Abstract
This paper proposes an iterative score-based generative model for solving the automatic colorization problem. Although unsupervised learning methods have shown the capability to generate plausible color, inadequate exploration of detailed information and data dimensions still limit the performance of the colorization model. Considering that the number of samples in score-based generative model has influence on estimating the target gradients and the gradient map possesses important latent information of the image, the inference process of the generative modeling is conducted in joint intensity–gradient domain for colorization. Specifically, a set of intensity–gradient formed high-dimensional tensors are trained, via the score matching, to attain the gradient of data distribution in joint intensity–gradient domain. As the score function is determined, data samples are generated by means of annealed Langevin dynamics, forming an iterative colorization procedure. Furthermore, the joint intensity–gradient constraint in data-fidelity term is proposed to limit the degree of freedom within generative model at the iterative colorization stage, thus being conducive to edge-preserving colorization effect. Experimental results conveyed the remarkable performance and diversity of our proposed method.

Keywords
Automatic colorization · Unsupervised learning · Generative model · Intensity–gradient domain · Gradient constraint

1 Introduction
Colorization is the task of assigning color to each pixel of the target grayscale image. It has abundant applications in a variety of computer vision tasks, such as image compression [1–4], outline and cartoon creation colorization [5, 6], and remote sensing images colorization [7, 8]. Practically, coloring grayscale image is hard to address because the colorization results could be multi-viable for a single grey-level image [9]. Different colors can be assigned to the gray pixels of an input image, making the problem ill-posed and inherently ambiguous. For example, the color of an apple can be red, green or irrational purple. Besides, the colorization problem also inherits the challenges of image enhancement, such as changes in illumination, variations in viewpoints, and occlusions. Therefore, colorization is still confronting challenges in image processing and computer vision.

At present, many methods have been developed to tackle the issues of colorization, which can be roughly divided into two categories: Traditional interactive methods [10–17] and automatic methods based on deep learning [18–29]. Generally speaking, traditional scribble-based methods require users to provide considerable scribbles, then determine the colors of the remaining pixels automatically [11]. The results of the colorization depend upon how the color scribbles are chosen. However, the scribbling process is time-consuming and requires expertise. Example-based colorization is another reliable interactive method, which automatically searches for the most similar image patch or pixel for colorization [12]. The key point to the success of example-based colorization is how similar the reference image is to the grayscale image. Regrettifully, finding a suitable reference image becomes an obstacle for users.

To address these limitations, researchers have also explored deep learning in supervised or unsupervised manners for automatic colorization. These methods have achieved fantastic success in large-scale data. In addition, some methods introduced auxiliary tasks like image segmentation and fusion [23], auxiliary input text [13, 14], etc., to improve the colorization performance. Larsson et al. [27] exploited a self-supervised visual representation learning
contains both low-level and semantic representations to predict per-pixel color histograms. Iizuka et al. [56] presented an end-to-end fashion to automatically colorize grayscale images that combines both global priors and local image features. However, one major challenge of supervised learning is the requirement of abundant reference images with labels for network training. Unsupervised learning is widely considered as the future direction of the colorization problem. It has the characteristics of obtaining implicit prior information and generating natural coloring results with diversity [30]. For instance, Vitoria et al. [25] proposed an automatic end-to-end adversarial approach that joins the advantages of generative adversarial networks (GANs) with semantic class distribution learning.

Most existing automatic colorization methods are usually optimized for intensity-level information. On this basis, some methods try to extract semantic level information to improve colorization performance [23, 31]. Unfortunately, this kind of high-level image understanding is difficult to be captured and may result in unsuitable effects. In the colorization task, the information captured from the grayscale image itself is limited, thus the exploration of more valid object is urgent. In this work, we attempt to extract the features in gradient domain of the grayscale image as “guidance” information.

On the one hand, recent study has shown that the human visual system cannot perceive absolute pixel values but relies on the local contrast ratio [32], which is directly related to the gradient information. Thus, learning the gradient distribution of natural images can improve the subjective quality of images. Gradient domain information has also been widely applied in many works in image processing and has achieved great success in image enhancement [33–35], image restoration [36, 37], and image fusion [38], etc. On the other hand, as one of the basic information of the image, conducting unsupervised learning in gradient domain can expand the dimensionality of the data density distribution. Especially, this strategy adds more prosperous prior information on the learning and inference processes, which adapts to the characteristics of the unsupervised learning process. For instance, Pan et al. [39] proposed a simple yet effective L0-regularized prior based on intensity and gradient for text image deblurring. They pointed out this image prior was considered as the future direction of the colorization problem. However, one major challenge of supervised learning is the requirement of abundant reference images with labels for network training. Unsupervised learning is widely considered as the future direction of the colorization problem. It has the characteristics of obtaining implicit prior information and generating natural coloring results with diversity [30]. For instance, Vitoria et al. [25] proposed an automatic end-to-end adversarial approach that joins the advantages of generative adversarial networks (GANs) with semantic class distribution learning.

Most existing automatic colorization methods are usually optimized for intensity-level information. On this basis, some methods try to extract semantic level information to improve colorization performance [23, 31]. Unfortunately, this kind of high-level image understanding is difficult to be captured and may result in unsuitable effects. In the colorization task, the information captured from the grayscale image itself is limited, thus the exploration of more valid object is urgent. In this work, we attempt to extract the features in gradient domain of the grayscale image as “guidance” information.

On the one hand, recent study has shown that the human visual system cannot perceive absolute pixel values but relies on the local contrast ratio [32], which is directly related to the gradient information. Thus, learning the gradient distribution of natural images can improve the subjective quality of images. Gradient domain information has also been widely applied in many works in image processing and has achieved great success in image enhancement [33–35], image restoration [36, 37], and image fusion [38], etc. On the other hand, as one of the basic information of the image, conducting unsupervised learning in gradient domain can expand the dimensionality of the data density distribution. Especially, this strategy adds more prosperous prior information on the learning and inference processes, which adapts to the characteristics of the unsupervised learning process. For instance, Pan et al. [39] proposed a simple yet effective L0-regularized prior based on intensity and gradient for text image deblurring. They pointed out this image prior was based on distinctive properties of text images and did not require any heuristic edge selection methods which were critical to state-of-the-art edge-based deblurring methods.

In this work, we study how to effectively employ generative model in joint intensity–gradient domain to improve the performance of automatic colorization and the visual perception of colorized images. As far as we know, gradient information has already been exploited in supervised learning manner for colorization [40], but only enforced in loss function. The application of prior information in gradient domain to unsupervised automatic colorization is still a blank area worth exploring.

