computed show less artefacts and favour better interpretability of the data by art historians.

Furthermore, when combined with additional infrared data, the proposed techniques integrate and restore image contents effectively thus providing useful piece of information for subsequent analysis.

Through this interdisciplinary project combining art history, mathematical image processing, and AI, we aim to better understand the historical data and later interventions on medieval images. By doing so, we hope to chronicle the life of the paintings and gain insights into their impact and evolution within past societies.

Author contributions

OA and PS collected the data. FM developed the computational methods and processed the data. FM, PS, EM and LC analysed the results. PS, RMD provided the historical and artistic background for the project. PS, FM, EM, ELP and LC wrote the manuscript.

Funding

PS, FM, OA, RMD and LC acknowledge the financial support received by the CNRS project PRIME Imag'In and the UCA project Arch-AI-story. LC and EM acknowledge the support received by the Academy 1 of UCA, program IDEX JEDI for invited researchers. LC acknowledges the support received by the ANR JCJC project TASKABILE (ANR-22-CE48-0010). Research partially supported by the Future AI Research (FAIR) project of the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.3 funded from the European Union - NextGenerationEU.

Availability of data and materials

The datasets analysed during the current study are available in the PA'INT [62] repository. The source code used for DIP inpainting is openly accessible in a dedicated GitHub repository [63].

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 24 June 2023 Accepted: 12 December 2023

Published online: 05 February 2024

References

- Dessi RM. Spectres d'art du Trecento: à propos de quelques peintures de personnages couronnés (Giotto, Simone Martini, Lippo Memmi et Ambrogio Lorenzetti). *Images Re-Vues Hist Anthropol Théorie Art*. 2018. <https://doi.org/10.4000/imagesrevues.5461>.
- Acquier O, Pasqualini A. Base de données (SQL) : Peintures murales du sud de l'Arc alpin associant des Images et des Textes (2022). <https://doi.org/10.34847/nkl.916b60t3>
- Bertalmio M, Sapiro G, Caselles V, Ballester C. Image inpainting. In: *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '00*, pp. 417–424. ACM Press/Addison-Wesley Publishing Co, USA (2000). <https://doi.org/10.1145/344779.344972>
- Fornasier M, March R. Restoration of color images by vector valued by functions and variational calculus. *SIAM J Appl Mathemat*. 2007;68(2):437–60.
- Baatz W, Fornasier M, Markowich P, Schönlieb C-B. Inpainting of ancient austrian frescoes. In: *Proceedings of Bridges*, 2008; pp. 150–156
- Calatroni L, d'Autume M, Hocking R, Panayotova S, Parisotto S, Ricciardi P, Schönlieb C-B. Unveiling the invisible: mathematical methods for restoring and interpreting illuminated manuscripts. *Heritage Sci*. 2018;6:1–21.
- Bugeau A, Bertalmio M, Caselles V, Sapiro G. A comprehensive framework for image inpainting. *IEEE Trans Image Process*. 2010;19(10):2634–45. <https://doi.org/10.1109/TIP.2010.2049240>.
- Schönlieb C-B. *Partial Differential Equation Methods for Image Inpainting*. Cambridge: Cambridge University Press; 2015.
- Ballester C, Bugeau A, Hurault S, Parisotto S, Vitoria P. An analysis of generative methods for multiple image inpainting. *arXiv*. 2022. <https://doi.org/10.48550/arXiv.2205.02146>.
- Ulyanov D, Vedaldi A, Lempitsky V. Deep image prior. *Int J Comp Vision*. 2020;128(7):1867–88. <https://doi.org/10.1007/s11263-020-01303-4>.
- Acquier O. *Écriture épigraphique et sermons dans les peintures murales des lieux de culte du sud de l'arc alpin du XVe au XVIe siècle (Provence orientale, Ligurie, Piémont)*. PhD thesis, Université Côte d'Azur (2021)
- Galli R. EOS For Astro. EOS for Astro (2021). <https://www.eosforastro.com/>
- Cosentino A. *Technical Photography*. Cultural Heritage Science Open Source (2023). <https://chsopensource.org/product-category/technical-photography/>
- Boust C, et al. Images scientifiques pour le patrimoine. *Hypothèse* (2015)
- Elharrouss O, Almaadeed N, Al-Maadeed S, Akbari Y. Image inpainting: a review. *Neural Proc Lett*. 2020;51:2007–28.
- Chan TF, Shen J. Nontexture inpainting by curvature-driven diffusions. *J Visual Commun Image Represent*. 2001;12(4):436–49. <https://doi.org/10.1006/jvci.2001.0487>.
- Papafitsoros K, Schönlieb CB. A combined first and second order variational approach for image reconstruction. *J Mathemat Imag Vision*. 2014;48(2):308–38.
- Caselles V, Morel J-M, Sbert C. An axiomatic approach to image interpolation. *IEEE Trans Image Proc*. 1998;7(3):376–86. <https://doi.org/10.1109/83.661188>.
- Bertalmio M, Bertozzi AL, Sapiro G. Navier-stokes, fluid dynamics, and image and video inpainting. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, p. (2001). IEEE
- Telea A. An image inpainting technique based on the fast marching method. *J Graph Tools*. 2004;9(1):23–34. <https://doi.org/10.1080/10867651.2004.10487596>.

