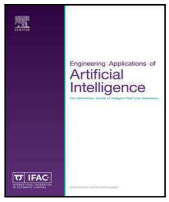
Engineering Applications of Artificial Intelligence 114 (2022) 105006

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Survey paper

# Deep learning for image colorization: Current and future prospects

Shanshan Huang [a], Xin Jin [b,c,*], Qian Jiang [b,c], Li Liu [a,**]

[a] School of Big Data & software Engineering, Chongqing University, Chongqing, 400000, China
[b] Engineering Research Center of Cyberspace, Yunnan University, Kunming, 650000, China
[c] School of Software, Yunnan University, Kunming 650000, China

ARTICLE INFO

*Keywords:*
Image colorization
Deep learning
Convolutional neural network
Generative adversarial network
Transformer

ABSTRACT

Image colorization, as an essential problem in computer vision (CV), has attracted an increasing amount of researchers attention in recent years, especially deep learning-based image colorization techniques(DLIC). Generally, most recent image colorization methods can be regarded as knowledge-based systems because they are usually trained by big datasets. Unlike the existing reviews, this paper adopts a unique deep learning-based perspective to review the latest progress in image colorization techniques systematically and comprehensively. In this paper, a comprehensive review of recent DLIC approaches from algorithm classification to existing challenges is provided to facilitate researchers' in-depth understanding of DLIC. In particular, we review DLIC algorithms from various perspectives, including color space, network structure, loss function, level of automation, and application fields. Furthermore, other important issues are discussed, such as publicly available benchmark datasets and performance evaluation metrics. Finally, we discuss several open issues of image colorization and outline future research directions. This survey can serve as a reference for researchers in image colorization and related fields.

## 1. Introduction

*The art challenges the technology, and the technology inspires the art. — John Lasseter*

Image colorization is the process of assigning RGB color value to each pixel of a grayscale image to obtain colorized images, which is a prospective image processing technique in computer vision (CV). Colorized images have a better visual experience and are widely used in image recognition (Cordonnier et al., 2021; Chen et al., 2022; Wu et al., 2021a), object detection (Tang et al., 2022; Dai et al., 2021), and other fields (Deng et al., 2021; Jin et al., 2021c; Valanarasu et al., 2021). Therefore, image colorization methods have been extensively studied, including but not limited to animation scene design (Ci et al., 2018; Zou et al., 2019; Zhang* et al., 2018; Yoo et al., 2019; Ramassamy et al., 2019), historical photograph restoration (Larsson et al., 2016; He et al., 2018; Zhang* et al., 2017; Su et al., 2020; Zhang et al., 2016), infrared image colorization (Kuang et al., 2020; Suarez and Sappa, 2017; Suarez et al., 2018; Xu et al., 2021; Dong et al., 2018; Zhong et al., 2020), remote sensing image processing (Ji et al., 2020; Gravey et al., 2019; Dias et al., 2020; Song et al., 2017), and so on (Dong et al., 2022; Xuan et al., 2021; Bian et al., 2021; Liang et al., 2021; Morra et al., 2021; Yu et al., 2020; Guo et al., 2021; Mathur et al., 2021). Image colorization is a multimodal problem, that is, the same target object has different colorization schemes. For example, a pair of shoes can be white, red, yellow, or other colors. In general, image colorization is a challenging and interesting research problem.

In recent years, the powerful feature extraction ability of deep learning in image processing has shown great application potential with rapid development. Particularly, the first deep learning-based image colorization (DLIC) method was proposed in 2015 (Cheng et al., 2015), DLIC algorithms have rapidly shown superior performance over conventional solutions and are constantly improving to the state of the art. Various deep learning techniques have been applied to image colorization tasks, including conventional convolutional neural networks (CNNs) (Larsson et al., 2016; He et al., 2018; Zhang* et al., 2017; Su et al., 2020; Zhang et al., 2016; Dias et al., 2020; Dong et al., 2022; Xuan et al., 2021; Iizuka et al., 2016; Dabas et al., 2020; Chybicki et al., 2019; Zhang et al., 2021a; Khanolkar et al., 2021; Endo et al., 2021; Xiao et al., 2019b; An et al., 2020; Varga and Sziranyi, 2016; Larsson et al., 2017a; M.H. Baig, 2017b), generative adversarial networks (GANs) (Zou et al., 2019; Zhang* et al., 2018; Kuang et al., 2020; Chen and Hays, 2018; Zhang et al., 2019; Hensman and Aizawa, 2017; Seo and Seo, 2021; Cao et al., 2017), capsule neural networks (CapsNet) (Zbulak, 2020), Transformer (Manoj et al., 2021), and so on (Su et al., 2018; Zhao et al., 2020; Liang et al., 2016). Fig. 1 shows the general development trajectory of image colorization methods based on deep learning since 2015.

* Corresponding author at: School of Software, Yunnan University, Kunming 650000, China.
** Corresponding author.
*E-mail addresses:* xinxin_jin@163.com (X. Jin), dcsliuli@cqu.edu.cn (L. Liu).

https://doi.org/10.1016/j.engappai.2022.105006
Received 13 March 2022; Received in revised form 16 May 2022; Accepted 25 May 2022
Available online xxxx
0952-1976/

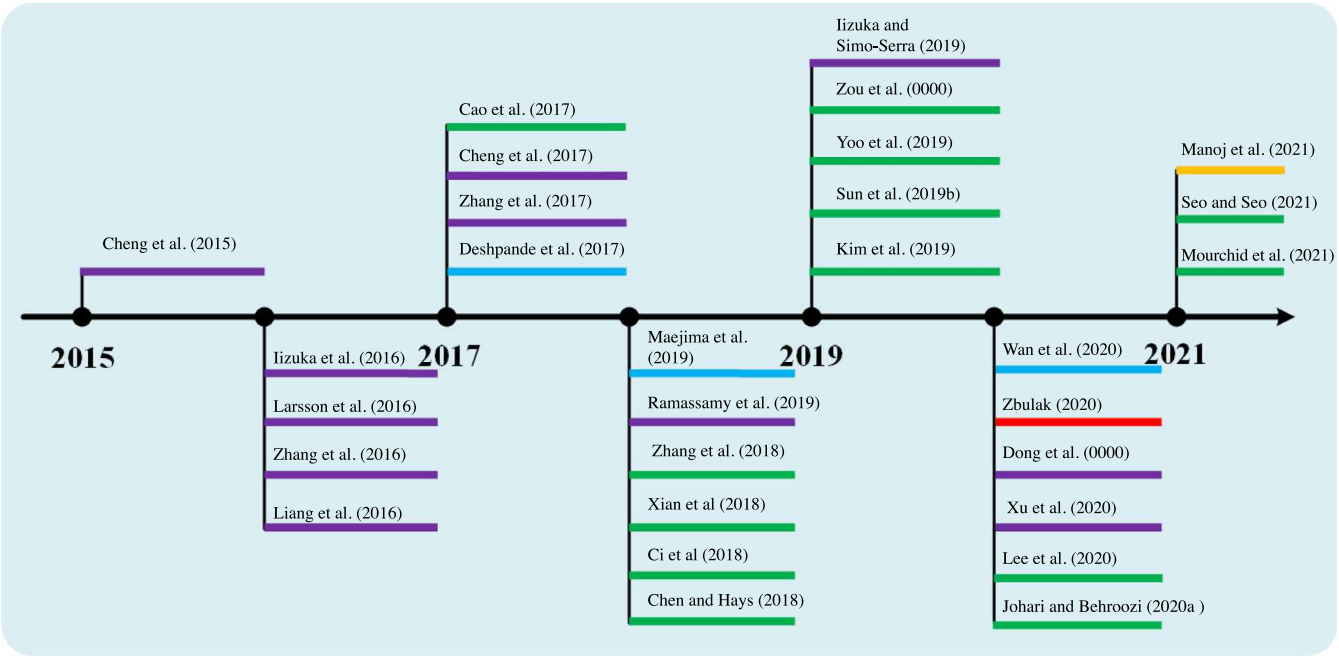



**Fig. 1.** Development of DLIC. Purple: CNN-based method. Green: GANs-based method. Red: CapsNet-based image colorization method. Yellow: Transformer-Based Method.

Due to the increasing requirements of image controllability and automation, the DLCI methods also can be divided into fully automatic colorization methods and semi-automatic colorization methods. For the former, the color image is obtained through an end-to-end model, and there is no need for human intervention and preprocessing or post-processing. However, this method often requires a large number of image datasets to conduct model training, and the color of obtained image is often relatively simple and uncontrollable. The latter can be guided by user guidance, including scribbles, text description, reference images, etc., to make up for the color uncontrollable problem of fully automatic colorization methods. However, because the semi-automatic colorization methods are heavily rely on human guidance, it is not suitable for the novice. In addition, semi-automatic methods are often difficult for users to provide accurate color information to guide natural scenarios image colorization, so the color of the colorized image look unnatural. Therefore, how to implement the fully automatic and controllable image colorization still requires further exploration.

The research on colorization in theory and application has been rapidly developing in recent years. However, several certain problems require attention. Existing DLIC methods often require large-scale image datasets, and the obtained colorization results are often unsatisfactory, such as color uneven or unsaturation, and lack of color diversity. Although the control of colorized results can be achieved through human interaction, it largely depends on people's aesthetic and experience, so it is not suitable for novice. In addition, there are some problems, such as artifact and loss of detail information. Therefore, image colorization is an interesting and challenging research that deserves further exploration.

In this paper, we give a comprehensive review of recent advanced DLIC methods. In the existing review literature, there are few review articles on DLIC methods. Most of the early review articles focus on conventional non-deep learning image colorization methods, and the existing reviews on DLIC methods are often not comprehensive enough. Such as, Recently, Žeger et al. (2021) summarized the methods of image colorization based on deep learning, and described several commonly used objective image quality evaluation metrics. However, this review only focuses on the colorization of natural images, but ignores the colorization methods in other fields (Line art images, infrared images, remote sensing images, etc.). Moreover, this review is not comprehensive enough, and some open issues existing in the colorization methods are not discussed. More recently, Anwar et al. (2022) reviewed image colorization methods from the perspective of domain type, network structure, etc., but this survey did not review the image colorization methods in the last two years, and only reviewed the single-image colorization methods. Therefore, this paper adopts a unique deep learning-based perspective to systematically and comprehensively review the latest progress of colorization techniques. The main contributions of this paper include the following three aspects:

(1) We comprehensively review the existing advanced DLIC methods from various perspectives. The existing DLIC methods are classified and sorted out from four aspects: color space, loss function, network structure, and application field. The proposed taxonomy is intended to help researchers gain a deeper understanding of the key Characteristics of DLIC models.
(2) We summarize the key issues of DLIC methods, including problem definition, datasets, image quality assessment. In addition, we have conducted a series of comparative experiments and comprehensively evaluated the performance of different colorization models by various objective evaluation metrics.
(3) We discuss existing challenges and some open-minded issues, and identify the development trend and future research direction of DLIC methods, to provide insightful guidance for further research.

The rest of this paper is organized as follows. Section 2 explains the problem definition of the image colorization task, and classifies DLIC methods according to different color spaces and loss functions, respectively. Section 3 summarizes the existing method from the perspective of network structure. Section 4 provides an overview of the DLIC methods according to the level of automation. Section 5 overviews the different application fields of image colorization. Sections 6 and 7 summarizes the existing public datasets and image quality evaluation criteria for image colorization respectively, and presents the experimental results several representative colorization methods. Section 8 discusses the challenges and future research directions of colorization. Finally, Section 9 summarizes the work of this article. Fig. 2 shows the structure of this survey.

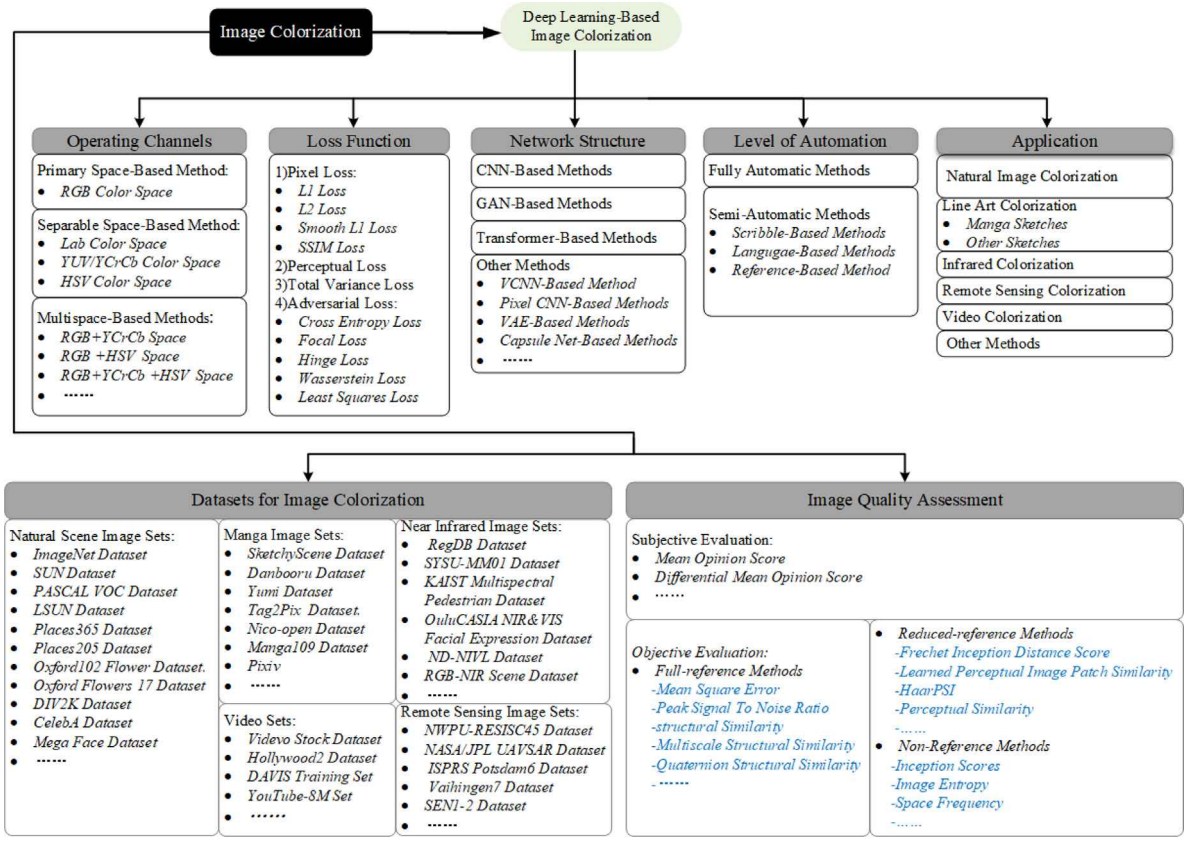



**Fig. 2.** Structure of this survey.

## 2. Problem setting and terminology

This section introduces the problem definition and terminology of image colorization, and summarizes the existing DLIC methods from the perspective of color space and loss function. Table 1 shows some representative DLIC methods.