This study focuses on the field of automatic image colorization and presents a novel Joint intensity–gradient domain guided Generative Model for satisfactory colorization, coined JGM. In this work, we approach this task from the viewpoint of generative modeling by annealed Langevin dynamics [41]. At first, score matching is used to estimate the gradients of data distribution in joint intensity–gradient domain. Then, the colorization problem can be viewed as an image generation procedure that enforced with linear constraints to generate a satisfactory result. The rationality of the underlying idea in JGM is listed in Fig. 1. In Fig. 1b, the generative model in joint intensity–gradient field achieves a more colorful result than only in intensity domain. Moreover, by enforcing linear constraints in both intensity and gradient domains, the degree of freedom of the generative model is better exploited, thus the colorization result in Fig. 1c exhibits more natural effects.

The major contributions of this paper are:

- **Iterative generative modeling in joint intensity–gradient domain**: A novel automatic colorization via score-based generative modeling is used for exploring the prior information in joint intensity–gradient domain. Learning prior knowledge in redundant and high-dimensional subspace paves the way for producing more chances to attain diversity and possible colorization.
- **Enforcing linear constraints in image intensity–gradient field**: Aid by the linearity property of the gradient operator, linear data-fidelity constraint is also enforced in gradient domain. The constraints in joint intensity–gradient domain leverage the visual appearance of results, leading to preferable colorization in high-dimensional space.

### 2 Related work

#### 2.1 Colorization

Traditional interactive colorization methods depend on user-provided guidance, such as color scribbles [11, 13, 14] or reference images [12, 15–17], which are the key point of obtaining the satisfactory result of colorization. These methods can be roughly divided into two categories: scribble-based and example-based.

Scribble-based methods, such as Levin et al. [11], consider colorization with interactive user scribbles, which require manually specifying desired colors of certain regions. The user has to apply several color scribbles to an image and the colors are then propagated through the image by means of
Fig. 1 Visualization of the colorization results with different generative models. a The reference grayscale image. b The top line is the colorization result with generative modeling in intensity domain, and the bottom line is with the generative modeling in joint intensity–gradient domain. c The colorization result of JGM. Particularly, the generative modeling of JGM is conducted in 9-channel (intensity–gradient) domain, largely reduces the color ambiguity in intensity and attains a more natural and realistic result

minimizing a quadratic cost function. In general, scribble-based methods formulate the colorization as a constrained optimization problem that propagates user-specified color scribbles based on some low-level similarity metrics.

Unlike the scribble-based methods, example-based colorization methods compute the correspondences between the reference and input image based on some low-level similarity metrics. Transfer-based methods rely on availability of related reference image, from which color is transferred to the target grayscale image. Mapping between source and target is established automatically, using correspondences between local descriptors, or in combination with manual intervention. Welsh et al. [12] proposed the first automatic colorization method using a color transfer approach from a reference color image to a destination grayscale image. The method transfers the chromatic information from the reference image to the destination image based upon local matches of the weighted average of a pixel’s luminance and neighborhood statistics which keeps the luminance of the destination image unchanged. Irony et al. [58] developed a colorization method that takes account of the context of pixels rather than attempting to colorize a pixel based upon its neighborhood statistics alone. The approach first segments the reference color image by using a supervised classification scheme, then a mapping is made between small neighborhood areas and points in feature space. The performance of these methods is highly dependent on how similar the reference image is to the input grayscale image. However, finding a suitable reference image is also a severe obstacle.

Deep learning for automatic approaches has shown the ability to capture more intricate color properties on a huge amount of grayscale and color image pairs [18–29]. Cheng et al. [20] proposed the first deep neural network model for fully automatic image colorization. It used adaptive image clustering technique to incorporate the global image information through joint bilateral filtering. Zhang et al. [24] utilized the self-supervised classification network that has learned color probability distributions to achieve automatic colorization in Lab color space. Besides, generative model [43, 44] is essentially an unsupervised learning tool to describe the dataset. ChromaGAN [25] exploited an adversarial learning colorization approach coupled with semantic information. Messaoud et al. [26] developed a conditional random field-based variational autoencoder formulation to achieve diversity. The generative model is capable of learning probability distributions or adversarial generation over different color spaces of data and has been used for many tasks.

2.2 Score-based generative model

Lately, Song et al. [41] introduced a new score-based generative model named noise conditional score networks (NCSN). The model estimates the gradients of data distribution by denoising score matching (DSM) [42] and then produces samples progressively via Langevin dynamics. The
framework allows flexible model architectures, avoids using adversarial methods and does not require sampling during training. Besides, it also provides a learning objective that can be used for principled model comparisons.

Score matching is to learn a non-normalized statistical model through samples of unknown data distribution. One of the best advantages of score matching is that one can directly train a score network \( s_0(x) \) to estimate score function \( \nabla_x \log p(x) \) without training a model to estimate \( p(x) \) first.

Denoising score matching (DSM) is a variant of score matching which completely circumvents first. To the following:

Although NCSN has achieved satisfactory results, there are still two major deficiencies. Firstly, it is unable to fully utilize the detailed image information only with prior learning in intensity domain, while learning the gradient distribution of natural images can improve the subjective quality of images. Second, another deficiency is that low data density regions and the manifold hypothesis [60, 61]. In regions of low data density, the original DSM may not have enough evidence to estimate score functions accurately, due to the lack of data samples. Thus, to further enriches the efficiency of DSM, we introduce a high-dimensional joint intensity–gradient strategy to boost the representation diversity of generative modeling that favors avoiding falling into local optima. The main idea depends on Theorem 1.

**Theorem 1** [59]. Let \( F \) be a class of \( \mathbb{R}^d \)-valued functions, all of which are \( M/2 \)-Lipschitz, bounded coordinate wise by \( R > 0 \), containing arbitrarily good approximations of \( \nabla \log p_\sigma \) on the ball of radius \( R \). Let \( \sigma < \sigma_{\text{max}} \) and support we have \( n \) \( i.i.d \) samples from \( p_\sigma, x_1, \ldots, x_n \). Let

\[
\hat{s} \in \arg\min_{S \in F} \frac{1}{n} \sum_{i=1}^{n} \| s(x_i) - \nabla \log p_\sigma(x_i) \|_2^2
\]

Then with probability at least \( 1 - 4\delta - C n e^{-R^2/n \sigma} \) on the randomness due to the sample,

\[
E_{p_\sigma}[[||\hat{s}(x) - \nabla \log p_\sigma(x)||_2]^2] \leq C (\log^3 n \cdot \beta_n^2(F) + \beta_n d)
\]

where \( C \) is a universal constant. Both \( \beta_n^2(F) \) and \( \beta_n \) contain the factor of \( 1/n \).

From Theorem 1, it is concluded that the representation boundary is related to the sample number \( n \) and space dimension. Particularly, the larger the sample number is, the less the representation error will be. This observation motivates us to boost the performance via sampling in high-dimensional space. The high-dimensional tensor solves the problem that DSM losses the estimated accuracy due to insufficient samples by increasing the number of samples.