21. Bertalmio M, Bertozzi A, Sapiro G. Navier-stokes, fluid dynamics, and image and video inpainting, 2001; vol. 1, p. 355. <https://doi.org/10.1109/CVPR.2001.990497>
22. Ballester C, Bertalmio M, Caselles V, Sapiro G, Verdera J. Filling-in by joint interpolation of vector fields and gray levels. *IEEE Trans Image Proc.* 2001;10(8):1200–11. <https://doi.org/10.1109/83.935036>.
23. Chan TF, Shen J. Nontexture inpainting by curvature-driven diffusions. *J Visual Commun Image Represent.* 2001;12(4):436–49.
24. Masnou S, Morel J-M. Level lines based disocclusion. In: *Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No.98CB36269)*, pp. 259–2633 (1998). <https://doi.org/10.1109/ICIP.1998.999016>
25. Criminisi A, Pérez P, Toyama K. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans Image Proc.* 2004;13(9):1200–12. <https://doi.org/10.1109/TIP.2004.833105>.
26. Aujol J-F, Ladjal S, Masnou S. Exemplar-based inpainting from a variational point of view. *SIAM J Mathemat Anal.* 2010;42(3):1246–85. <https://doi.org/10.1137/080743883>.
27. Arias P, Facciolo G, Caselles V, Sapiro G. A variational framework for exemplar-based image inpainting. *Int J Comp Vision.* 2011;93(3):319–47.
28. Barnes C, Shechtman E, Finkelstein A, Goldman DB. Patchmatch: a randomized correspondence algorithm for structural image editing. *ACM Trans Graph.* 2009;28(3):24.
29. Newson A, Almansa A, Fradet M, Gousseau Y, Pérez P. Video inpainting of complex scenes. *SIAM J Imaging Sci.* 2014;7(4):1993–2019. <https://doi.org/10.1137/140954933>.
30. Newson A, Almansa A, Gousseau Y, Pérez P. Non-Local Patch-Based Image inpainting. *Image Proc On Line.* 2017;7:373–85. <https://doi.org/10.5201/ijol.2017.189>.
31. Oncu AI, Deger F, Hardeberg JY. Evaluation of digital inpainting quality in the context of artwork restoration. In: *Fusiello A, Murino V, Cucchiara R, editors. Computer Vision - ECCV 2012. Workshops and Demonstrations.* Berlin, Heidelberg: Springer; 2012. p. 561–70.
32. Köhler R, Schuler CJ, Schölkopf B, Harmeling S. Mask-specific inpainting with deep neural networks. In: *German Conference on Pattern Recognition (2014)*
33. Pathak D, Krahenbuhl P, Donahue J, Darrell T, Efros AA. Context encoders: Feature learning by inpainting. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2536–2544. IEEE Computer Society, Los Alamitos, CA, USA (2016). <https://doi.org/10.1109/CVPR.2016.278>. <https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.278>
34. Liu G, Reda FA, Shih KJ, Wang T-C, Tao A, Catanzaro B. Image inpainting for irregular holes using partial convolutions. In: *Ferrari V, Hebert M, Sminchisescu C, Weiss Y, editors. Computer Vision - ECCV 2018.* Cham: Springer; 2018. p. 89–105.
35. Wang Y, Tao X, Qi X, Shen X, Jia J. Image inpainting via generative multi-column convolutional neural networks. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems. NIPS'18*, pp. 329–338. Curran Associates Inc., Red Hook, NY, USA (2018)
36. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. In: *Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems, vol. 27.* Curran Associates, Inc., ??? (2014). NIPS
37. Iizuka S, Simo-Serra E, Ishikawa H. Globally and locally consistent image completion. *ACM Trans Graph.* 2017. <https://doi.org/10.1145/3072959.3073659>.
38. Liu H, Jiang B, Xiao Y, Yang C. Coherent semantic attention for image inpainting. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4169–4178. IEEE Computer Society, Los Alamitos, CA, USA (2019). <https://doi.org/10.1109/ICCV.2019.00427>.
39. Liu H, Wan Z, Huang W, Song Y, Han X, Liao J. Pd-gan: Probabilistic diverse gan for image inpainting. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9371–9381 (2021)
40. Lahiri A, Jain AK, Agrawal S, Mitra P, Biswas PK. Prior guided gan based semantic inpainting. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020)*
41. Hedjazi MA, Genc Y. Efficient texture-aware multi-gan for image inpainting. *Knowledge-Based Syst.* 2021;217: 106789. <https://doi.org/10.1016/j.knosys.2021.106789>.
42. Ren Y, Yu X, Zhang R, Li TH, Liu S, Li G. Structureflow: Image inpainting via structure-aware appearance flow. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 181–190. IEEE Computer Society, Los Alamitos, CA, USA (2019). <https://doi.org/10.1109/ICCV.2019.00027>.