### 2.1. Problem definitions

Before reviewing recent DLIC models, we first provide a common definition of image colorization. Image colorization refers to the restoration of corresponding color images from grayscale or line art images. In practice, it is difficult to obtain a large number of gray image datasets to train a colorization model, so the gray image $I_g$ is usually modeled as the output of the following equation:

$$I_g = \Phi\left(I_r\right) \tag{1}$$

where $I_r$ represents the color image. For the line art image, the conventional edge detection algorithm, such as eXtended Difference-of-Gaussians (XDoG) (Ci et al., 2018; Zou et al., 2019; Lee and Lee, 2020; Kim et al., 2019; Liu et al., 2017), Canny Edge Detection algorithm (Seo and Seo, 2021; Thasarathan and Ebrahimi, 2019; Sun et al., 2019b; Li et al., 2021b), is used to process the true color image. Researchers need to colorize the obtained grayscale image or line art image, i.e. given an input grayscale image or line art image $I_g$ with a size of $W \times H$, the input gray image $I_g$ is mapped into a color image $I_c$ through the image colorization model $f$. The equation is as follows:

$$I_c = f\left(I_g\right) \tag{2}$$

For DLIC methods, the model $F$ is usually obtained by learning a collection of training samples. i.e., given a grayscale image collection $G = \left\{I_g \in \mathbb{R}^{W\times H\times 1}\right\}$ and corresponding real color image collection $C = \left\{I_c \in \mathbb{R}^{W\times H\times 3}\right\}$, find a model $F$ which can minimize prediction errors $L$.

$$\hat{\theta} = \arg\min_{\theta} L\left(I_c - I_r\right) + \lambda\Psi\left(\theta\right) \tag{3}$$

$L$ here is usually certain distance measurement (such as L1 distance, L2 distance) or a combination of various distance measurements. $F$ is the set of potential mapping functions. $\Psi\left(\theta\right)$ is the regularization term, and $\lambda$ is the compromise parameter. DLIC methods are typically modeled through deep learning networks, which are discussed later in Section 3.

### 2.2. Color space for image colorization

In the past few years, color space plays an important role in DLIC (Larsson et al., 2016; Zhang* et al., 2017; Su et al., 2020; Zhang et al., 2016; Cheng et al., 2015; Iizuka et al., 2016; Manoj et al., 2021; Deshpande et al., 2017; Xu et al., 2020). Color space is the theoretical basis of color information research, which quantifies color from people's subjective feelings to concrete expression, and provides a powerful basis for the computer record and performance of color. The selection of different color spaces has a great influence on whether the image colorization methods are effective. Therefore, choosing a suitable color space is an important issue in ensuring the performance of the colorization models.

Color space can be expressed in various forms, which can be divided into two categories according to the basic structure, i.e., the primary color space and color-bright separable color space. The former is typically RGB color space, while the latter includes YUV/YCrCb, CIELAB, HSV, and other spaces. Correspondingly, according to the different color spaces used, the image colorization method can be roughly divided into RGB space-based methods, separable space-based methods, and multispace-based methods. The main difference between the two methods is that the former uses single-channel grayscale image $X \in \mathbb{R}^{H\times W\times 1}$ as the input to predict the three-channel color image

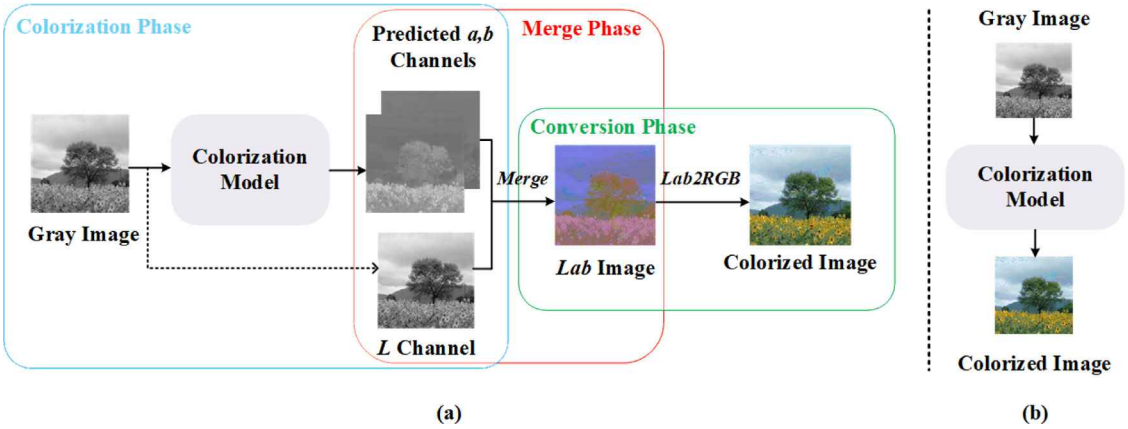



**Fig. 3.** Processing steps of color space-based colorization method. (a) Separable space-based method, here take CIELAB Color space as an example. (b) RGB space-based method.

$Y \in \mathbb{R}^{H\times W\times 3}$, while the latter predicts two missing color channels $Y \in \mathbb{R}^{H\times W\times 2}$.

Next, we will focus on commonly used color spaces. On this basis, the image colorization methods based on deep learning are classified and reviewed.

### 2.2.1. RGB space-based method

RGB color space is the most well-known color space, which is widely used in various fields of image processing. This color space consists of three channels: red (R), green (G), and blue (B). Each color channel has 256 gray values (0–255), and each channel can be combined in a specific proportion to present different colors. However, there are some problems with the RGB color space. For example, the color changes with the value of each channel; the gray value of each channel of a certain color is difficult to express accurately.

RGB color space is the most widely used in the line art image colorization task (Ci et al., 2018; Zhang* et al., 2018; Seo and Seo, 2021; Kim et al., 2019; Thasarathan and Ebrahimi, 2019; Zhang et al., 2017; Xie* et al., 2020), which is mainly because, compared with gray image, line art image only has simple line composition, neither gray value nor semantic information, so it is difficult to realize the colorization in the color-light separable color space. For example, Zhang* et al. (2018) proposed a semi-automated colorization model based on RGB color space, which solves artifacts such as watercolor blur color distortion and dark texture to some extent. Similarly, a GANs-based line art colorization method was proposed by Seo and Seo (2021). This method achieves good colorization performance in RGB color space. In addition, RGB color space is also used in gray image colorization tasks, and has achieved a favorable colorization effect (Ramassamy et al., 2019; Johari and Behroozi, 2020a,b; Mourchid et al., 2021).

However, there are some limitations to the image colorization method based on RGB space, that is, it is necessary to predict R, G, B channels with a given grayscale image, which increases the difficulty of the colorization task. For the representation of color predictions, using RGB is overdetermined, as lightness is already known.

### 2.2.2. Separable space-based method

The Human Visual System (HVS) is less sensitive to color than to brightness. In the RGB color space, three primary colors are equally important, but the brightness information is ignored. In color-light separable color space, the chroma information and the brightness information of the image can be separated, so that we can handle the chroma and brightness information, separately. This kind of color space is closer to human vision and more convenient for color editing. In general, the separable space-based method consists of three stages, as shown in Fig. 3, and CIELAB color space here is taken as an example. Firstly, the brightness channel L (grayscale image) is input into the colorization model to obtain the two missing channels A and B. Then, the complete CIELAB color image is obtained by combining the obtained chrominance AB with the input brightness channel L. Finally, the merged image is converted to RGB color space through color space conversion to obtain the final color image.

Compared with RGB space-based methods, the separable space-based methods only need to predict the other two missing channels except for the brightness channel, which makes the model training more stable (Cao et al., 2017). We will introduce three commonly used color-light separable color spaces, including CIELAB, YUV, and HSV color space as follows.

**(a) CIELAB Color Space.** Different from RGB color space, CIELAB is a color space independent of equipment, and any colors can be expressed in CIELAB color space. This color space can describe the human visual experience in a digital manner. In CIELAB space, the L channel is independent of the color information and only contains the brightness information. The AB channel only contains color information, where A represents the red-to-green range and B represents the blue-to-yellow range. CIELAB space and RGB space can be converted into each other through XYZ space, which can be calculated by the following equations.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4125 & 0.3576 & 0.1804 \\ 0.0193 & 0.1192 & 0.9502 \\ 0.0193 & 0.1192 & 0.9502 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}; \tag{4}$$

$$\begin{aligned} L &= 116 f\left(Y/Y_n\right) - 16 \\ A &= 500\left[f\left(X/X_n\right) - f\left(Y/Y_n\right)\right] \\ B &= 200\left[f\left(Y/Y_n\right) - f\left(Z/Z_n\right)\right] \end{aligned}; \tag{5}$$

$$f(t) = \begin{cases} t^{1/3} & if \quad t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{4}{29} & otherwise \end{cases}; \tag{6}$$

where the default value of $X_n$, $Y_n$, $Z_n$ is 95.047, 100.0, 108.883.

CIELAB color space has been widely used in image colorization because of its perceptual consistency with human color vision (Yoo et al., 2019; Zhang et al., 2019; Lee et al., 2020; Chen et al., 2018a; Kong et al., 2021; Kim et al., 2021; Li et al., 2021c). In the research of Iizuka et al. (2016), it is shown that more sensible colorization results can be obtained in CIELAB color space compared with RGB color space. This space is also adopted in Xian et al. (2018) to solve the problem that color constraints input in RGB form need to struggle with semantic understanding of the network. This method converts ground truth images into CIELAB space, and constraints were applied in the L channel and AB channel respectively to obtain higher quality color images without introducing obvious visual artifacts. Similar colorization methods are also found in Cao et al. (2017), Zhao et al. (2020). In general, image colorization based on CIELAB space can obtain colorized images unified with human color visual perception.

**Table 1**
Description of several representative methods.

| Methods | Color space | Loss function | Network structure | Application | Automaticity | Description | Journal/ Conference |
|---|---|---|---|---|---|---|---|
| Cheng's (Cheng et al., 2015) | YUV | l2 loss | DNNs | Natrual image | fully-automatic | end-to-end, the first deep learning color method, huge reference images required | ICCV 2015 |
| Iizuka's (Iizuka et al., 2016) | CIELAB | cross-entropy loss+mse loss | CNNs | Natrual image | fully-automatic | end-to-end; data-driven; user uncontrollable; insufficient generalization | ACM TOG 2016 |
| Zhang's (Zhang* et al., 2017) | CIELAB | cross-entropy loss+smooth-l1 loss +regression loss | CNNs | Natrual image | fully-automatic+semi-automatic | end-to-end, data driven, interactive colorization, color bleeding | ACM TOG 2017 |
| Ci's (Ci et al., 2018) | RGB | adversarial loss+perceptual loss | cGANs | Line Art | semi-automatic | interactive colorization; color overflow; user guidance is required | ACM MM 2018 |
| Zhang's (Zhang* et al., 2018) | RGB | adversarial loss+mae loss | GANs | Sketch image | semi-automatic | interactive colorization; Two stage; not suitable for complex sketches | ACM TOG 2018 |
| Zhang's (Zhang et al., 2019) | CIELAB | adversarial loss+perceptual loss+temporal consistency loss+l1 loss+smooth loss | GANs | Video | semi-automatic | first end-to-end network for exemplar-based video colorization; reference image required | CVPR 2019 |
| Kim's (Kim et al., 2019) | RGB | adversarial loss+reconstruction loss+classification loss+changing loss | ACGANs | Line Art | semi-automatic | two stages; text tag-based | ICCV 2019 |
| Xu's (Xu et al., 2020) | CIELAB | huber loss | CNNs | Natural image | semi-automatic | end-to-end; reference-based; pretrained VGG19 | CVPR 2020 |
| Lee's (Lee et al., 2020) | CIELAB | similarity-based triplet loss+adversarial loss+perceptual loss+l1 loss+style loss | GANs | Sketch Image | semi-automatic | reference-Based, augmented-self reference is utilized | CVPR 2020 |
| Zhang's (Zhang et al., 2021a) | RGB | mse loss | CNNs | Line Art | semi-automatic | user-guided; split filling mechanism; simple but effective | CVPR 2021 |
| Manoj's (Manoj et al., 2021) | RGB | negative log-likelihood | Transformer | Natrual image | fully-automatic | First application of transformers for image colorization; self-attention; diverse colorization | ICLR 2021 |

**(b) YUV (YCrCb) Color Space.** YUV color space, also known as YCrCb, the Y component represents brightness, and the Cr and Cb components represent chrominance, which describes the color and saturation of an image, respectively. Similarly, YUV and RGB color spaces can be converted to each other by the conversion equation is shown below.

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & 0.331 & 0.5 \\ 0.5 & 0.419 & -0.081 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{7}$$

YUV color space takes human perception into account and is therefore more suitable for image colorization tasks. This color space can minimize the correlation between the three coordinate axes of the color space. Based on this feature, the colorized image has higher accuracy and better visual effects (Cheng et al., 2015; Xiao et al., 2019a). Since U and V are independent chroma signals, Liang et al. (2016) used two networks of the same architecture to output U and V respectively, which simplified the network structure and improved the accuracy of the structure. Cao et al. (2017) showed that colorized images obtained in RGB space suffer from structural loss due to an additional trade-off between L1 loss and GAN loss, while the obtained images in the YUV color space are smoother and more natural. To sum up, YUV color space is very popular in the application of image colorization tasks, and often can achieve good colorization effects.