Besides, from the perspective of generative modeling, unlike the traditional generative models that often encode the input to be a latent variable in lower-dimensional spaces (e.g., CVAE [28, 45]), JGM is exploited in multi-feature spaces (e.g., Intensity domain, gradient domain) jointly. It benefits from the high-dimensional inference process of generative modeling. It is different from the score-based way of iGM [46] which is used in multi-color space (for example, RGB, Ycbcr) in combination. The progressive process of iGM is to obtain the information of Ycbcr color space, which is a time-consuming and inefficient coloring algorithm. It should be mentioned that the high-dimensional learning schemes have been studied in several recent works [46, 48]. They pointed out that the prior information learned from high-dimensional tensors is more effective than the information obtained from low-dimensional objects.

The core innovation behind JGM is that it paves a way to directly conduct generative modeling in high-dimensional

3 Proposed JGM model

3.1 Motivation of JGM

Although NCSN has achieved satisfactory results, there are still two major deficiencies. Firstly, it is unable to fully utilize the detailed image information only with prior learning in intensity domain, while learning the gradient distribution of natural images can improve the subjective quality of
In Fig. 2a, CV AE learns a multi-modal conditional model into the automatic colorization. In short, we estimate learning \[ 18, 25, 28 \].

In recent years, likelihood-based models (e.g., variational 

3.2 JGM: prior learning in intensity–gradient domain

In Fig. 2b, iGM expands the freedom of color in multi-color space by means of the image gradient map, and can reduce the representation error by increasing the number of samples. To facilitate the understanding of the underlying contributions in JGM that differ from previous approaches, the visual comparison of CV AE [28], iGM [46] and JGM is shown in Fig. 2.

In the iterative colorization stage, under the linear constraint of data distribution in the joint intensity–gradient domain. At the prior learning stage, a set of high-dimensional tensors are used to learn the prior information from the gradient of target vectors is modeled by mixture density networks. The one-to-many mapping allows the target vectors to take multiple values for the same input vector, which provides diversity. In Fig. 2b, iGM expands the freedom of color in multi-color space and adopts the intermediate variable to obtain linearly autocorrelative constraint that guides the direction of the color generation more precisely. By contrast, the proposed JGM carries out high-dimensional generative modeling via constructing intensity–gradient tensors directly.

The diagram of JGM is depicted in Fig. 2c. As seen, JGM consists of prior learning and iterative colorization stages. At the prior learning stage, a set of high-dimensional tensors are used to learn the prior information from the gradient of data distribution in the joint intensity–gradient domain. At the iterative colorization stage, under the linear constraint of the intensity–gradient domain, we utilize Langevin dynamics to generate more accurate colorizations.

target. Particularly by taking advantage of the joint intensity–gradient field, the proposed JGM learns prior information and then colorizes image in iterative manner.

\[
\nabla_1 x = (\nabla_1 x_r, \nabla_1 x_g, \nabla_1 x_b) \\
\nabla_2 x = (\nabla_2 x_r, \nabla_2 x_g, \nabla_2 x_b)
\]

Accordingly, for every continuously differentiable probability density \( p(X) = [x, \nabla x] = [x, \nabla_1 x, \nabla_2 x] \), we call \( \nabla_X \log p(X) \) as its score function. Through score functions, the score-based generative model progressively denoises an initial white noise into an image with the output of score-based network. A score network \( s_\theta(X) \) is directly trained to estimate the gradient of data prior \( \nabla_X \log p_{\text{data}}(X) \) instead of data prior \( p_{\text{data}}(X) \). The score function is learned via DSM [49, 50]. In this process, the key insight is to perturb the data using multiple noise levels to help the score network capturing both coarse and fine-grained image features. Because learning the score function with the single-noise perturbed data distribution will lead to inaccurate score estimation in the low data density region on high-dimensional data space, which could be severe due to the low-dimensional manifold assumption. The noise distribution is chosen to be \( p_\sigma(\tilde{X}|X) = N(\tilde{X}|X, \sigma^2I) \); therefore \( \nabla_\tilde{X} \log p_\sigma(\tilde{X}|X) = -(\tilde{X} - X)/\sigma^2 \). The score-based generative model aims to train a conditional score network to estimate the scores of all perturbed data distributions jointly, i.e., under a sequence of noise levels \( \{\sigma_i\}_{i=1}^L \). More specifically, for a given \( \sigma \), it estimates the score function of each \( p_\sigma(X) \) by training a single neural network \( s_\theta(X, \sigma) \) with the DSM loss:

\[
\ell(\theta; \sigma) = \frac{1}{2} E_{p_{\text{data}}(X)} \left[ \| s_\theta(X, \sigma) - \nabla_X \log p_\sigma(X) \|^2 \right] \\
= \frac{1}{2} E_{p_{\text{data}}(X)} E_{p_\sigma(\tilde{X}|X)} \left[ s_\theta(\tilde{X}, \sigma) - \nabla_{\tilde{X}} \log p_\sigma(\tilde{X}|X) \right]^2
\]
Then, Eq. (6) is combined for all $\sigma \in \{\sigma_i\}_{i=1}^{L}$ to get one unified objective:

$$L(\theta; \{\sigma_i\}_{i=1}^{L}) \triangleq \frac{1}{L} \sum_{i=1}^{L} \lambda(\sigma_i) \ell(\theta; \sigma_i)$$

(7)

where $\lambda(\sigma_i) > 0$ is a coefficient function depending on $\sigma_i$. As a conical combination of DSM objectives, one can imagine that Eq. (7) achieves the minimum value if and only if $s_{\theta^*}(X, \sigma_i) = \nabla_X \log p_{\sigma_i}(X)$ for all $i \in \{1, 2, \cdots, L\}$. Indeed, perturbing the data with various noise levels and training a single conditional score network by simultaneously using the estimated scores corresponding to all the levels largely improve the performance of score-based generative model.

The essence of stacking to be $X$ is to obtain more data information at high-dimensional manifold and high-density regions, thus avoiding some difficulties in score estimation of DSM [41, 42]. Subsequently, the JGM is trained over data $X$ in high-dimensional space as network input and the parameterized $s_{\theta^*}(X, \sigma)$ is obtained. The visualization of the prior learning stage is depicted in the top region of Fig. 3. Generally speaking, the joint image domain and the gradient domain prior stacked in the channel as the input of the network to assist generative the model in training stage.

### 3.3 JGM: colorization in intensity–gradient field

The entire colorization procedure mainly involves three ingredients in the testing phase: annealed Langevin dynamics, intensity–gradient constraints, and linear minimization. Annealed Langevin dynamics is introduced as a sampling approach. The Langevin dynamics can produce high-dimensional samples from probability density $p(X)$ of a joint intensity–gradient field using only the score function $\nabla_X \log p(X)$. In the iterative colorization process, samples of each disturbance noise level are used as the initial input of next noise level until reaching the smallest. The process provides samples for the network to generate final colorization result gradually. Specifically, given a step size $\alpha > 0$,

![Fig. 3](image-url)
the total number of iterations $T$, and an initial sample $X_t$ from any prior distribution, Langevin dynamics iteratively evaluate the following:

$$X_{t+1} \leftarrow X_t + \alpha \nabla X \log p(X_t) + \sqrt{\alpha} z_t$$

(8)

where $\forall t : z_t \sim N(0, I)$.