43. Xiong W, Yu J, Lin Z, Yang J, Lu X, Barnes C, Luo J. Foreground-aware image inpainting. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5833–5841. IEEE Computer Society, Los Alamitos, CA, USA (2019). <https://doi.org/10.1109/CVPR.2019.00599>.
44. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. *Adv Neural Inform Proc Syst.* 2020;33:6840–51.
45. Goodfellow IJ, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. *arXiv.* 2015. <https://doi.org/10.48550/arXiv.2205.02146>.
46. Lugmayr A, Danelljan M, Romero A, Yu F, Timofte R, Van Gool L. Repaint: Inpainting using denoising diffusion probabilistic models. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11461–11471 (2022)
47. Chen L, Zhou L, Li L, Luo M. Crackdiffusion: crack inpainting with denoising diffusion models and crack segmentation perceptual score. *Smart Mater Struct.* 2023;32(5): 054001.
48. Wang S, Saharia C, Montgomery C, Pont-Tuset J, Noy S, Pellegrini S, Onoe Y, Laszlo S, Fleet DJ, Soricut R, Baldridge J, Norouzi M, Anderson P, Chan W. Imagen editor and editbench: Advancing and evaluating text-guided image inpainting. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18359–18369 (2023)
49. Li W, Lin Z, Zhou K, Qi L, Wang Y, Jia J. Mat: Mask-aware transformer for large hole image inpainting. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10758–10768 (2022)
50. Suvorov R, Logacheva E, Mashikhin A, Remizova A, Ashukha A, Silvestrov A, Kong N, Goka H, Park K, Lempitsky V. Resolution-robust large mask inpainting with fourier convolutions. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 2149–2159 (2022)
51. Wang N, Wang W, Hu W, Fenster A, Li S. Thangka mural inpainting based on multi-scale adaptive partial convolution and stroke-like mask. *IEEE Trans Image Proc.* 2021;30:3720–33. <https://doi.org/10.1109/TIP.2021.3064268>.
52. Lv C, Li Z, Shen Y, Li J, Zheng J. Separafill: two generators connected mural image restoration based on generative adversarial network with skip connect. *Heritage Sci.* 2022;10(1):135.
53. Deng X, Yu Y. Ancient mural inpainting via structure information guided two-branch model. *Heritage Sci.* 2023;11(1):131.
54. Drozdal M, Vorontsov E, Chartrand G, Kadoury S, Pal C. The importance of skip connections in biomedical image segmentation. In: *Carneiro G, Mateus D, Peter L, Bradley A, Tavares JMRS, Belagiannis V, Papa JP, Nascimento JC, Loog M, Lu Z, Cardoso JS, Cornebie J, editors. Deep Learning and Data Labeling for Medical Applications.* Cham: Springer; 2016. p. 179–87.
55. Orhan E, Pitkow X. Skip connections eliminate singularities. In: *International Conference on Learning Representations (2018)*. <https://openreview.net/forum?id=HkwBEMWCZ>
56. Evangelista D, Morotti E, Piccolomini EL, Nagy J. Ambiguity in solving imaging inverse problems with deep-learning-based operators. *J Imaging.* 2023. <https://doi.org/10.3390/jimaging9070133>.
57. Cascarano P, Sebastiani A, Comes MC, Franchini G, Porta F. Combining weighted total variation and deep image prior for natural and medical image restoration via ADMM. In: *2021 21st International Conference on Computational Science and Its Applications (ICCSA)*, pp. 39–46 (2021). <https://doi.org/10.1109/ICCSA54496.2021.00016>
58. Liu J, Sun Y, Xu X, Kamilov US. Image restoration using total variation regularized deep image prior. In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7715–7719 (2019). <https://doi.org/10.1109/ICASSP.2019.8682856>
59. The GIMP Development Team: GIMP. <https://www.gimp.org>
60. Sha'asua A, Ullman S. Structural saliency: The detection of globally salient structures using a locally connected network. In: *[1988 Proceedings] Second International Conference on Computer Vision*, pp. 321–327 (1988). <https://doi.org/10.1109/CCV.1988.590008>
61. Desolneux A, Moisan L, Morel J-M. From Gestalt Theory to Image Analysis: A Probabilistic Approach. *Interdisciplinary Applied Mathematics*, 2008; vol. 34. Springer, ??? . <https://doi.org/10.1017/CBO9780511734304>
62. <https://www.cepam.cnrs.fr/datas/paint.php>. Accessed 24 Jun 2023

63. Merizzi F. https://github.com/fmerizzi/Deep_image_prior_inpainting_of_ancient_frescoes. Accessed 24 Jun 2023.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.