**(c) HSV Color Space.** HSV color space is an intuitive color model, which is widely used in image editing tools. It is composed of three components: Hue (H), Saturation (S), and Value (V). The H component is represented by angle and ranges from 0 to 360. The S component represents the degree of similarity between the colors and the spectrum. The value of the S component ranges from 0 to 1, with the larger the value, the more saturated the color. The V component is the brightness of the color, usually ranging from 0 to 1, i.e., black to white. HSV and RGB color space can be converted to each other by equation set as follow.

$$H = \begin{cases} 0^\circ \&\ \ , if \& \Delta = 0 \\ 60^\circ \times \left( \frac{G'-B'}{\Delta} + 0 \right) \& \ \ , if \& C_{\max} = R' \\ 60^\circ \times \left( \frac{B'-R'}{\Delta} + 2 \right) \&, \ \ if \& C_{\max} = G' \\ 60^\circ \times \left( \frac{R'-G'}{\Delta} + 4 \right) \& \ \ , if \& C_{\max} = B' \end{cases} \tag{8}$$

$$S = \begin{cases} 0 \&, C_{\max} = 0 \\ \frac{\Delta}{C_{\max}} \&, C_{\max} \neq 0 \end{cases}$$

$$V = C_{\max}$$

where $R' = R/255, G' = G/255, B' = B/255$, and $C_{\max} = \max\left(R', G', B'\right)$, $C_{\min} = \min\left(R', G', B'\right), \Delta = C_{\max} - C_{\min}$. Since there are few colorization methods for images based on HSV color space in practical applications, these methods will not be described in detail here.

### 2.2.3. Multi-color space-based methods

Fig. 4 shows the representation of the same image in different color spaces. Observing Fig. 4, we can find that the color and brightness information of an image is separable in YUV, CIELAB, and HSV spaces, while in the RGB color space the color and brightness information are mixed. Different color spaces have their advantages and disadvantages,

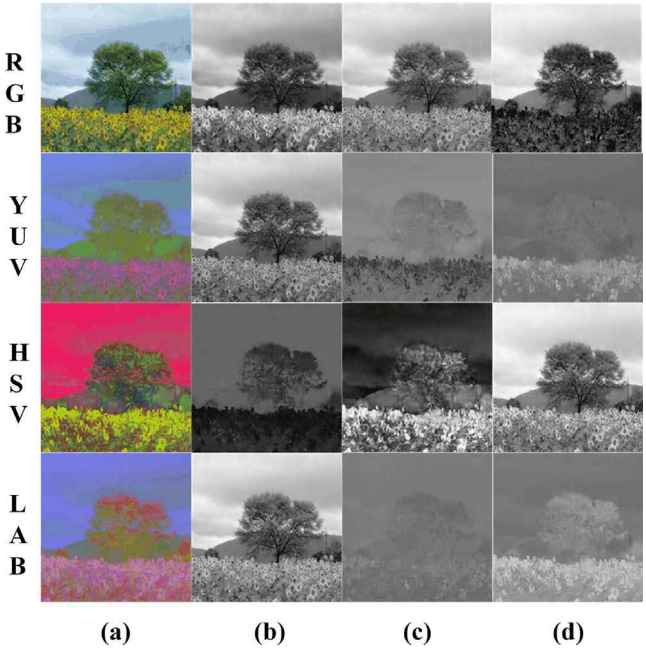



**Fig. 4.** The representation of images in different color spaces. (a) represents images in different color spaces; (b), (c), and (d) respectively represent three different components. Taking the first row as an example, (b), (c), and (d) respectively represent three components R, G, and B.

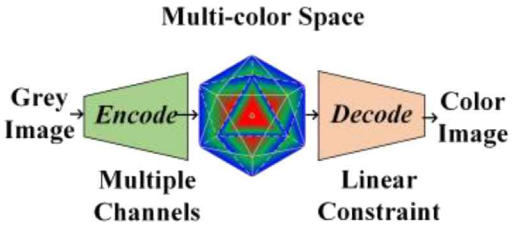



**Fig. 5.** Multi-color Space-Based Method, image from Zhou et al. (2020). First, the freedom of color is extended to a high-dimensional tensor, and then the linear autocorrelation constraint is obtained by using the intermediate results to guide the global direction of color more accurately. The use of multiple color spaces facilitates subsequent accurate colorization.

so many researchers began to study the DLIC method based on the combination of multiple color spaces in recent years. For instance, Hensman and Aizawa (2017) combined HSV with RGB color space to improve the reliability and accuracy of the colorization model. In Zhou et al. (2020), the prior information of multiple color spaces, including RGB, YCrCb, and HSV, was used as effective constraints to improve the performance of the colorization model (see Fig. 5). In general, model learning in multi-color spaces provides more redundancy and related samples for training, so that more reliable and accurate models can be obtained. The adoption of multi-color space, to a certain extent, expands the freedom of color space, and provides a richer color saturation. Therefore, multi-color space-based methods have strong development potential and are worth further exploration and research

### 2.3. Loss function for image colorization

In the field of image colorization, the loss function is used to evaluate the degree of inconsistency between the predicted image and ground truth, and it is also the objective function of optimization in neural networks. The smaller the loss function is, the better the robustness of the model will be. In the early stages, researchers usually used pixel-level loss (such as L1 loss and L2 loss) (Cheng et al., 2015; Su et al., 2018), but later we found that such loss could not accurately measure the difference between colorized image and ground truth. Therefore, other loss functions, such as perceptual loss (Ci et al., 2018; Chen and Hays, 2018) and total variation (TV) loss (Kuang et al., 2020; Liu et al., 2017; Johari and Behroozi, 2020a), are studied to better assess the error between the colorized image and ground truth image, to obtain more realistic and natural, higher-quality colorization results. The selection of loss function is an important factor affecting the model colorization performance. Next, we will introduce the loss function widely used in DLIC methods, and make a simple induction of the existing colorization methods. Some representative publications corresponding to commonly used loss functions are listed in Table 2.

In most studies, the researchers attributed the images colorization to a regression problem and solved it with a deep learning model. Specifically, a mapping function $F$ is learned using a series of gray-color pairs, which can then be used to transform a new grayscale image into a color image. To obtain great colorization performance, it is necessary to select an appropriate loss function to optimize the model.

#### 2.3.1. L1 loss

L1 loss, also known as Mean Absolute Error (MAE), calculates the absolute sum of the difference between the target image and the predicted image. This loss function can be used to measure the distance between the predicted image and the ground truth. The calculation is shown as follows.

$$L_1(\hat{Y}, Y) = \sum_{i=1}^{n} \left| \hat{Y}_i - Y_i \right| \tag{9}$$

where $\hat{Y}_i$ is the output of the colorization model, i.e., the pixel matrix of the colorized image. $Y_i$ is the pixel matrix of the true color image. L1 loss function is widely used in image processing, including image super-resolution, image restoration, image colorization, and other tasks. In Xiao et al. (2019a), L1 loss is used as a constraint condition to ensure that the gray image of the colorized image is consistent with the original gray image. In addition, the colorization results obtained by Seo and Seo (2021), Thasarathan and Ebrahimi (2019), Hou et al. (2019) show that the use of L1 loss can reduce the blur degree of the colorized image to a certain extent. However, because L1 loss only calculates the global error and ignores the local information, the colorized image lacks the fine texture information and is prone to artifacts (Berg et al., 2018). In addition, the colorization method using

**Table 2**
Different colorization method categories and representative publication.

| Category | Sub-category | Representative publication |
|---|---|---|
| Loss function | L1 Loss | Kuang et al. (2020), Xu et al. (2021), Ji et al. (2020), Dabas et al. (2020), Seo and Seo (2021), Cao et al. (2017), Su et al. (2018), Lee and Lee (2020), Kim et al. (2019), Zhang et al. (2017), Xie* et al. (2020), Mourchid et al. (2021), Chen et al. (2018a), Xiao et al. (2019a), Iizuka and Simo-Serra (2019), Yin et al. (2021) |
| | L2 Loss | Ci et al. (2018), Suarez et al. (2018), Dong et al. (2018), Cheng et al. (2015), Dabas et al. (2020), Su et al. (2018), Liang et al. (2016), Xu et al. (2020), Xian et al. (2018), Lee and Lee (2020), Liu et al. (2017), Du et al. (2021), Zhong et al. (2020) |
| | Smooth L1 loss | Zou et al. (2019), Yoo et al. (2019), He et al. (2018), Zhang* et al. (2017), Su et al. (2020), Zhang et al. (2019) |
| | Perceptual Loss | Ci et al. (2018), He et al. (2018), Kuang et al. (2020), Xu et al. (2021), Chen and Hays (2018), Zhang et al. (2019), Xu et al. (2020), Lee et al. (2020), Lei and Chen (2019), Yin et al. (2021) |
| | Total Variance Loss | Kuang et al. (2020), Liu et al. (2017), Johari and Behroozi (2020a,b) |
| | Adversarial Loss | Zhang* et al. (2018), Yoo et al. (2019), Suarez and Sappa (2017), Xu et al. (2021), Endo et al. (2021), Chen and Hays (2018), Zhang et al. (2019), Cao et al. (2017), Liu et al. (2017), Zhang et al. (2017), Johari and Behroozi (2020b), Mourchid et al. (2021), Xiao et al. (2019a), Hou et al. (2019), Zhong et al. (2020), Furusawa et al. (2017), Li et al. (2021c), Wu et al. (2021b), Yin et al. (2021) |
| Level of automation | Fully Automatic | Ramassamy et al. (2019), Zhang et al. (2016), Bian et al. (2021), Iizuka et al. (2016), Chybicki et al. (2019), Varga and Sziranyi (2016), Chen and Hays (2018), Seo and Seo (2021), Zbulak (2020), Liu et al. (2017), Mourchid et al. (2021), Hou et al. (2019), Kiani et al. (2020), Yu et al. (2015), Wan et al. (2020) |
| | Scribble-Based Methods | Ci et al. (2018), Zhang et al. (2021a), Lee and Lee (2020), Liu et al. (2017), Thasarathan and Ebrahimi (2019), Chen et al. (2019), Min et al. (2020), Li et al. (2020), Furusawa et al. (2017) |
| | Language-Based | Zou et al. (2019), Kim et al. (2019), Chen et al. (2018a), Bahng et al. (2018), Manjunatha et al. (2018) |
| | Reference-Based | Larsson et al. (2016), He et al. (2018), Kuang et al. (2020), Xuan et al. (2021), Zhang et al. (2019), Sun et al. (2019b), Xu et al. (2020), Zhang et al. (2017), Lee et al. (2020), Kong et al. (2021), Chakraborty (2019), Lee and Cho (2020), Iizuka and Simo-Serra (2019), Chen et al. (2020), Li et al. (2021c) |
| Application fields | Natural Image Colorization | Ramassamy et al. (2019), Larsson et al. (2016), He et al. (2018), Zhang* et al. (2017), Su et al. (2020), Zhang et al. (2016), Cheng et al. (2015), Iizuka et al. (2016), Dabas et al. (2020), An et al. (2020), Cao et al. (2017), Su et al. (2018), Deshpande et al. (2017), Xu et al. (2020), Johari and Behroozi (2020a), Mourchid et al. (2021), Chen et al. (2018a), Cheng et al. (2017), Bahng et al. (2018), Jin et al. (2021b), Li et al. (2021c), Wu et al. (2021b), Yin et al. (2021) |
| | Line Art Colorization | Ci et al. (2018), Zou et al. (2019), Zhang* et al. (2018), Yoo et al. (2019), Ramassamy et al. (2019), Zhang et al. (2021a), Varga and Sziranyi (2016), Chen and Hays (2018), Seo and Seo (2021), Lee and Lee (2020), Kim et al. (2019), Thasarathan and Ebrahimi (2019), Sun et al. (2019b), Xie* et al. (2020), Lee et al. (2020), Xian et al. (2018), Chen et al. (2020), Furusawa et al. (2017), Zhang et al. (2021), Cao et al. (2021), Casey et al. (2021) |
| | Infrared Colorization | Kuang et al. (2020), Suarez and Sappa (2017), Suarez et al. (2018), Xu et al. (2021), Dong et al. (2018), Zhong et al. (2020), Limmer and Lensch (2016), Yang et al. (2022) |
| | Remote Sensing Colorization | Ji et al. (2020), Gravey et al. (2019), Dias et al. (2020), Song et al. (2017), Huang et al. (2021), Li et al. (2018), Doi et al. (2020), Poterek et al. (2020), Ozcelik et al. (2020) |
| | Video Colorization | Endo et al. (2021), Zhang et al. (2019), Thasarathan and Ebrahimi (2019), Lei and Chen (2019), Vondrick et al. (2018), Iizuka and Simo-Serra (2019), Shi et al. (2020), Akimoto et al. (2020) |
| | Other Colorization | Dong et al. (2022), Xuan et al. (2021), Bian et al. (2021), Liang et al. (2021), Morra et al. (2021), Yu et al. (2020), Guo et al. (2021), Hou et al. (2019), Klein et al. (2020), Maejima et al. (2019), Aizawa et al. (2019) |
| Evaluation Metric | MSE | Zhang* et al. (2018), Suarez et al. (2018), Chybicki et al. (2019), Min et al. (2020), M.H. Baig (2017b), Teng et al. (2021), Chen et al. (2020) |
| | PSNR | Ramassamy et al. (2019), Larsson et al. (2016), Su et al. (2020), Kuang et al. (2020), Xu et al. (2021), Dong et al. (2018), Ji et al. (2020), Dong et al. (2022), Bian et al. (2021), Chybicki et al. (2019), Endo et al. (2021), Xiao et al. (2019b), M.H. Baig (2017b), Zbulak (2020), Su et al. (2018), Zhao et al. (2020), Thasarathan and Ebrahimi (2019), Johari and Behroozi (2020b), Lee et al. (2020), Kong et al. (2021), Zhou et al. (2020), Cheng et al. (2017), Lei and Chen (2019), Du et al. (2021), Chen et al. (2019), Min et al. (2020), Wan et al. (2020), Teng et al. (2021), Larsson et al. (2017b), Li et al. (2021c), Zhang et al. (2021), Wu et al. (2021b) |
| | SSIM | Ramassamy et al. (2019), Su et al. (2020), Kuang et al. (2020), Suarez et al. (2018), Xu et al. (2021), Ji et al. (2020), Dong et al. (2022), Bian et al. (2021), Dabas et al. (2020), Chybicki et al. (2019), Varga and Sziranyi (2016), Thasarathan and Ebrahimi (2019), Mourchid et al. (2021), Kong et al. (2021), Zhou et al. (2020), Du et al. (2021), Chen et al. (2019), Min et al. (2020), Wan et al. (2020), Li et al. (2021c), Zhang et al. (2021), Wu et al. (2021b) |
| | MS-SSIM | Kuang et al. (2020), Chybicki et al. (2019), M.H. Baig (2017b), Johari and Behroozi (2020b), Min et al. (2020), Teng et al. (2021) |
| | IS | Chen and Hays (2018), Johari and Behroozi (2020a) |
| | FID | Ci et al. (2018), Ji et al. (2020), Chybicki et al. (2019), An et al. (2020), Zhang et al. (2019), Manoj et al. (2021), Lee and Lee (2020), Kim et al. (2019), Thasarathan and Ebrahimi (2019), Johari and Behroozi (2020a), Lee et al. (2020), Cao et al. (2021), Wu et al. (2021b) |
| | EN | Xu et al. (2021), Kong et al. (2021) |
| | LPIPS | Yoo et al. (2019), Su et al. (2020), Seo and Seo (2021), Lei and Chen (2019), Kim et al. (2021) |
| | HaarPSI | Chybicki et al. (2019) |
| | Other | Ci et al. (2018), Kuang et al. (2020), Xu et al. (2021), Ji et al. (2020), Chybicki et al. (2019), Varga and Sziranyi (2016), Lee and Lee (2020), Kong et al. (2021), Min et al. (2020), Teng et al. (2021) |

L1 loss often obtains a color image with low color saturation, this is mainly because L1 loss tries to predict the pixel difference between the image and the ground truth on average (Dabas et al., 2020). Directly minimizing the L1 loss between the colorized image and ground truth greatly inhibits the color diversity (Chen and Hays, 2018).