Suppose we have a neural network $s_{\theta_i}(X)$ parameterized by $\theta_i$ and it has been trained such that $s_{\theta_i}(X, \sigma_i) \approx \nabla X \log p(X, \sigma_i)$ for all $i \in \{1, 2, \cdots, L\}$. By decreasing the $\sigma_i$-value (accordingly $\sigma_i$-value), we can approximately generate samples from $p(X)$ using annealed Langevin dynamics by replacing $\nabla X \log p_{\theta_i}(X_{t-1})$ with $s_{\theta_i}(X_{t-1}, \sigma_i)$ iteratively, i.e.,

$$X_t \leftarrow X_{t-1} + \frac{\alpha_i}{2} s_{\theta_i}(X_{t-1}, \sigma_i) + \sqrt{\alpha_i} z_{t-1}$$

(9)

Because of the great randomness of generative model, a new constraint called data-consistency flow that imposes on the iteration process is proposed to make the colorization process more controllable and the colorization effect more natural. Thus, the modified annealed Langevin dynamics combined with intensity–gradient data-consistency flow can be expressed as:

$$X_t \leftarrow X_{t-1} + \beta_{\alpha} \nabla X \log p(X) - \lambda DC(\nabla X_{t-1}) + \sqrt{\alpha_i} z_{t-1}$$

(10)

where the data-consistency term $DC(X)$ contains the linear data-fidelity constraints in intensity and gradient domains, respectively, i.e., $DC(X) = DC_1(X) + DC_2(X)$ and

$$DC_1(X) = DC_1(x) = \sum_{i=1}^3 F x_i - y$$

(11)

$$DC_2(X) = DC_2(\nabla X) = \sum_{i=1}^3 F \nabla C x_i - \nabla y$$

The iterative formula of annealed Langevin dynamics in JGM can be rewritten as:

$$X_t \leftarrow X_{t-1} + \frac{\alpha_i}{2} [s_{\theta_i}(X_{t-1}, \sigma_i) - \lambda \nabla X DC(\nabla X_{t-1})] + \sqrt{\alpha_i} z_{t-1}$$

(12)

The visualization of the iterative colorization stage is depicted in the bottom region of Fig. 3.

Finally, we apply the intermediate result to update the image $x_t$ and the gradient maps $[\nabla_1 x_t, \nabla_2 x_t]$. Intuitively, colorized image from joint intensity–gradient field can be attained by direct intensity–gradient integration with solving Poisson equation [51]. However, considering that solving the Poisson equation is an ill-posed problem and may introduce a large error, we sort to a weighting strategy. The approach is a linear minimization with gradient fidelity and image consistency. Obviously, the linear summation incorporates the local information in gradient domain and global information in intensity domain simultaneously. Denoting the final output as $x$, the mixed minimization is formulated as:

$$\min_x \{ ||x - x_t||^2 + \beta \left[ ||\nabla_1 x - \nabla_1 x_t||^2 + ||\nabla_2 x - \nabla_2 x_t||^2 \right] \}$$

(13)

where the first term is for the color confidence to use the latest image to guide the image colorization. The second term is the gradient confidence that is close to the intermediate gradient results. $\beta$ is a parameter balancing the two-loss functions. Contrast to Poisson integration that relies purely on the computed gradients with a boundary condition, our iterative colorization finds a balance between reference image and computed gradients, which is especially useful when gradients are not integrable.

In summary, as explained in Algorithm 1, the whole colorization procedure consists of two-level loops: The outer loop handles $\nabla X \log p_{\theta_i}(X)$ to approximate $\nabla X \log p_{\text{data}}(X)$ of the intensity and gradient domains with decreasing $\sigma_i$-value, while the inner loop decouples to be an alternating process of the updating estimated gradient of data prior $\nabla X \log p_{\theta}(X)$ and the least square scheme.

To intuit the procedure of annealed Langevin dynamics, we provide intermediate samples and the associate PSNR values in Fig. 4. As can be seen, the samples are generated from the high-dimensional noisy data distribution, evolving from pure random noise to colorful images.

---

**Algorithm 1 JGM for Iterative Colorization**

| Training stage |
|----------------|
| **Dataset**: Intensity-gradient domain dataset: $X:=[x, y, z] \in \mathbb{R}_{x,y,z}$ |
| **Output**: Trained JGM $s_{\theta_i}(X, \sigma)$ |

| Iterative colorization stage |
|-----------------------------|
| **Setting**: $\sigma \in \{\sigma_1, \cdots, \sigma, \cdots, \sigma_L\}$, $x_t$, and $x$ |
| For $t \leftarrow 1$ to $T$ do |
| $\alpha_t \leftarrow \alpha_0 \cdot \frac{\sigma_i}{\sigma_L} \cdot \frac{\sigma_t}{\sigma_1}$ |
| For $t \leftarrow 1$ to $T$ do |
| Stack $X_{\omega} = [x_{\omega}, \nabla x_{\omega}]$ and Sample $z_{\omega} \sim N(0, I)$ |
| $X_t \leftarrow X_{\omega} + \frac{\alpha_t}{2} [s_{\theta_i}(X_{\omega}, \sigma_t) - \lambda DC(X_{\omega})] + \sqrt{\alpha_t} z_{\omega}$ |
| End for |
| $x_{\omega} = \arg \min_x \{ ||x - x_{\omega}|| + \beta \left[ ||\nabla_1 x - \nabla_1 x_{\omega}|| + ||\nabla_2 x - \nabla_2 x_{\omega}|| \right] \}$ |
| End for |
| Return $x_{\omega}$ |

---

* Springer
4 Experimental results

In this section, the proposed method is evaluated both quantitatively and qualitatively. First, the performance of JGM is compared to several state-of-the-art automatic colorization methods on different datasets. Then, two ablation studies are performed to evaluate some of the key factors in the model. Finally, many sets of colorization results are exhibited to validate the robustness and diversity ability of JGM. For convenient reproducibility, the source code of JGM is available at https://github.com/yqx7150/JGM.

4.1 Experiment setup

4.1.1 Datasets

LSUN [52] is a large color image dataset. It contains around one million labeled images for each of 10 scene categories and 20 object categories. Among them, we choose the indoor scene LSUN-bedroom dataset and the outdoor scene LSUN-church dataset, which both have enough samples more than 3 million and various colors to verify the effectiveness of JGM. Furthermore, to verify the robustness of JGM, JGM tested on both COCO-stuff [53] and ImageNet [54] datasets.

At the stage of prior learning, we use 150,000 images randomly picked from LSUN dataset. We reshape each image into 128 × 128 pixels as pre-processing. At the test stage, we randomly choose 100 images from each dataset for testing.

4.1.2 Model training

During the learning phase, we use a high-dimensional tensor as the network input and disturb it simultaneously via random Gaussian noise of various amplitudes. Additionally, RefineNet [41] with instance normalization, dilated convolutions and UNet-type architectures [48] was selected as the network structure. Noting that JGM allows flexible model architectures. Adam is chosen as an optimizer with a learning rate of 0.005 and halved every 5,000 iterations. Subsequently, the JGM model is trained for iterations with the batch size of 8 that takes around 20 h. The model is performed with PyTorch interface on 2 NVIDIA Titan XP GPUs, 12 GB RAM.

Quality Metrics: The main metrics used for comparisons among different methods in all reported experiments are the PSNR (Peak-Signal-to-Noise-Ratio), SSIM (Structural Similarity) and Naturalness [55]. Denoting \( x \) and \( \hat{x} \) to be the colorized image and ground-truth, the PSNR is defined as:

\[
\text{PSNR}(x, \hat{x}) = 20 \log_{10} \frac{\text{Max}(\hat{x})}{\|x - \hat{x}\|_2}
\]

(14)

The SSIM is defined as:

\[
\text{SSIM}(x, \hat{x}) = \frac{(2x_\mu \hat{x}_\mu + c_1)(2\sigma_{x\hat{x}} + c_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)}
\]

(15)

Besides, as the ultimate goal of image colorization is to obtain satisfactory results for the observer, we introduce the user study method to test reality and Naturalness of results directly. During the experiment, we invited 100 users, and
The PSNR result is almost 1 dB higher than Su et al. of 26.46 dB and 0.9365 on the LSUN-bedroom dataset. PSNR and SSIM indexes, JGM achieves the highest value which highlights the effectiveness of our approach. For both JGM is slightly lower than Su et al. in terms of PSNR, SSIM, and running time (seconds) performances.

**Table 1** Colorization comparison of JGM to state-of-the-art techniques, in terms of PSNR, SSIM, and running time (seconds) performances.

| Algorithm   | LSUN-bedroom | LSUN-church | Runtime (s) |
|-------------|--------------|-------------|-------------|
| Zhang et al. [24] | 20.89/0.8946 | 23.65/0.9228 | 1.896       |
| MemoPainter | 22.92/0.8975 | 21.66/0.8767 | 0.244       |
| ChromaGAN   | 24.16/0.8899 | 24.63/0.9106 | 0.378       |
| Su et al.   | 25.94/0.9309 | **26.21/0.9433** | 0.123       |
| Deoldify    | 23.93/0.9110 | 24.29/0.9065 | 0.181       |
| iGM-6C      | 24.83/0.9221 | 23.12/0.8953 | 0.055/iter  |
| NCSN        | 21.92/0.8169 | 23.14/0.9252 | 0.031/iter  |
| JGM         | **26.46/0.9365** | 23.76/0.8931 | 0.063/iter  |

The best value of each column is highlighted with bold.

they need to rate on the Saturability, Semantic Correctness and Edge Keeping, respectively. The full score for each indicator is ten (7–10 points is considered good), and statistical the score for the excellent proportion. This criterion will further test whether the algorithm can produce realistic, natural and reasonable coloring effects.

### 4.1.3 Compared methods

JGM is compared with four state-of-the-art methods, including, Zhang et al. [24], MemoPainter [29], ChromaGAN [25], Su et al. [23], Deoldify [62], iGM [46] and NCSN [41].

### 4.2 Comparisons with state-of-the-arts

1. **Quantitative metrics**

   Quantitative comparisons are performed for the LSUN dataset to evaluate the compared algorithms. We randomly use 100 grayscale as input for each method and take the average score of PSNR and SSIM of the 100 final colorized images as the final result.

   We report the quantitative comparisons on LSUN datasets in Table 1. The first block of the results shows models trained on the LSUN datasets. JGM performs favorably against several recent methods [23–25, 29, 46, 62] on LSUN datasets, which highlights the effectiveness of our approach. For both PSNR and SSIM indexes, JGM achieves the highest value of 26.46 dB and 0.9365 on the LSUN-bedroom dataset. The PSNR result is almost 1 dB higher than Su et al. and iGM. As the LSUN-church dataset contains more diverse and challenging scenes, we show the results finetuned on the LSUN-church training set in the second block. Although JGM is slightly lower than Su et al. and Deoldify in values, the coloring effect is commendable. Since colorization is regarded as a one-to-many task that multiple feasible colorized results may be given under the same grayscale input, JGM does not necessarily aim to restore the ground-truth color of the original image to pursue high PSNR and SSIM.

   The coloring effect pursued by JGM is not so similar to the ground-truth, but there is no sense of violation in the eyes of humans. The results present a significant performance boost gained by our method in all metrics, which further highlights the contribution of automatic colorization to the improved performance.

   The colorization results of different approaches are shown in Fig. 5 for the qualitative comparison. For JGM, all of this was done by exploiting joint intensity–gradient and training with a very large set of real images, which allowed our model to generate very realistic colorizations. On the one hand, it can be found that in the LSUN-bedroom dataset, the results of both the others are gloomy and only locally colorized. Similarly, an inharmonious yellow appears in whole image in the third result of Zhang et al. [24]. Compared with the unnatural results of previous methods, JGM provides a more natural result.

   On the other hand, compared with the results of the LSUN-bedroom dataset with unclear color contrast, the results of LSUN-church dataset show that all of the methods are limited to images in which the semantic features work well: sky, sea, buildings, etc.... This limits their application in outdoor and indoor scenes. These methods can achieve suitable colorization, the deficiency of unclear color range and low saturation still exists.

   Figure 6 gives a better sense of the zoomed parts by JGM. To reduce other interference as much as possible, we randomly magnified some areas on results. JGM can separate the small and dense complex patterns from the surroundings.

2. **User study**

   Generally speaking, the purpose of image colorization is to visually present a natural and reasonable effect, rather than recover the ground-truth precisely. We performed a user study asking the question “Does this image look natural to you?” to evaluate the naturalness of the ground-truth validation images, the results of our baseline, and the results of our model. Images are randomly chosen and shown to the users one-by-one. The study was done with 100 different users, each shown roughly 10 images of each type for a total of 1,000 images. Indications were given to the users to use their gut feeling and try to look at the details of the images. More specifically, we show the observers the results and ask them to evaluate the Naturalness of images based on the three indicators, including Saturability, Semantic Correctness and Edge Keeping. Among the three indicators, Saturability is an examination of reasonable colorization of method. In the user study, we ask the observers the results and ask them to evaluate the Naturalness of images based on the three indicators, including Saturability, Semantic Correctness and Edge Keeping. Among the three indicators, Saturability is an examination of reasonable colorization of method.
Fig. 5 Visual comparisons with the state-of-the-arts. From left to right: Grayscale, Ground-truth, Zhang et al. [24], MemoPainter, ChromaGAN, Su et al., Deoldify, iGM, NCSN and JGM. Our method with gradient domain and high-dimensional can predict pleasing colors visually. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 6 The zoomed version of colorization results in JGM. It can be seen that the colorized images by JGM exhibit showing high naturalness and contrast and Edge Keeping can describe the degree of color overflow and edges blur.