#### 2.3.2. L2 loss

L2 loss, also known as Mean Squared Error (MSE) loss, is calculated as the square error between the pixel matrix of the predicted image and the target image. The calculation equation is described as follows:

$$L_2(\hat{Y}, Y) = \sum_{i=1}^{n} \left(\hat{Y}_i - Y_i\right)^2 \tag{10}$$

Compared with L1 loss, L2 loss is easier to solve, and the model converges faster with the same learning rate. Therefore, L2 loss is widely used in CV, such as style transfer (Chen et al., 2017, 2018b) photo-realistic image synthesis (Chen and Koltun, 2017), image super-resolution (Sajjadi et al., 2017). In the first DLIC method (Ramassamy et al., 2019), L2 loss function is used to minimize the square distance between the colorized image and ground truth to achieve model optimization. The use of L2 loss function can obtain a more natural image colorization effect, but it is not robust to the inherent multimodal properties in the image colorization problem (Larsson et al., 2016; Zhang et al., 2016). In addition, because L2 loss cannot correctly learn the global background of the image, the colorized image has obvious colorization errors (Iizuka et al., 2016). To sum up, L1 loss and L2 loss have their pros and cons. while the occurrence of huber loss (smooth-l1 loss) combines the advantages of both.

#### 2.3.3. Smooth-l1 loss

Smooth-l1 loss combines the advantage of L1 and L2 loss, which can be calculated by:

$$L_\delta(\hat{Y}, Y) = \begin{cases} \frac{1}{2}\left(\hat{Y} - Y\right)^2 & for \left|\hat{Y} - Y\right| \le \delta \\ \delta\left|\hat{Y} - Y\right| - \frac{1}{2}\delta & otherwise \end{cases} \tag{11}$$

where $\hat{Y}$ is the colorized image and $Y$ is ground truth. In general, $\delta = 1$. L1 and L2 loss are in many cases replaced by smooth-l1 loss. For instance, Zou et al. (2019) replaced L1 loss with Smooth-l1 loss to overcoming the problem of excessive color differences due to slight differences in corresponding RGB values. In He et al. (2018), Zhang* et al. (2017), Smooth-l1 loss was used as a distance measure to avoid the usage of average solutions in fuzzy colorization problems.

#### 2.3.4. Perceptual loss

Perceptual loss (Justin et al., 2016) was originally proposed for single-image super-resolution and style transfer tasks. Perceptual loss is more robust to measures image similarity than pixel loss functions, such as L1, L2 loss, and others mentioned above. The initial perceptual loss includes two parts: the feature reconstruction loss and the style reconstruction loss, which are used to measure the content and style differences between images, respectively. The feature reconstruction loss $L_f^{\phi,j}$ and the style reconstruction loss $L_s^{\phi,j}$ can be calculated by the Eqs. (12) and (14).

$$L_f^{\phi,j}\left(\hat{Y} - Y\right) = \frac{1}{C_j H_j W_j}\left\|\phi\left(\hat{Y}\right) - \phi(Y)\right\|_2^2 \tag{12}$$

$$G_j^{\phi}(X)_{c,c'} = \frac{1}{C_j H_j W_j}\sum_{h=1}^{H_j}\sum_{w=1}^{W_j}\phi_j(X)_{h,w,c}\phi_j(X)_{h,w,c'} \tag{13}$$

$$L_s^{\phi,j}\left(\hat{Y} - Y\right) = \frac{1}{C_j H_j W_j}\left\|G_j^{\phi}\left(\hat{Y}\right) - G_j^{\phi}(Y)\right\|_F^2 \tag{14}$$

where $\hat{Y}$ and $Y$ represent the predicted image and the real image, respectively. $G_j^{\phi}$ computes the Gram matrix of $C_j \times C_j$, $\phi_j(Y)$ is the output of input $Y$ at the $j$th layer in the loss network $\phi$, and its shape is $C_j \times H_j \times W_j$. When perceptual loss is used for image colorization, only feature reconstruction loss is usually used. This loss function can make up for the defects of the pixel-level loss function and better measure the perceptual and semantic differences between images. In Kuang et al. (2020), the perceptual loss was applied to the thermal image colorization task to recover texture information. For the multimodal problem of video colorization, Lei and Chen (2019) used a perceptual loss with diversity to distinguish the various modality in the solution space. Perceptual loss measures semantic differences caused by unnatural colorization, which is robust to appearance differences caused by two plausible colors. However, perceptual loss also has some limitations, such as the inability to use unusual or artistic colors to color images (He et al., 2018).

#### 2.3.5. Total variance loss

TV loss (Rudin et al., 1992) is defined as the sum of absolute differences between adjacent pixels, which can be calculated by the following equation.

$$L_{tv} = \sqrt{\left(y_{i+1,j} - y_{i,j}\right)^2 + \left(y_{i,j+1} - y_{i,j}\right)^2} \tag{15}$$

To suppress the noise in the generated image, Liu et al. (2017) introduced TV loss in image colorization. This loss avoids the color mutation problem in the output image, this is because TV loss can constrain the pixel changes in the generated results and improve the smoothness level of the image. Johari and Behroozi (2020a) also found that the training process of GANs can be stabilized by using the TV loss function.

#### 2.3.6. Adversarial loss

In recent years, GANs have been widely used in various image processing tasks because of its powerful generative capability, such as image generation, image inpainting, image super-resolution, and image colorization. In general, GANs consist of two parts: generator and discriminator, and the model is optimized by adversarial loss. Four adversarial loss functions that are widely used in DLIC methods, including GANs (Goodfellow et al., 2014), Least Squares GANs (LSGANs) (Mao et al., 2017), Wasserstein GANs (WGANs) (Arjovsky et al., 2017), and WGAN-GP (Ishaan et al., 2017) are introduced below.

The initial adversarial loss is represented by the cross-entropy loss, which can be calculated by:

$$\begin{aligned}\min_G \max_D V(D,G) = {} & E_{x\sim P_{data}(x)}\left[\log D(x)\right] \\ & + E_{z\sim P_z(z)}\left[\log\left(1 - D(G(z))\right)\right].\end{aligned} \tag{16}$$

For discriminator D, the equation is maximized as much as possible, and for generator G, the opposite is true. The training for GANs is an iterative process that repeats the following steps: (a) Fix G, train D's discrimination skills; (b) Fix D, train G's generation capability; until G and D reach dynamic equilibrium, i.e., the image generated by G is consistent with the distribution of the real image, D cannot distinguish between the generated image from real image. Compared with conventional CNNs, GANs improve the quality of generated images through adversarial learning between generator and discriminator.

However, because cross-entropy loss only depends on the authenticity of the generated image, and ignores the pixel difference between the generated image and the real image, the gradient disappearance problem occurs in generator. To solve this problem, Mao et al. (2017) used the least-squares loss function instead of the cross-entropy loss to obtain higher quality generation results and more stable training process. The least-squares adversarial loss can be calculated by the following equation set.

$$\begin{aligned}\min_D V_{LSGAN}(D) = {} & \frac{1}{2}E_{x\sim P_{data}(x)}\left[(D(x) - b)^2\right] \\ & + \frac{1}{2}E_{z\sim P_z(z)}\left[(D(G(z)) - a)^2\right] \\ \min_G V_{LSGAN}(G) = {} & \frac{1}{2}E_{z\sim P_z(z)}\left[(D(G(z)) - c)^2\right]\end{aligned} \tag{17}$$

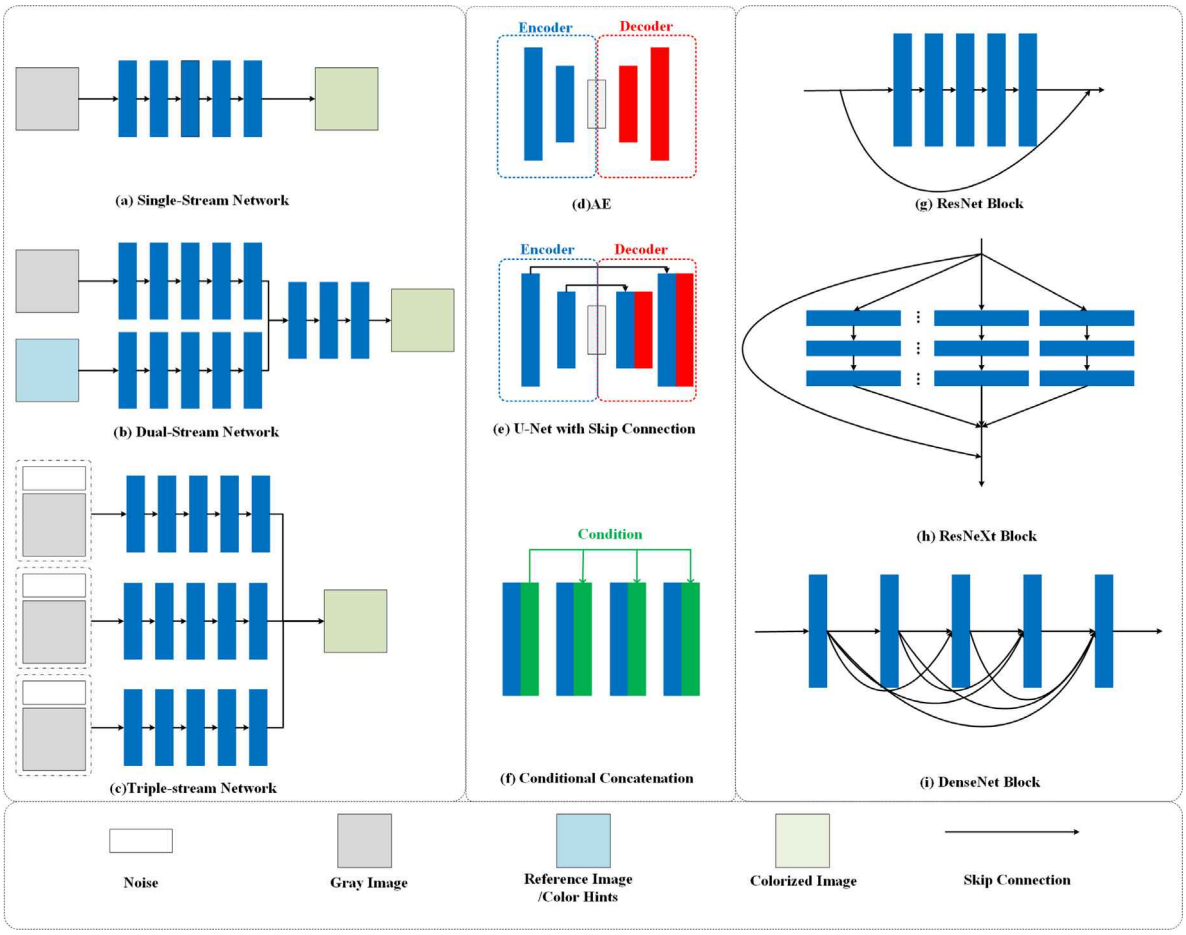



**Fig. 6.** Typical network structures and key module structures. (a) Single-stream network, where a grayscale image is input into an end-to-end network to obtain a colorized image. (b) Dual-stream network, where a reference image or other auxiliary information (color hints, text descriptions, etc.) is input in addition to the grayscale image, through two networks with similar or different structures to obtain a colorized image. (c) Triple-stream network, i.e., with three inputs, typically R, G, and B images and noise vectors concatenated together into a network with the similar structure to obtain the colorized image. (d) AE, i.e., encoder–decoder structure. (e) U-Net with skip connection, i.e., skip connection is added between encoder–decoder. (f) Conditional concatenation, i.e., adding conditions to each layer of the network. (g) ResNet Block. (h) ResNext Block. (i) DenseNet Block.

where $z$ is random noise, $P_{data}(x)$ is the probability distribution obeyed by the real data $x$, $E_{x\sim P_{data}(x)}$ is the expectation. Constants $a$ and $b$ represent the markup of the real image and the generated image, respectively. In general, $a = c = 1$ and $b = 0$. Similarly, a Wasserstein distance-based adversarial loss was proposed by Arjovsky et al. (2017) to solve the problem of GANs training instability. This loss function can ensure the diversity of generated samples. Further, on the basis of Wasserstein adversarial loss, Ishaan et al. (2017) set up an additional loss to limit the gradient of the discriminator, as equation set (18), allowing the generation model to converge faster and produce higher quality samples.

$$\begin{aligned} L(D) = & \underbrace{-E_{x\sim P_{data}(x)}(D(x)) + E_{x\sim P_{data}(x)}[D(G(x))]}_{Wasserstein \quad adversarial \quad loss} \\ & + \underbrace{\lambda E_{x\sim P_{penalty}(x)}[\|\nabla D(x)\| - 1]^2}_{gradient \quad penalty} \\ L(G) = & -E_{x\sim P_{data}(x)}[D(G(x))] \end{aligned} \tag{18}$$

In the image colorization, the adversarial loss is mainly responsible for supervising the quality of the generated color image and making it conform to the true distribution (Isola et al., 2017). While the adversarial loss makes the network produce clear and realistic colorized images, it inevitably brings some problems, such as the network's understanding of color sometimes conflicts with the user's color constraints. For example, the user provides a rainbow color constraint for the ocean, but the adversarial network thinks it looks fake and prevents the generator from producing such an output, but in fact the colorful ocean is reasonable (such as the ocean under a sunset glow). For this limitation, Xian et al. (2018) only applied adversarial loss to grayscale images, making the discriminator focus on generating sharp realistic details while ignoring the color information of the image. In addition, it is well known that the usage of adversarial loss tends to produce distorted textures in image synthesis tasks, this problem also occurs in image colorization tasks. Therefore, the adversarial loss is usually combined with loss functions, such as pixel-level loss and perceptual loss, to improve the colorization performance of the image colorization model (Ci et al., 2018; Chen and Hays, 2018; Zhang et al., 2019; Cao et al., 2017; Liu et al., 2017; Zhang et al., 2017; Johari and Behroozi, 2020b; Mourchid et al., 2021; Chen et al., 2018a).