During the experiment, we invited 100 users, and they need to score 1–10 points on the three indicators in each pair. The scoring is based on the following three points: (i) whether the color of the image is dim; (ii) whether the image has unreasonable colors; (iii) whether the image will overflow at the edge. Finally, it is converted to a percentage based on the number of people and the score. After these indicators are evaluated, the Naturalness is the equally weighted sum of the three values. The results are recorded in Table 2. To facilitate observation, we draw the error bar figures for comparing the images generated by different approaches as in Fig. 7.

The Naturalness of Zhang et al. [24], MemoPainter, ChromaGAN and iGM are 88.84%, 89.94%, 90.44% and 88.96%, respectively. This phenomenon indicates that all the methods can achieve the goal of global Naturalness but being less sensitive to local colorization.

In addition, other colorization methods face great challenges in terms of color saturation and semantic correctness, all of which have achieved low values. As can be seen from the results that the iGM effects are dim, Zhang et al. [24] and ChromaGAN appear desaturated yellow, MemoPainter cannot achieve spatially consistent colorization.

It is commendable that JGM reaches the best results of 93.97%, 92.89% and 92.66% on the three indicators, indicating high saturation, contrast and semantic accuracy,
Table 2  Comparison of naturalness between JGM and other state-of-the-arts. The mean value of each criterion is recorded.

| Algorithms   | Saturability (%) | Semantic correctness (%) | Edge keeping (%) | Naturalness (%) |
|--------------|------------------|--------------------------|-----------------|----------------|
| Zhang et al. | 88.48            | 89.02                    | 89.02           | 88.84          |
| MemoPainter  | 90.02            | 90.12                    | 89.68           | 89.94          |
| ChromaGAN   | 90.30            | 90.95                    | 90.06           | 90.44          |
| iGM-6C       | 87.38            | 88.47                    | 90.04           | 88.96          |
| JGM          | **93.97**        | **92.89**                | **92.66**       | **93.17**      |

The best value of each column is highlighted with bold.

Fig. 7 The exemplar of the Naturalness of the state-of-the-art automatic colorization methods includes three aspects: Saturability, Semantic Correctness and Edge Keeping. JGM can produce more natural results. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

without edge overflow. It can achieve distinct colors on different objects. The model with gradient information helps to achieve the highest performance and produce convincing colorizations while retaining global color consistency. In summary, the results obtained by JGM have a better perception of the object color and generate the highest Naturalness effect, which can be considered as the real color by subjects.

4.3 Ablation study

There are two main components critical to our final performance: prior learning in joint or separate mode, and iteration with or without gradient data-fidelity. Here we conduct two ablation studies to validate these important designs.

(1) Comparison of joint/separate learning

In the algorithm of the JGM-divide, two networks $s_{θ_1}(x)$ and $s_{θ_2}(∇x)$ are trained in the image domain and gradient domain separately. Then, using the annealed Langevin dynamics with the prior information, image domain and gradient domain samples can be obtained from these trained networks.

Since the use of two separate models, both the image domain $x$ and the gradient domain $∇x$ need to be taken into consideration during the iterative colorization process. Thus, the modified annealed Langevin dynamics combined with DC flow can be rewritten as follows:

$$x_{t+1} \leftarrow x_t + \frac{α_t}{2} [s_{θ_1}(x_t) - λ_1 DC(x_t)] + √α_t z_1$$

$$∇x_{t+1} \leftarrow ∇x_t + \frac{α_t}{2} [s_{θ_2}(∇x_t) - λ_2 DC(∇x_t)] + √α_t z_2$$

(16)

where the $DC(x) = (Fx - y)$ is the data-consistency of image domain, and $DC(∇x) = (F∇x - ∇y)$ is the data-consistency of gradient domain.

Compared with the standard JGM, although JGM-divide can alleviate the issues of uneven colorization and insufficient contrast to some extent, it needs to train two prior information of gradient domain and image domain separately. Thus, in turn, it increases the training complexity and extends the computational cost for colorization. More importantly, considering the lack of data samples in areas with low data density, DSM may not have enough evidence to accurately estimate the scoring function and the complexity of multiple models. This observation prompts us to improve performance by sampling in a high-dimensional embedding space. It can be seen from Table 3 that sampling in the high-dimensional embedding space can produce better results, and the contrast is higher. In brief, the visual effect and rationality of the standard JGM are better than JGM-divide.
Table 3  Colorization comparison of separate and joint learning under gradient constraints, in terms of PSNR and SSIM performance

| Algorithms    | JGM-divide | JGM       |
|---------------|------------|-----------|
| Church        | 23.26/0.9269 | 23.76/0.8931 |
| Bedroom       | 22.23/0.9059 | 25.46/0.9365 |
| COCO-stuff    | 19.45/0.8477 | 20.72/0.7254 |
| ImageNet      | 15.13/0.8227 | 18.71/0.6821 |

The best value of each row is highlighted with bold

4.4 Robustness test

4.4.1 Colorizing legacy black-and-white photos

Colorization is an ill-posed problem due to the large degree of freedom during assigning color information. Mathematically, the common approach in the literature to solve the problem relies on the following linear observation model:

\[ y = Fx + n \]  

(17)

where \( y \in \mathbb{R}^N \) and \( x \in \mathbb{R}^N \) denote the gray-level image and the original color image, respectively. \( n \in \mathbb{R}^N \) denotes the additive white Gaussian noise. \( F \in \mathbb{R}^{N \times N} \) denotes the degradation matrix for graying \( x \). Regarding the forward operator, some researchers assume the lightness channel \( L \) in CIE Lab color space [24] or the luminance channel \( Y \) in YUV color space as the grayscale observation [18]. In more general cases, we can only observe the grayscale image \( y \) without knowing its forward model \( F \). In this circumstance, the task of “blind” colorization is more challenging.

In this experiment, a prevailing processing method of forming \( F \) is chosen:

\[ F(x) = (x_R + x_G + x_B)/3.0 \]  

(18)

As we observed in Fig. 9, convincing results are generated by JGM. Due to the lack of similar images in the training dataset, the problem of color overflow still exists. Take the fourth image for example, the result is realistic, but the ground contains obvious artifacts.

Fig. 8  Colorization visualization results of the JGM with different variants. a Grayscale, b JGM-divide, c JGM without data-fidelity constraint in gradient domain and d JGM. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.5 Diversity of colorization results

Diverse colorization aims to generate different colorized images rather than restore the original color, which is often achieved via GANs [43] or VAE [44]. Due to the joint high-dimensional intensity and gradient of image information, JGM can improve the original performance of the generative model, generating multiple feasible colorizations.