## 3. Representative network architectures

Nowadays, network design is an important issue in deep learning. In image colorization, researchers apply various design strategies to construct the network structures. In this section, we divide image colorization methods into four categories from the perspective of network structures, including CNNs-based, GANs-based, Transformer-based, and other methods. In addition, we have a simple classification of the backbone structure and key modules in the colorization model, which is shown in Fig. 6.

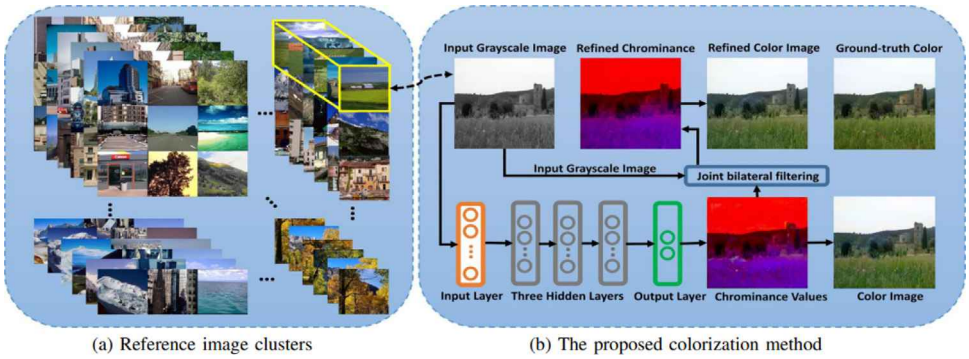



**Fig. 7.** The pipeline of Deep colorization (Cheng et al., 2015). The first deep learning based image colorization method.

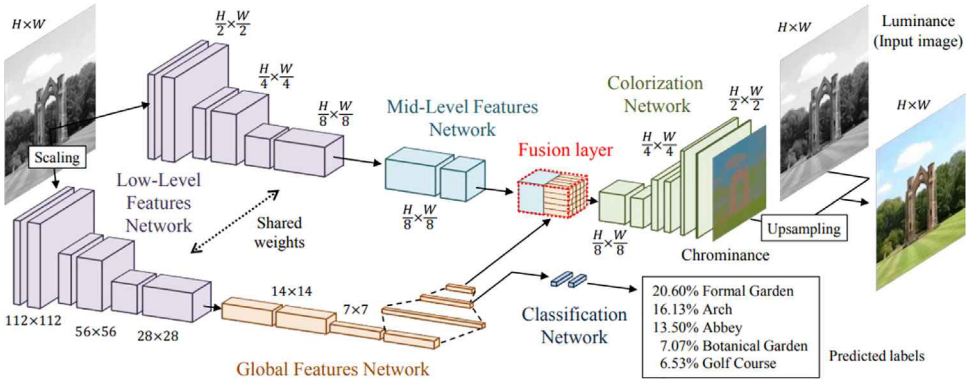



**Fig. 8.** The pipeline of Iizuka et al. (2016). A representative study of dual-stream networks.

### 3.1. CNNs-based methods

In recent years, with the development of deep learning, CNNs has made great achievements in image processing with its strong feature learning ability. Particularly, CNNs-based colorization methods are also proliferating and achieving impressive results.

#### 3.1.1. Deep Colorization (Cheng et al., 2015)

In 2015, the first DLIC method based on a deep neural network was proposed by Cheng et al. (2015), and the pipeline of this method is shown in Fig. 7. The method first divides the reference images into different clusters using the proposed adaptive graph clustering method, and then trains a colorization model for each cluster image collection separately. Specifically, given a grayscale image, the proposed method first automatically searches for the closest clusters and the corresponding pre-trained models; then each pixel point is extracted with feature descriptors as input to the neural network, and the output is the chromaticity of the corresponding pixel, followed immediately by combining the output with the grayscale pixel values to obtain the corresponding color values. Finally, a joint bilateral filter (using the grayscale image as a guide) is used to further adjust the output color image. Although this method performs well for natural image colorization, its application is largely limited by the need to use reference images.

#### 3.1.2. Let there be Color! (Iizuka et al., 2016)

Iizuka et al. (2016) is a representative work based on a dual-stream network structure. This work proposes a method for image colorization combining global features with local features. The proposed model consists of four main components: a low-level feature network, a mid-level feature network, a global feature network, and a colorization network. All components are tightly coupled and trained in an end-to-end manner. The output of the model is the chromaticity of the image, which is fused with the luminance to form the colorized image. The proposed method can be generalized to many types of images because this method learns information from a large dataset in an end-to-end manner. Fig. 8 shows the pipeline of Iizuka et al. (2016)

Colorization methods based on CNNs have also been applied by other researchers (Larsson et al., 2016; Zhang et al., 2016). CNNs-based colorization methods usually require a large-scale reference image dataset to train the learning model to realize image colorization. However, it is difficult to obtain the image dataset containing all the objects to train the neural network model in the actual training process, which greatly limits the performance of this method. Another disadvantage is that such methods tend to assign only one color to the same object, whereas in practice there are multiple potential colors.

### 3.2. GANs-based methods

GANs was proposed in 2014, and it has demonstrated great application potential in CV with its powerful generation capabilities (Goodfellow et al., 2014; Mao et al., 2017; Arjovsky et al., 2017; Ishaan et al., 2017). GANs also evolved from the original model to conditional GANs(cGANs) (Isola et al., 2017), CycleGANs (Zhu et al., 2017), and most recently StyleGAN3(Karras et al., 2021), and so on. With the development of GANs, image colorization methods based on GANs have achieved good colorization performance. In GANs-based colorization models, the main difference usually is the network structure, especially the generator structure and the loss function. Most of these colorization methods are based on classic generative models (such as cGANs and CycleGANs) to make these models more suitable for image colorization. Next, we will summarize the general architecture and key modules of the generator by category.

Generator structures in colorization models can be divided into single-stream networks, dual-stream networks, and multi-stream networks. The structure of single-stream networks is shown in Fig. 6(a), which can be seen in the research (Ci et al., 2018; Xu et al., 2021; Seo and Seo, 2021; Cao et al., 2017; Liu et al., 2017; Li et al., 2021b;

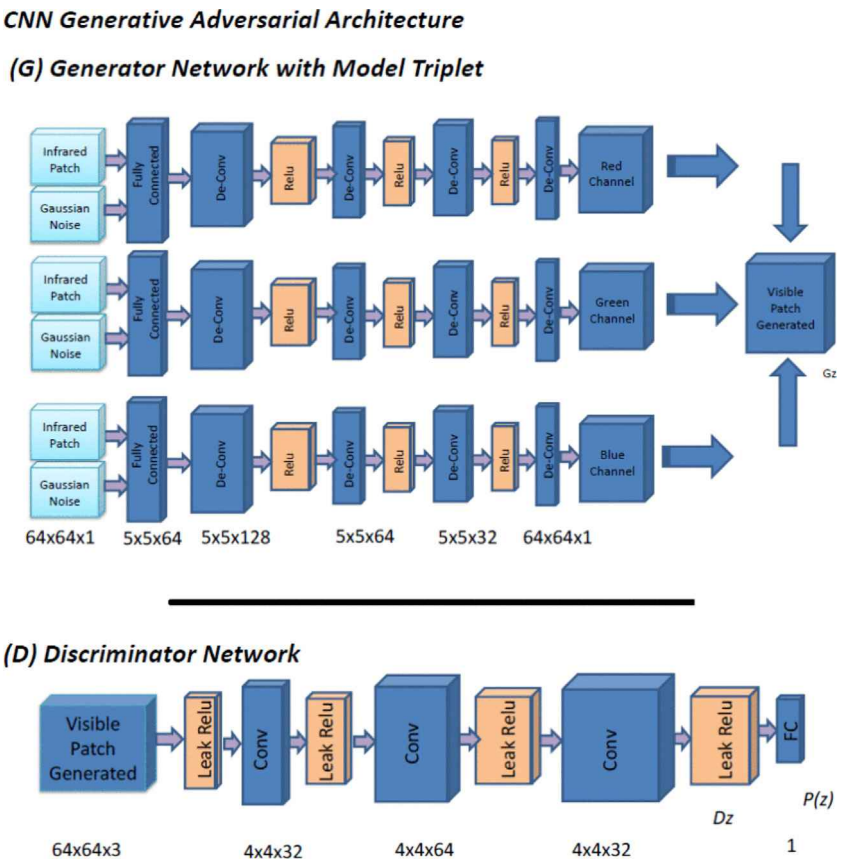



**Fig. 9.** The pipeline of Suarez and Sappa (2017). A representative study of triple-stream networks.

Mourchid et al., 2021; Silva et al., 2019; Huang et al., 2021). Single-stream network usually takes a grayscale image as input directly and outputs a colorized image after a series of convolution, pooling, activation, and deconvolution operations. Unlike single-stream networks, dual-stream networks have two inputs, one of which is a grayscale image and the other input is an auxiliary condition that provides color information (which may be a reference image, a color scribble or a text description); correspondingly, there are also two branch sub-networks, which may have similar or very different structures. One branch network is used to extract input grayscale images for feature extraction, and the other branch network is used to extract features of reference images or other color cues (Kuang et al., 2020; Iizuka et al., 2016; Sun et al., 2019b; Lee et al., 2020; Kong et al., 2021; Du et al., 2021). The adoption of a dual-stream network results in a colorized image with richer and more natural color information, due to the reference image as well as other additional information. However, the access to additional information often has high aesthetic and professional requirements for the user, thus largely limiting the application of this type of network. The structure of dual-stream networks is shown in Fig. 6(b). The multi-stream network is not limited to triple-stream networks, as shown in Fig. 6(c), but also includes networks with more branches. For example, the three-branch colorization method proposed by Suarez and Sappa (2017), which is shown in Fig. 9. In this method, infrared image block and gaussian noise are input into three branch networks to obtain R, G, and B channels respectively. This network structure is also seen in Zhang* et al. (2017), Suarez et al. (2018).

Furthermore, there are some differences between the backbone structure and key modules in the model, which greatly affects the color performance of the model. Next, we will focus on the backbone network and key modules that are widely used in image colorization models. The backbone network includes the autoencoder (AE), U-Net (Ronneberger et al., 2015) with skip connection, and conditional concatenation. Key modules include ResNet block, ResNeXt block, DenseNet block, as shown in Fig. 6(g), Fig. 6(h), and Fig. 6(i) respectively.

#### 3.2.1. AE

AE consists of an encoder and a decoder, as shown in Fig. 6(d). In the generator of image colorization model, it is often combined with other modules, such as ResNet block (Ji et al., 2020). The usage of these modules can deepen the network and improve the feature extraction capacity of the network. However, the colorized images are not satisfied, because of the lack of the underlying feature (Chybicki et al., 2019).

#### 3.2.2. U-Net

The U-Net network was originally proposed for the image segmentation (Huang et al., 2021), which adds layer-by-layer connections between encoders and decoders to form a U-shaped structure, as shown in Fig. 6(e). In colorization models, skip connection can help deconvolution to reconstruct the color image by fusing the low-level features and high-level features. In addition, ResNeXt block (Cao et al., 2017; Lee and Lee, 2020), ResNet block (Ji et al., 2020; Seo and Seo, 2021; Lee et al., 2020), and other modules (Xu et al., 2021; Li et al., 2021b) are also used in U-Net structure to further improve the quality of the colorized image. For example, Xu et al. (2021) combined DenseNet block with U-Net for colorization of near-infrared face images. In Lee and Lee (2020), ResNeXt block was combined with U-Net structure to enhance the quality of colorized images. A similar approach can be seen in Seo and Seo (2021).

#### 3.2.3. Conditional concatenation

However, the encoder–decoder structure is more inclined to extract global features with or without the addition of skip connections, which is more suitable for global shape transformation tasks. However, spatial local guidance is as important as global features in image colorization. Local guidance can ensure that target boundaries in colorized images are accurately separated by different generated colors. Therefore,

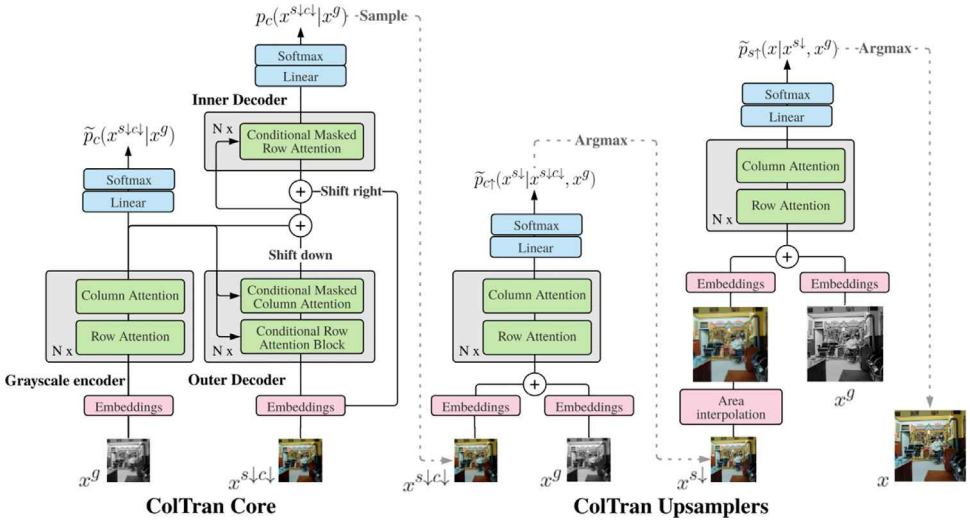



**Fig. 10.** Structure of the ColTran (Manoj et al., 2021). This model consists of 3 individual models: an autoregressive colorizer (left), a color upsampler (middle) and a spatial upsampler (right). ColTran core (autoregressive colorizer) is an instantiation of Axial Transformer with conditional transformer layers and an auxiliary parallel head.

conditional concatenation module inputs auxiliary information (text description, noise, category labels, etc.) layer by layer in the generative network to enhance the color diversity of the generated image. For example, Cao et al. (2017) input grayscale images as conditions into each layer of the network to provide continuous condition supervision, as Fig. 6(f). Similarly, gaussian noise was input into each layer of the generator to increase the diversity of colors in Suarez et al. (2018).

### 3.3. Transformer-based methods

Transformer (Carion et al., 2020) was originally proposed as a sequence-to-sequence model for machine translation. Recent studies have shown that Transformer-based pre-training model performs better in a variety of tasks, including CV, audio processing, and even chemistry and life sciences. Manoj et al. (2021) introduced Transformer into the image colorization task for the first time, realizing high-fidelity image colorization. The method first uses a conditional self-regression Transformer to generate a low-resolution coarse-colorized image, and then the obtained image is up-sampled through two completely parallel networks to obtain the fine color high-resolution images. Fig. 10 shows network structure of Manoj et al. (2021). More recently, Casey et al. (2021) introduced Transformer into the colorization of animation images, named Animation Transformer (AnT), which uses a Transformer based architecture to learn the spatial and visual relationships between segments across a sequence of images. This method provides a practical and advanced AI-assisted colorization method for professional animation workflow.

### 3.4. Other methods

In addition to the colorization methods mentioned above, a few image colorization methods based on other network structures have also been proposed. For instance, Zbulak (2020) adopted a Capsule Network (CapsNet) that was studied to solve the image colorization problem. Zhao et al. (2020) used conditional PixelCNN to generate pixel-by-pixel distribution and realize pixel semantic image colorization. In Liang et al. (2016), an image colorization method based on vectorized convolutional neural network(VCNN) was proposed. In addition, some colorization methods adopt the idea of transfer learning (Kiani et al., 2020) to transfer the pre-trained network model to realize image colorization.

## 4. Level of automation

In order to more comprehensive understanding of the DLCI methods, this section divides colorization methods into automatic colorization and semi-automatic colorization based on the automation level of colorization. Further, according to the different input of the model, the semi-automatic image colorization methods can be divided into scribble-based method, language-based methods, and reference-based methods. Several representative colorization methods with different levels of automation are shown in Table 2.

### 4.1. Fully automatic methods

The fully automatic image colorization method does not require any human intervention, nor does it require the colorization method of image pre-processing and post-processing operations. This method often requires learning a direct mapping from grayscale images to color images on large-scale datasets without artificially providing reference images or other color cues, such as color scribble and color palette. The fully automatic method can combine low-level detail information and semantic information to receive realistic natural colorization results. In particular, the data-driven deep networks have relieved users of the burden of retrieving high-quality training images with the development of large datasets.

Although automatic colorization methods have yielded impressive automatic color results (Zhang et al., 2016; Iizuka et al., 2016; Chen and Hays, 2018; Seo and Seo, 2021; Zbulak, 2020; Liu et al., 2017), there are some limitations: this kind of method can only cover a small part of scenes, because it is difficult to find a large dataset with all target objects or scenes for model training. In addition, another limitation of fully automatic colorization is that it does not provide a favorable colorization effect, and it is difficult to meet the specific requirements of different users, i.e., users cannot manipulate the output image with the colors they want. More fundamentally, the color of an object is essentially ambiguous, such as leaves, which can be green, brown, or yellow, but the existing fully automatic method often can only select a single color, and the colorization effect is not necessarily reasonable. Therefore, the fully automatic image colorization method has a lot of research space, it is worth further exploration.

### 4.2. Semi-automatic methods

The Semi-automatic colorization method allows users to control the color of the output image, which can be roughly divided into three methods: scribble-based methods, language-based methods, and reference-based methods.

#### 4.2.1. Scribble-based methods

Early scribble-based methods employed user-provided scribble or palette to color adjacent pixels with similar intensity values, which relied heavily on user input and usually required a large amount of user interaction to achieve fine colorization effects. Subsequent studies have improved such methods, especially as deep learning has evolved, and learning-based interactive scribble-based methods have begun to appear. For example, Ci et al. (2018) proposed a method to generate color illustrations based on a given line art image and color scribble. In Chen et al. (2019), a method of automatically generating scribble was proposed by Chen et al. the generated scribble was placed in the region of minimum entropy, which improved the credibility of color propagation. Recently, a colorization model based on the total variation of natural vectors was proposed by Min et al. (2020), this method solves the color overflow problem that exists in the scribble-based method to some extent. More recently, Zhang et al. (2021a) proposed a line art plane fill method that can calculate the "impact area" of the user's color scribble, this method reduced color leakage/pollution between scribbles.

These methods have data-centric properties that reduce the artist's burden and produce visually pleasing colorization results. Inevitably, there are some limitations in these methods, which are often used in animation images, line art images, and other artistic images, but they are not applicable for real natural images. For real natural scene colorization, it is difficult to provide accurate color even if users know what color the target object should be. An example is that it is often difficult for humans to accurately describe the colors of sunset glow because they are often colorful and variable.

#### 4.2.2. Reference-based method

Unlike the scribble-based method, the reference-based methods use the color information of the reference image to realize the target image colorization. In particular, large datasets have become more accessible in recent years, largely eliminating the time-consuming problem of selecting reference images. Therefore, reference-based colorization methods are revitalized in the deep learning era. Sun et al. (2019b) used reference images to specify the structure and color style of the icon to achieve the purpose of customizing the icon. In He et al. (2018), He et al. adopted an image retrieval algorithm, which takes into account the semantic information and low-level brightness information of the image, to randomly recommend the reference image, and then used the selected image's color information to colorize the image. Similar reference-based methods have been proposed in Kuang et al. (2020), Xuan et al. (2021), Xu et al. (2020), Kong et al. (2021), Chakraborty (2019), Lee and Cho (2020).

In general, the colorization results obtained by the reference-based methods largely depend on the selection of the reference images. Therefore, the key of the reference-based methods is to find the reference image which is highly related to the gray image in content (such as objects, illumination, and viewpoint). However, the existing methods are either very sensitive to the selection of reference images or require a large amount of time and resources that are difficult to apply to real-time colorization.

#### 4.2.3. Language-based methods

The language-based image colorization method employed text description and gray image to generate the specified color image. Compared with the scribble-based methods, these methods can reuse the same set of instructions to achieve consistent colorization of a group of sketches that contain similar objects. This is a challenge for scribble-based methods because there is a direct and fixed relationship between scribble and specific sketch areas. Therefore, the language-based methods can be considered as a complement to the scribble-based methods. The first colorization method based on text descriptions was proposed by Chen et al. (2018a), which used attention mechanisms to integrate natural language description and image features, and realizes language-based image colorization. Subsequently, Bahng et al. (2018) proposed a method that can use text semantics to generate a palette, and then used the generated palette to obtain the colorized image, which is more focused on using text to generate palettes than directly colorizing images. For the color artifact problem, Zou et al. (2019) designed a language-based interactive colorization system for scene sketches. The system allows users to interactively locate specific instances of foreground objects, and then through language instructions to meet various colorization needs progressively.

However, language-based colorization methods also have some limitations. Most of the existing methods can only deal with individual language instructions, and cannot understand context information. In addition, some information contained in the input instruction cannot be recognized and processed, like blonde hair, the "wheels" of the car. More fundamentally, the colorized image obtained by the language-based methods often has artifacts and uneven color effects.

## 5. Applications of image colorization

In order to have a more comprehensive understanding of the development of colorization, this section introduces the application field of colorization in detail. DLIC methods can be roughly divided into six categories according to different application fields: natural image colorization, line art image colorization, infrared image colorization, remote sensing image colorization, video colorization, and other colorization methods. We present representative work for each of these six colorization application areas in Table 2.

### 5.1. Natural image colorization

According to the different scenes, the natural image colorization method can be divided into indoor scene colorization and outdoor scene colorization. Indoor scene mainly includes bedroom, restaurant, and other scenes (Cheng et al., 2015; Cao et al., 2017; Xiao et al., 2019a), the outdoor scene is not limited to the sky, ocean, mountains, grasslands, deserts, and other natural scenes (He et al., 2018; Zhang* et al., 2017; Xiao et al., 2019b; An et al., 2020; Su et al., 2018; Johari and Behroozi, 2020a; Kong et al., 2021), but also buildings, billboards, transportation, and other artificial objects (Iizuka et al., 2016; Zbulak, 2020; Deshpande et al., 2017; Zhou et al., 2020; Silva et al., 2019), as well as human, animal (Larsson et al., 2016; Xu et al., 2020; Johari and Behroozi, 2020b; Xian et al., 2018; Kiani et al., 2020; Lee and Cho, 2020). Natural scene images are various, which contain different objects, so natural image colorization is a challenging research topic. Especially, the fully automatic image colorization method is usually data-driven, that is, it needs a large number of real color image datasets to carry out model training. However, it is difficult to obtain a color natural image dataset that contains all objects of all scenes. In addition, the task of natural image colorization, except historical images, is not particularly meaningful because the acquisition of color images is no longer a problem contemporarily. However, in terms of historical images, how to obtain color images that not only retain the real historical scenes but also restore the history as far as possible is still a problem to be solved urgently. This is not just because large reference image datasets are hard to obtain, but more because things that existed in old photographs do not exist today.

### 5.2. Line art colorization

Line art images exist in the fields of interior design, animation creation, and video editing (Lee and Lee, 2020; Kim et al., 2019; Liu et al., 2017; Thasarathan and Ebrahimi, 2019; Sun et al., 2019b; Li et al., 2021b). However, there is a limit to using only black and white line art to convey the complex emotional changes and atmosphere of a scene, so it is customary to use color image to convey ideas. The colorization of line art not only relies on the designer's imagination to match the color of objects, but also needs to consume a lot of time and effort. According to the different application scenes, the line art colorization methods can be divided into manga colorization and other line art colorization. The former is mostly utilizing color scribble or color clues to achieve image colorization. The latter is similar to image synthesis or image-to-image translation, that is, the sketch image is inputted into the deep learning model to generate colorized natural images with rich detailed texture information. These two methods are described in the next two sections.

#### 5.2.1. Mange colorization

Mange colorization is not only an interesting research topic, but also has potential applications in digital entertainment. However, creating an impressive and expressive animation requires a good color composition and proper use of textures and shadows, which means that even experienced artists can spend a huge amount of time and effort. Therefore, both fully automatic and semi-automatic methods can reduce the artist's workload to a certain extent. Unlike natural images colorization, mange sketch colorization is more challenging because sketch images contain only a few lines and no texture or shadow information. In addition, it is difficult to obtain the ground truth-sketch image pairs due to the limitation of comic copyright, which also increases the uncertainty of model generalization largely.

Therefore, most manga colorization methods employed color cues to achieve fully automatic or semi-automatic colorization. These methods are based on prior knowledge learned on composite sketches to realize the line art image colorization. For example, Ci et al. proposed a scribble-based user-guided animation line-art colorization method (Ci et al., 2018). Zhang et al. (2021a) proposed the user-guided colorization method, which can calculate the influence area of the user's color scribble and can well avoid color overflow. However, the colorized results obtained by the scribble-based methods are often too monotonous and even contain unrealistic colors or artifacts. On this basis, other methods use two-stage model to realize the colorization of line-art images (Zhang* et al., 2018; Silva et al., 2019). In the first stage, these methods guess the color areas based on given color clues and splashing colorful colors on the sketch to get a colorized draft. The purpose of this stage is to increase the richness of the color scheme. Although the generated sketch image may contain more coloring errors and blurred textures, it is full of rich, vibrant color schemes. The second stage is to refine the obtained drafts, including the detection of unnatural colors and color enhancement, to obtain a high-quality colorized image with natural realism. Fig. 11 shows the two-stage line art image colorization process.

#### 5.2.2. Other sketch colorization

In addition to the manga colorization mentioned in the previous section, line art images are widely used in other fields, such as interior design and icon design. For the human face and animal sketch, Lee et al. (2020) used the same image with geometric distortion as a virtual reference to realize the transformation from sketch image to ground truth image. Similarly, Li et al. (2021b) proposed an automatic colorization model in interior design, which can generate interior design images of different styles based on reference images. In Sun et al. (2019b), colorization techniques were applied to the contours of icons (such as banners, signboards, billboards, homepages, and mobile applications). In general, although the existing line art colorization method is not perfect, its extensive application scenarios and huge application potential. It is worth researchers to further study to explore the application of line art colorization methods in more fields.

### 5.3. Infrared colorization

With the continuous development of sensors, the thermal infrared images have gradually expanded from the initial military reconnaissance to modern environmental monitoring, security monitoring, and other fields. This section introduces the main technical areas and limitations of infrared image colorization to understand their characteristics and performance. Fig. 12 shows some typical infrared image colorization models.

Unlike the RGB cameras, infrared imaging can not only work all-weather, but also benefit from the super penetration of infrared radiation, which can overcome some visual obstacles such as clouds and fog, to obtain more information (such as pedestrians, animals, road and roadside information). However, the gray value of infrared images obtained in low-light or night vision environments is seriously homogenized, and has the disadvantages of low resolution and poor interpretability. In addition, compared with visible image colorization that only estimates the chromaticity of the image, the infrared image colorization needs to estimate the brightness and chromaticity simultaneously. Moreover, the objective feature (thermal feature) of the thermal infrared image has no necessary relationship with its visible appearance (perceived color), which further increases the difficulty of colorization.