As shown in Fig. 10, the diverse colorizations have reasonable effects for indoor and outdoor scenes. We observe different tones and background colors for indoor scenes and different rivers, sky for outdoor scenes. In order to reflect the diversity of the results, we relaxed the intensity–gradient constraints of the single colorization. Although JGM can produce diversified results, we have observed that some of the results have the problem of color overflow. Compared with Fig. 5, it not only reflects the diversity of JGM, but also proves the effectiveness of the constraints.
5 Conclusion

In this work, an iterative generative model for automatic colorization with two major characteristics was proposed, namely JGM, prior learning in joint intensity–gradient domain and iterative colorization under data-consistency constraints. More specifically, generative modeling in high-dimensional space has addressed major limitations of existing methods and provided diverse probable colorizations. Additionally, enforcing data-fidelity in joint intensity–gradient field produced substantially improved the generative procedure and the final colorization results, as validated by extensive comparisons with state-of-the-art methods. The proposed JGM only involved a couple of insensitive parameters, which were fixed in all experiments. In the forthcoming study, we hope to apply these critical factors in wavelet domain, etc. Besides, extending the core idea used in colorization to hyperspectral imaging is also a promising direction [57].

Acknowledgements The authors sincerely thank the anonymous reviewers for their valuable comments and constructive suggestions that are very helpful in the improvement of this paper. This work was supported by National Natural Science Foundation of China (61871206, 61601450).