Existing visible image colorization methods can also be applied to near-infrared image colorization, but the obtained image is often difficult to present the image with the real environment color and lacks the high-frequency detail information. Therefore, many infrared image colorization methods have been proposed to obtain visually perceptive high-quality color infrared images (Kuang et al., 2020; Suarez and Sappa, 2017; Suarez et al., 2018; Xu et al., 2021; Dong et al., 2018). Suarez and Sappa (2017) proposed a near-infrared colorization method based on GANs, which used a triple architecture to learn the three color channels R, G, and B, separately. Furthermore, Suarez et al. (2018) included gaussian noise in each layer of the generator in order to ensure the color diversity of generated images. The same year, an end-to-end near-infrared image colorization method was proposed by Dong et al. (2018), which adopted S-Shape Net(S-Net) composed of ColorNet and EdgeNet. EdgeNet can not only enhance the edge, but also stabilize the color region, and then obtain rich and clear color RGB image. However, these methods often need to rely on a large number of infrared image-visible image pairs, and the absence of paired images limits the application of such colorization methods largely. To solve the problem of insufficient matching data, the CycleGANs-based colorization method is proposed (Nyberg et al., 2019; Sun et al., 2019a). There are still coloring errors and other color artifacts in the colorized images obtained by these methods, so there is still considerable effort needed to explore this research.

### 5.4. Remote sensing colorization

Remote sensing image contains various types of images, such as panchromatic image (PAN), multi-spectral (MS) image, hyperspectral (HS) image, and synthetic aperture radar (SAR) image. PAN images are single-channel and generally have high spatial resolution, but they cannot display the color of the ground object, i.e., the image has little spectral information. Therefore, PAN image colorization is a worthy research topic. Fig. 13 shows some typical colorization model structures of remote sensing images. For example, in Li et al. (2018), a colorization method of grayscale satellite images using multiple discriminators is proposed. Similarly, Ji et al. (2020) used multi-domain periodic consistent GANs (MC-GANs) for SAR image colorization to solve the problem of limited paired data. For the grayscale aerial images, Poterek et al. (2020) proposed a colorization model based on cGANs.

In addition to the remote sensing image colorization methods mentioned above, there are some methods that use image fusion methods

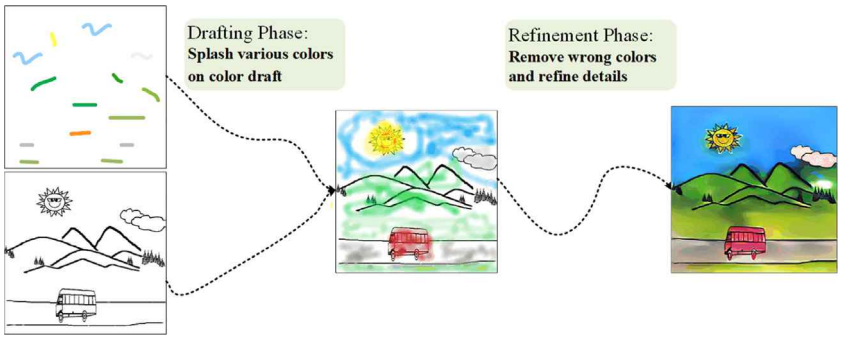



**Fig. 11.** The process of two-stage mange image colorization. The first stage: Drafting Stage, where the coloring does not exactly follow the lines of the sketch, but splashes the colors onto the canvas in a relatively spontaneous way. The second stage: Refinement Stage, which focuses on fixing coloring mistakes and retouching details to get the final colorized image.

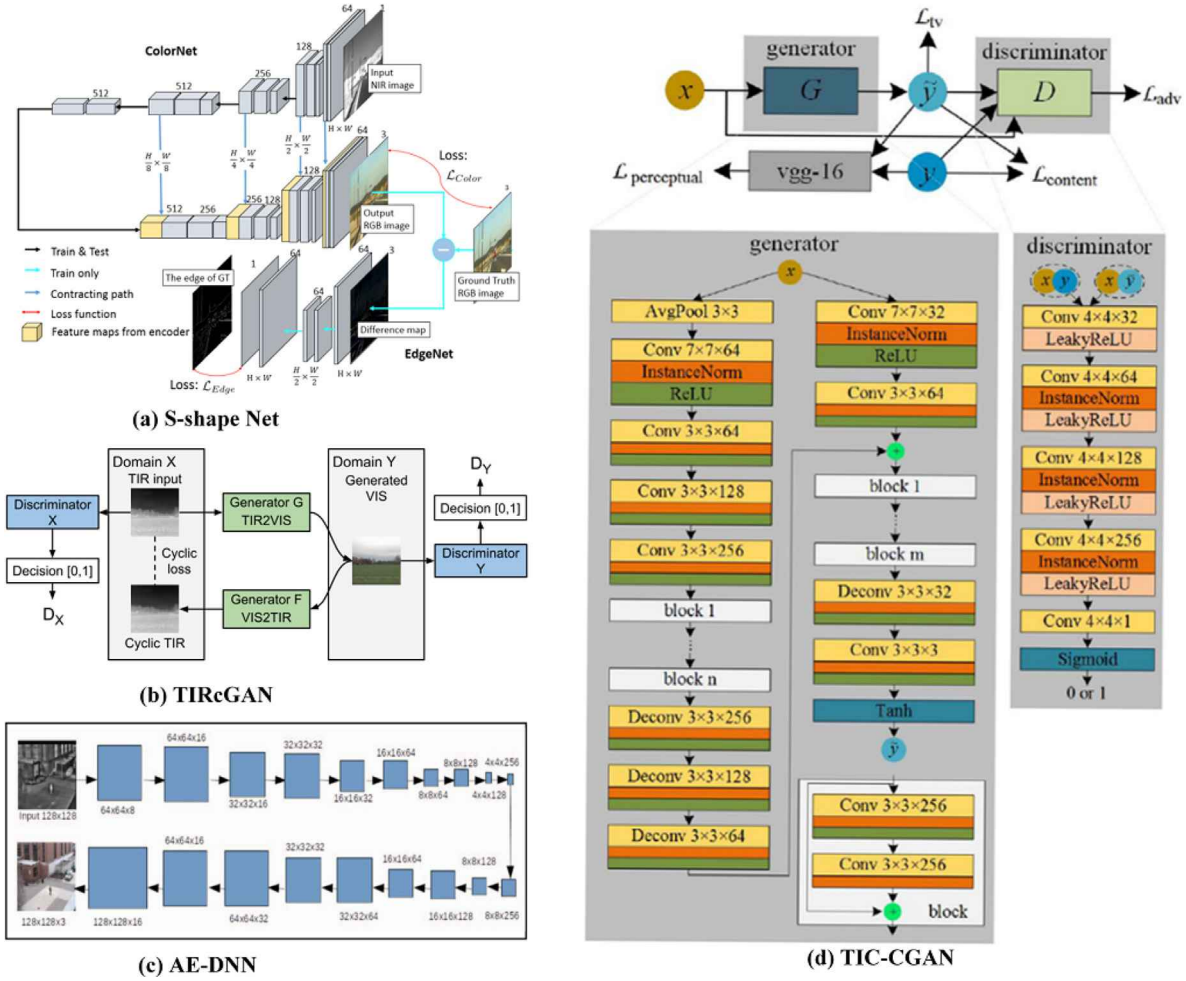



**Fig. 12.** Representative infrared image colorization models. (a) (b) (c) (d) show the network structure of Dong et al. (2018), Nyberg et al. (2019), Qayyum et al. (2018), Kuang et al. (2020) respectively.

to obtain high-quality color remote sensing images, also known as pan-sharpening. Specifically, pan-sharpening is a method of fusing high-spectral resolution and low-spatial resolution MS images with high-spatial resolution and low-spectral resolution PAN images to obtain high-resolution color fusion images. In recent years, many researchers have applied CNNs to pan-sharpening tasks, such as PNN (Masi et al., 2016), PNN+ (Scarpa et al., 2018), and PANNet (Yang et al., 2017). Although they have achieved good results, these methods often require additional supervision, which limits their performance to a certain extent. In addition, this method mainly uses the spectral information of MS images, but neglects the rich spatial information of PAN images, which makes the obtained color fused images lack of detail information. More recently, with the development of GANs, many GANs-based pan-sharpening methods have been proposed, such as (Liu et al., 2018a; Shao et al., 2019; Ma et al., 2020; Jin et al., 2021a). Liu et al. (2018a) and Shao et al. (2019) still need high resolution MS images for supervised learning. Ma et al. (2020) and Jin et al. (2021a) often obtain color fused images with high spatial resolution and high spectral resolution without the supervision of ground truth. In general, remote sensing colorization can provide powerful prior information for ground

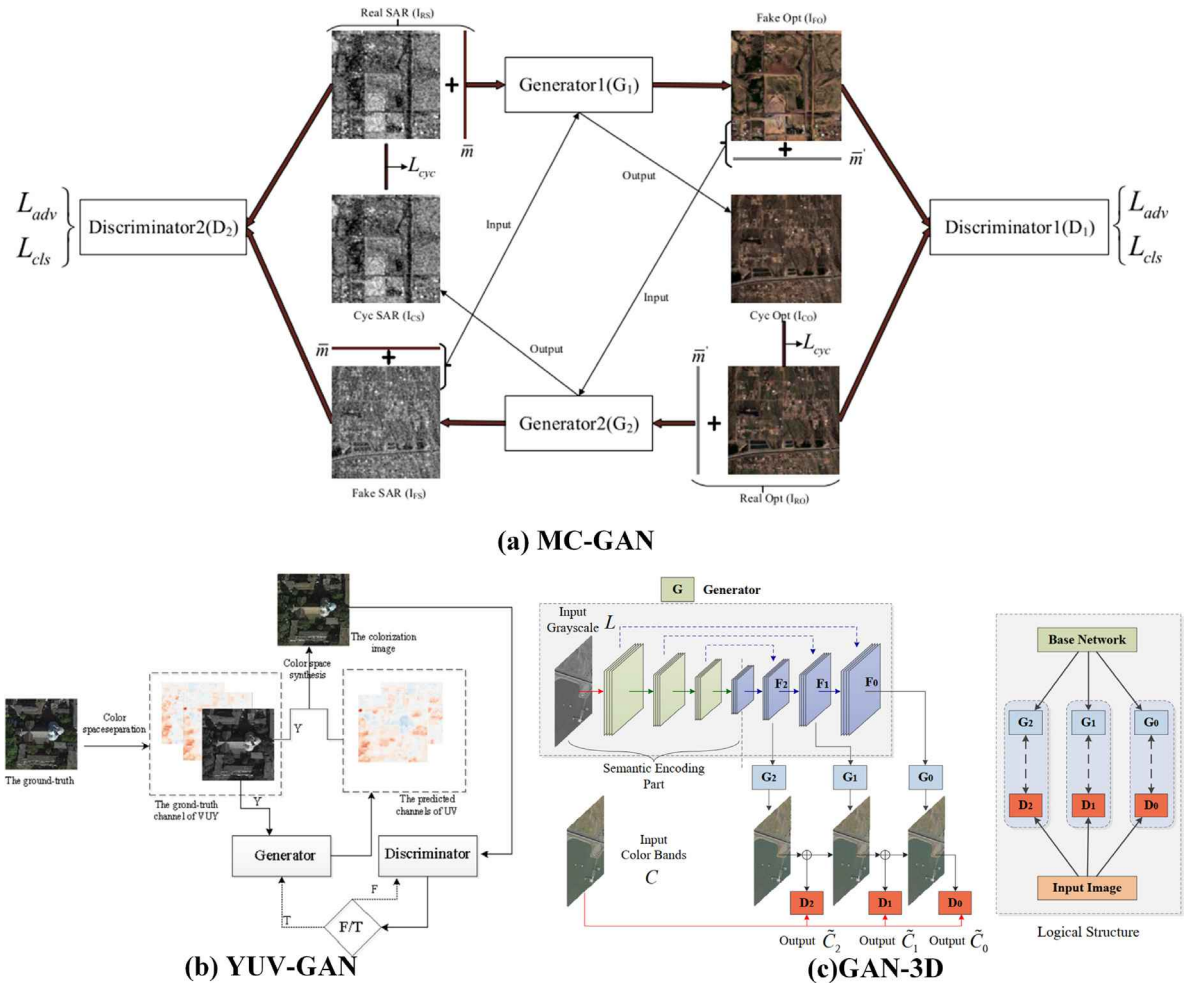



**Fig. 13.** Representative remote sensing image colorization models. (a), (b), and (c) correspond to the network structures proposed in Ji et al. (2020), Wu et al. (2020), Li et al. (2018).

scene classification and target detection, which is worthy of further research.

### 5.5. Video colorization

Video colorization is challenging because of its multimodality and global space–time consistency requirements. First, it is not reasonable to restore a true image in various situations. For example, given a grayscale image of a T-shirt, we often cannot predict the correct color of the T-shirt accurately, as it may be yellow, blue, or some other color. The goal of colorization is to produce a set of colorized results that look natural, not to restore the underlying colors. Second, it does not matter what color is assigned to a region (e.g., T-shirt), but should the entire region be spatially consistent. In addition to the above image colorization challenges, video colorization has other challenges, such as time consistency, cost, and user control. To be specific, video colorization requires that a particular object should be consistent between the previous frame and the current frame of the video or even throughout the video clip. For example, an orange cat named "Meow Meow" should appear orange throughout the video, rather than black one frame before and white the next. Therefore, the image colorization methods cannot be directly extended to video colorization. We show some typical video colorization models in Fig. 14.

In video colorization, the most intuitive method is to run a time filter on frame-by-frame colorization results during post-processing, which relieves flickering but can lead to color degradation and blur (Dogan et al., 2015; Paul et al., 2017; Lai et al., 2018). Another way is to spread color scribble between frames through explicit calculation of optical flow (Dogan et al., 2015; Vondrick et al., 2018; Jampani et al., 2017; Liu et al., 2018b; Meyer et al., 2018) or assume that the first frame is colorized, and then propagate the color frame by frame. This method may cause colorization error accumulation, and the number of propagable frames is limited, so it is only applicable to short videos. For the colorization error accumulation, Liu et al. (2018b) colorized each frame with the help of reference images, which greatly reduced the accumulation of errors. However, the above methods often require human intervention (color scribble or provide reference images), so in Lei and Chen (2019), an automatic video colorization method without label data and user guidance was proposed. Although this method alleviates the problem of manual intervention largely, the colorization effect is not ideal, and there are problems such as unsaturated color and uneven color. In addition, for specific videos to be colorized, such as animation and historical videos, the colorization results are often biased from the user's intention or historical facts. Therefore, video colorization is still a worthy and challenging study.