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Baig, M.H., Torresani, L.: Multiple hypothesis colorization and its application to image compression. Comput. Vis. Image Understand. pp. 111–123 (2017).
2. Liu, Q., Li, S., Xiong, J., Qin, B.: WpmDecolor: weighted projection maximum solver for contrast-preserving decolorization. Vis. Comput. 35(2), 205–221 (2019)
3. Zhang, X., Liu, S.: Contrast preserving image decolorization combining global features and local semantic features. Vis. Comput. 34, 1099–1108 (2018)
4. Wu, J., Shen, X., Liu, L.: Interactive two-scale color-to-gray. Vis. Comput. 28, 723–731 (2012)
5. Frans, K.: Outline colorization through tandem adversarial networks. arXiv:1704.08834 (2017).
6. Qu, Y., Wong, T.T., Heng, P.A.: Manga colorization. ACM Trans. Graph. 25(3), 1214–1220 (2006)
7. Limmer, M., Lensch, H.P.: Infrared colorization using deep convolutional neural networks. In: Proceedings of IEEE International Conference on Machine Learning and Application (ICMLA), pp. 61–68 (2016).
8. Guo, J., Pan, Z., Lei, B., Ding, C.: Automatic color correction for multisource remote sensing images with Wasserstein CNN. Remote Sensing 9(5), 483 (2017)
9. Lei, C., Chen, Q.: Fully automatic video colorization with self-regularization and diversity. In: Proceedings of IEEE Conference Computer Vision Pattern Recognition, pp. 3753–3761 (2019).
10. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of IEEE Conference Computer Vision and Pattern Recognition, pp. 1125–1134 (2017).
11. Levin, A., Lischinski, D., Weiss, Y.: Colorization using optimization. In: ACM SIGGRAPH 2004 Papers, Vol. 23, no. 3, pp. 689–694, (2004).
12. Walsh, T., Ashikhmin, M., Mueller, K.: Transferring color to greyscale images. ACM Trans. Graph. 21(3), 277–280 (2002)
13. Manjunatha, V., Iyyer, M., Boyd-Graber, J., Davis, L.: Learning to color from language. In: North American Chapter of the Association for Computational Linguistics (NAACL), pp. 764–769 (2018).
14. Bahng, H., Yoo, S., Cho, W., Keetae Park, D., Wu, Z., Ma, X., Choo, J.: Coloring with words: Guiding image colorization through text-based palette generation. In: Proceedings of Euro Conference Computer Vision, pp. 431–447 (2018).
15. Lee, J., Kim, E., Lee, Y., Kim, D., Chang, J., Choo, J.: Reference-based sketch image colorization using augmented-self reference and dense semantic correspondence. In: Proceedings of IEEE Conference Computer Vision and Pattern Recognition, pp. 5801–5810 (2020).
16. Li, B., Lai, Y.K., John, M., Rosin, P.L.: Automatic example-based image colorization using location-aware cross-scale matching. IEEE Trans. Image Process. 28(9), 4606–4619 (2019)
17. Xu, Z., Wang, T., Fang, F., Sheng, Y., Zhang, G.: Stylization-based architecture for fast deep exemplar colorization. In: Proceedings of IEEE Conference Computer Vision Pattern Recognition pp. 9363–9372 (2020).
18. Cao, Y., Zhou, Z., Zhang, W., Yu, Y.: Unsupervised diverse colorization via generative adversarial networks. In: Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery Databases. Springer, Cham, pp. 151–166 (2017).
19. Charpiat, G., Hofmann, M., Schölkopf, B.: Automatic image colorization via multimodal predictions. In: Proceedings of European Computer Vision Conference. Springer, Berlin, pp. 126–139 (2008).
20. Cheng, Z., Yang, Q., Sheng, B.: Deep colorization. In: IEEE International Conference on Computer Vision, pp. 415–423 (2015)
21. Guadarrama, S., Dahl, R., Bieber, D., Norouzi, M., Shlens, J., Murphy, K.: Pixcolor: Pixel recursive colorization. In: British Machine Vision Conference (BMVC) (2017).
22. Morimotoand, Y., Taguchii, Y., Naemura, T.: Automatic colorization of greyscale images using multiple images on the web. In: Proceedings of SIGGRAPH, pp. 59–59 (2009).
23. Su, J.W., Chu, H.K., Huang, J.B. (2020) Instance-aware image colorization. In: Proceedings of IEEE Conference Computer Vision Pattern Recognition, pp. 7968–7977 (2020).
24. Zhang, R., Isola, P., Efros, A.A.: Colorful image colorization. In: Proceedings of European Conference Computer Vision, pp. 649–666 (2016).
25. Vitoria, P., Raad, L., Ballester, C.: ChromaGAN: adversarial image colorization with semantic class distribution. In: IEEE Winter Conference on Applications of Computer Vision, pp. 2445–2454 (2020).
26. Messaoud, S., Forsyth, D., Schwing, A.G.: Structural consistency and controllability for diverse colorization. In: Proceedings of European Conference Computer Vision, pp. 596–612 (2018).
27. Larsson, G., Maire, M., Shakhnarovich, G.: Learning representations for automatic colorization. In: Proceedings of European Conference Computer Vision, pp. 577–593 (2016).
28. Deshpande, A., Lu, J., Yeh, M.C., Chong, M.J., Forsyth, D.: Learning diverse image colorization. In: Proceedings of IEEE Conference Computer Vision and Pattern Recognition, pp. 6837–6845 (2017).
29. Yoo, S., Bahng, H., Chung, S., Lee, J., Chang, J., Choo, J.: Coloring with limited data: Few-shot colorization via memory augmented networks. In: Proceedings of IEEE Conference Computer Vision Pattern Recognition, pp. 11283–11292 (2019).
30. Anwar, S., Tahir, M., Li, C., Mian, A., Khan, F.S., Muzaffar, A.W.: Image colorization: a survey and dataset. arXiv preprint arXiv:2008.10774 (2020).
31. Yatziv, L., Sapiro, G.: Fast image and video colorization using chrominance blending. IEEE Trans. Image Process. 15(5), 1120–1129 (2006)
32. Sheikh, H.R., Bovik, A.C.: Information theoretic approaches to image quality assessment. In: Handbook of Image and Video Processing, Academic Press, pp. 975–989 (2005).
33. Gong, Y., Shalzarinii, I.F.: Image enhancement by gradient distribution specification. In: Proceedings of Asian Conference Computer Vision, pp. 47–62 (2014).
34. Sun, Z., Feng, W., Zhao, Q., Huang, L.: Brightness preserving image enhancement based on a gradient and intensity histogram. J. Electron. Imag. 24(5), 053006–053006 (2015)
35. Mi, Z., Liang, Z., Wang, Y., Fu, X., Chen, Z.: Multi-scale gradient domain underwater image enhancement. In: Proceedings of OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO), pp. 1–5 (2018).
36. Wang, H., Chen, Y., Fang, T., Tyan, J., Attuia, N.: Gradient adaptive image restoration and enhancement. In: Proceedings of IEEE International Conference on Image Processing, pp. 2893–2896 (2006).
37. Cho, T.S., Züitick, C.L., Joshi, N., Kang, S.B., Szeliski, R., Freeman, W.T.: Image restoration by matching gradient distributions. IEEE. Trans. Pattern Anal. Mach. Intell. 34(4), 683–694 (2012).
38. Petrovic, V.S., Xydeas, C.S.: Gradient-based multisemiesolution image fusion. IEEE. Trans. Image Process. 13(2), 228–237 (2004)
39. Pan, J., Hu, Z., Su, Z., Yang, M.H.: L0-regularized intensity and gradient prior for deblurring text images and beyond. IEEE. Trans. Pattern Anal. Mach. Intell. 39(2), 342–355 (2016)
40. Gooch, A.A., Olsen, S.C., Tumblin, J., Gooch, B.: Color2gray: Salience-preserving color removal. ACM Trans. Graph. 24(3), 634–639 (2005)
41. Song, Y., Ermon, S.: Generative modeling by estimating gradients of the data distribution. In: Proceedings of Advances on Neural Information Processing System, pp. 11918–11930 (2019).
42. Vincent, P.: A connection between score matching and denoising autoencoders. Neural Comput. 23(7), 1661–1674 (2011)
43. Goodfellow, I., Abadie, J.P., Mirza, M., Xu, B., Farley, D.W., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In Proceedings of Advances in Neural Information Processing System, pp. 2672–2680 (2014).
44. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: International Conference on Learning Representations (ICLR), (2014)
45. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: International Conference on Learning Representations (2014).
46. Zhou, J., Hong, K., Deng, T., Wang, Y., Liu, Q.: Progressive colorization via iterative generative models. IEEE Signal Process. Letter 27, 2054–2058 (2020)
47. Bishop, C.M.: Mixture density networks. Aston University, Birmingham, U.K., Tech. Rep. NCRG/94/004 (1994).
48. Liu, Q., Yang, Q., Cheng, H., Wang, S., Zhang, M., Liang, D.: Highly undersampled magnetic resonance imaging reconstruction using autoencoding priors. Magn. Reson. Med. 83(1), 322–336 (2020)
49. Zhang, L., Wu, X., Buades, A., Li, X.: Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. J. Electron. Imag. 20(2), 023016 (2011)
50. Jayaram, V., Thickstun, J.: Source separation with deep generative priors. arXiv preprint arXiv:2002.07942 (2020).
51. Pérez, P., Gangnet, M., Blake, A.: Poisson image editing. In: ACM SIGGRAPH, pp. 313–318 (2003).
52. Yu, F., Seff, A., Zhang, Y., Song, S., Funkhouser, T., Xiao, J.: LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop. arXiv preprint arXiv:1506.03365 (2015).
53. Caesar, H., Uijlings, J., Ferrari, V.: Coco-stuff: Thing and stuff classes in context. In: Proceedings of IEEE Conference Computer Vision Pattern Recognition, pp. 1209–1218 (2018).
54. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, A., Karpathy, Z., Khosla, A., Bernstein, M., Berg, A.C., Li, F.: ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 115(3), 211–252 (2015)
55. Zhao, J., Liu, L., Snoek, C.G.M., Han, J., Shao, L.: Pixel-level semantics guided image colorization, arXiv preprint arXiv:1808.01597 (2018).
56. 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 Trans. Graph. 35(4), 1–11 (2016)
57. Liu, Q., Leung, H.: Variable augmented neural network for decolorization and multi-exposure fusion. Inf. Fusion 46, 114–127 (2019)
58. Iony, R., Cohen-Or, D., Lischinski, D.: Colorization by example. In: Proceedings of Euro graphics Symposium Rendering, vol. 2, pp. 201–210 (2005)
59. Block, A., Mroueh, Y., Rakhlin, A.: Generative modeling with denoising auto-encoders and Langevin sampling. arXiv preprint arXiv:2002.00107 (2020).
60. Narayanan, H., Mitter, S.: Sample complexity of testing the manifold hypothesis. In: Proceedings of Advance Neural Information Processing System (2010).
61. Rifai, S., Dauphin, Y., Vincent, P., Bengio, Y., Muller, X.: The manifold tangent classifier. In: Proceedings of Advance Neural Information Processing System (2011).
62. Antic, J.: A deep learning based project for colorizing and restoring old images (and video!). https://github.com/jantic/DeOldify, 2019. Online; Accessed 16 Oct 2019.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.
Weidong Liao received his Ph.D. degree in Industrial Economics from Jiangxi University of Finance and Economics in 2003. He was a visiting professor at University of Southampton, UK, from 2005 to 2006. Ever since 2007, he has been with School of Economics, Jiangxi University of Finance and Economics, Nanchang, China, where he is currently a professor. His current research interest includes artificial intelligence, industrial economics, etc.

Qiegen Liu received the Ph.D. degree in Biomedical Engineering from Shanghai Jiao Tong University. Since 2012, he has been with School of Information Engineering, Nanchang University, Nanchang, China, where he is currently a professor. During 2015-2017, he was also a postdoc in UIUC and University of Calgary. His current research interest is sparse representation, deep learning and their applications in image processing, computer vision and inverse imaging.