### 5.6. Other applications

In addition to the applications mentioned above, the image colorization method is also applied to medical images (Liang et al., 2021; Morra et al., 2021; Yu et al., 2020; Guo et al., 2021). For X-ray image colorization, a colorization method based on transfer learning was proposed by Morra et al. (2021). In Yu et al. (2020), Liu et al. proposed a (Positron Emission Tomography) PET data colorization method

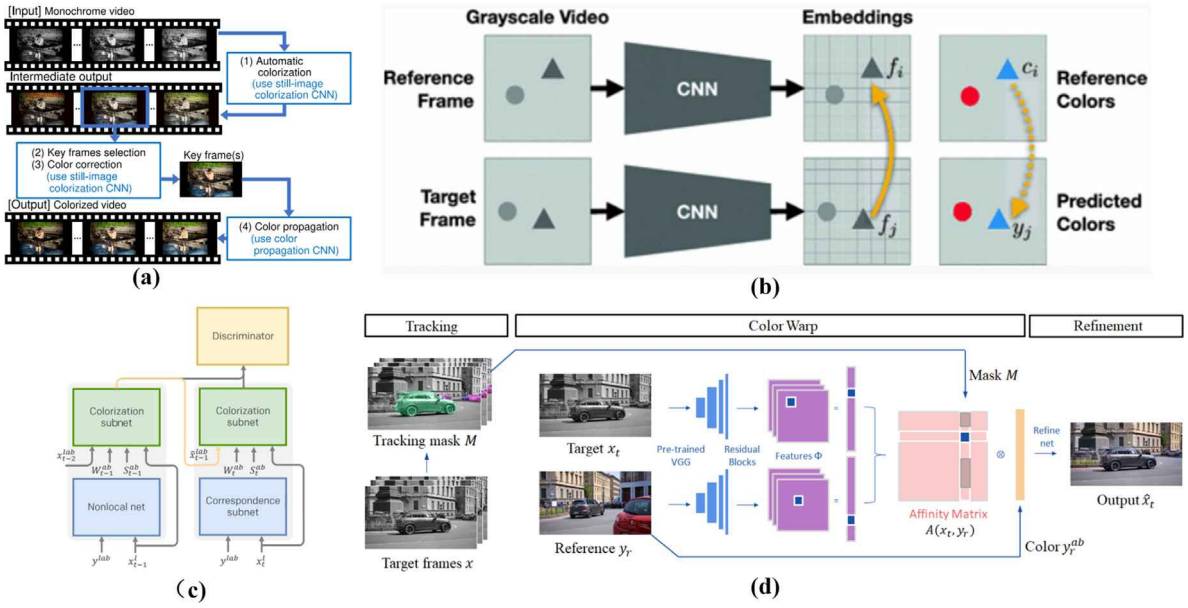



**Fig. 14.** Representative video colorization models. (a), (b), (c) and (d) respectively show the network structure of the colorization model in Endo et al. (2021), Vondrick et al. (2018), Zhang et al. (2019), Akimoto et al. (2020).

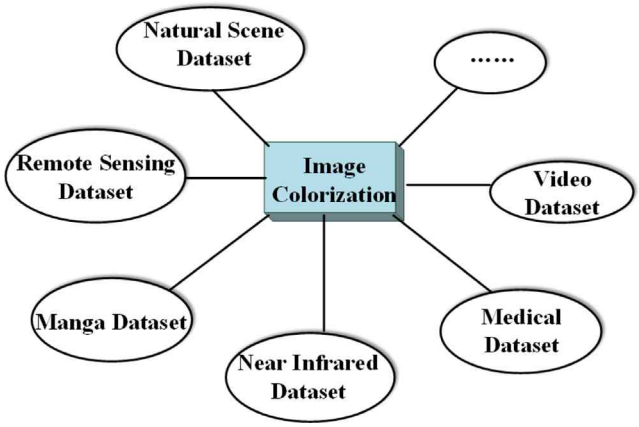



**Fig. 15.** Dataset for different Colorization tasks.

based on dual-threshold scheme, which applies a pair of high and low thresholds to colorize the PET image. In addition, it has applications in 3D point cloud data (Hou et al., 2019), archaeology (Klein et al., 2020), monochromatic microscopic images (Bian et al., 2021), and other fields (Dong et al., 2022; Xuan et al., 2021).

## 6. Datasets and evaluation metrics

In this section, we introduces the datasets commonly used in colorization model training(or testing) and the model performance evaluation system respectively. The types of datasets are sorted from five aspects: natural image, remote sensing image, infrared image, manga image and video, as shown in Fig. 15. In addition, we also present commonly used performance evaluation metrics for colorization models, including user studies and objective image quality evaluation metrics.

### 6.1. Datasets for colorization

Fig. 15 illustrates the common dataset classification for colorization methods. It is well known that different colorization methods use different datasets for model training when applied to different domains. They also differ significantly in terms of the type, number, resolution and diversity of images. Therefore, in order to help researchers to choose the right datasets for different colorization tasks, we sort out the commonly used datasets in terms of data type, data volume, number of scenes or target classes, and applicable research tasks, and also give the sources of different datasets, as shown in Table 3. Furthermore, we show the examples of different datasets in Figs. 16–19, to give the reader a clearer visualization of the actual situation of each dataset.

### 6.2. Evaluation metrics

To evaluate the performance of image colorization methods, researchers proposed several image quality evaluation methods, which can be divided into subjective evaluation and objective evaluation. The main purpose of image quality evaluation is to compare the performance of different colorization methods, which can be used as a guide for selecting colorization methods in practical applications, and also as a loss function. The subjective evaluation method is a popular, reliable and direct method to evaluate the quality of colorized images based on HVS, it plays an important role in image quality evaluation. Generally, trained raters score the colorized images by observing the image details, object integrity, and image distortion. Because of this, the subjective evaluation method has disadvantages such as long time consuming, high cost, and non-reproducibility. Therefore, the objective evaluation method which can quantitatively and automatically evaluate the image quality is proposed to overcome these problems. Compared with the subjective evaluation method, the objective evaluation method is highly consistent with the visual perception of human, and is not easy to be biased by observers. Objective evaluation methods can be further divided into three categories: full reference methods, reduced reference methods based on feature extraction, and no reference methods. In this section, we briefly introduce some representative image quality evaluation methods, and list the usage of different image quality evaluation methods in Table 2.

#### 6.2.1. Subjective evaluation

Mean Opinion Score (MOS) test is a commonly used subjective image quality evaluation method, which requires raters to assign perceived quality scores to the test images. In general, the score

**Table 3**
Description of the datasets and their sources.

| Category | Datasets | Quantity | Object/scenario classes | Instructions | Resource |
|---|---|---|---|---|---|
| Natural scene dataset (Zhu et al., 2017) | ImageNet (Huang et al., 2015) | 14,197,122 | 1000 | image colorization, image processing, image recognition | https://image-net.org/ |
| | SUN (Xiao et al., 2010) | 130,519 | 899 | image colorization,scenario understanding | https://vision.princeton.edu/projects/2010/SUN/ |
| | PASCAL VOC (Everingham et al., 2015) | 11,530 | 20 | image colorization, classification,recognition, object detection | https://host.robots.ox.ac.uk/pascal/VOC/ |
| | LSUN (Yu et al., 2015) | 1,000,000 | 10/20 | image colorization, scenario understanding | https://www.yf.io/p/lsun |
| | Places205 (Zhou et al., 2014) | 25,000,000 | 205 | multi-scene image colorization, scene classification | http://places2.csail.mit.edu/download.html |
| | Place365 (Zhou et al., 2014) | 10,000,000 | 400+ | multi-scene image colorization, scene classification | http://places.csail.mit.edu/downloadData.html |
| | Oxford Flower | 250,000+ | 17/205 | image colorization, classification | https://www.robots.ox.ac.uk/~vgg/data/flowers/ |
| Manga dataset | SketchyScene (Zou et al., 2018) | 40,000+ | 7,000/11,000 | image retrieval, sketch colorization | https://github.com/SketchyScene/SketchyScene#dataset |
| | Yumi's cell | – | – | line art image colorization | https://comic.naver.com/webtoon/list?titleId=651673 |
| | Danboroo (Danbooru-Community, 2018) | 3692578 | – | image colorization, classification | https://www.gwern.net/Danbooru2020 |
| | Manga109 (Matsui et al., 2017) | 109 | 109 | image colorization, object recognition and retrieval tasks | http://www.manga109.org/en/ |
| Near infrared dataset | KAIST Multispectral Pedestrian (Hwang et al., 2015) | 95,000 pairs | – | thermal infrared colorization, Multispectral pedestrian detection | https://github.com/SoonminHwang/rgbt-ped-detection/blob/master/data/README.md |
| | RGB-NIR Scene (Zhao et al., 2011) | 477 pairs | 9 | image colorization, image fusion,Scene Category Recognition. | http://matthewalunbrown.com/nirscene/nirscene.html |
| Remote sensing dataset | NWPU-RESISC45 | 31,500 | 45 | image colorization, image fusion, Remote Sensing Image Scene Classification | http://www.escience.cn/people/JunweiHan/NWPU-RESISC45.html |
| | ISPRS Potsdam6 | 38 | – | image colorization, Semantic segmentation | https://www2.isprs.org/commissions/comm2/wg4/benchmark/2d-sem-label-potsdam/ |
| | Vaihingen7 | 33 | – | image colorization, Semantic segmentation | https://www2.isprs.org/commissions/comm2/wg4/benchmark/2d-sem-label-vaihingen/ |
| | SEN1-2 (Schmitt et al., 2018) | – | 282,384 pairs | image colorization,SAR-Optical Data Fusion | https://mediatum.ub.tum.de/1436631 |
| Video | Videvo Stock | 351,331 | 25 | Video colorization | https://www.videvo.net/ |
| | Hollywood2 | 3669 | 69 | Video colorization | http://www.di.ens.fr/~laptev/actions/hollywood2/ |
| | DAVIS (Perazzi et al., 2016) | 3455 | 50 | Video colorization | https://davischallenge.org/ |
| | YouTube-8M (Abu-El-Haija et al., 2016) | 8,264,650 | 4,800 | Video colorization | https://research.google.com/youtube8m/ |

ranges from 1 (bad color) to 5(good color), the final MOS is the arithmetic mean of all the scores. Image subjective evaluation method can reflect the real intuitive quality of the colorized image, and the evaluation results are reliable, but there are many shortcomings: it cannot be described by the mathematical model, and it is difficult to achieve real-time quality evaluation. In practical application, some image colorization models have poor performance in the commonly used objective quality evaluation metrics, but are far better than other models in terms of perceived quality. In this case, the MOS test is the most reliable image quality evaluation method.

#### 6.2.2. Full-reference methods

The full reference image quality evaluation method is to evaluate the image quality by comparing the evaluated image with the real image. Commonly used quality evaluation metrics include MSE, PSNR, SSIM, MS-SSIM, and so on.

**(a) MSE.** Mean Squared Error (MSE) is used to calculate the pixel error between the colorized image and the source image. The smaller the MSE metric, the better the colorization performance, i.e., the closer the distance between the colorized image and the source image, and the more realistic and natural the colorized image is. The MSE can be

**Fig. 16.** Examples of Natural Scene datasets. Row 1: ImageNet Dataset; Row 2: SUN Dataset; Row 3: PASCAL VOC Dataset; Row 4: LSUN Dataset; Row 5 and 6: Places Dataset; Last two rows: Oxford Flower Dataset.

**Fig. 17.** Examples of Different Manga Datasets. First row: SketchyScene Dataset; Second row: Manga109 Dataset; Third row: Yumi's Cell; Last row: Danboroo Dataset.

calculated by the following equation:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\|T(i,j) - F(i,j)\|^2 \tag{19}$$

where $m$, $n$ represents the size of the real image $T$ and the colorized image $F$. MSE evaluates the pixel-level difference between the colorized image and the real image, but does not care about visual perception, so it is biased in evaluating the quality of the colorized image.

**(b) PSNR.** Peak Signal-to-Noise Ratio (PSNR) reflects the distortion degree of the colorized image. It is the most common and most widely used objective evaluation metric of image quality. The calculation equation is as follows.

$$PSNR = 10 \times \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{20}$$

where $MAX_I^2$ represents the maximum pixel value in the image, the higher the PSNR value is, the better the quality of the colorized image is. Like MSE, they are based on the error between the corresponding pixel points. Because PSNR does not take into account the visual characteristics of the human eyes, the evaluation results are often inconsistent with subjective feelings.

**(c) SSIM.** Since HVS is very sensitive to structural loss and distortion, a general image quality evaluation metric, structural similarity metric (SSIM), is proposed to measure the structural similarity between images, which is mainly composed of structural brightness and contrast. Eqs. (21)–(23) are brightness, contrast, and structural calculation equations, respectively.

$$l(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \tag{21}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \tag{22}$$

$$s(x,y) = \frac{(\sigma_{xy} + c_3)}{(\sigma_x\sigma_y + c_3)} \tag{23}$$

$$c_1 = (k_1 L)^2 \qquad c_2 = (k_2 L)^2 \qquad c_3 = c_2/2 \tag{24}$$

where $\mu_x$, $\mu_y$ are the mean values of images $x$ and $y$. $\sigma_x^2$, $\sigma_y^2$ are the variances of images and respectively is the covariance between images$x$ and $y$. Usually, $L$=255; $k_1$, $k_2$ are constant, $k_1$=0.01, $k_2$=0.03. The SSIM can be obtained by multiplying the Eqs. (21)–(23), which is expressed as follows.

$$\begin{aligned} SSIM &= [l(x,y)]^\alpha \times [c(x,y)]^\beta \times [s(x,y)]^\gamma \\ &= \frac{(2\mu_x\mu_y + c_1)(\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \end{aligned} \tag{25}$$

Here, $\alpha$=$\beta$=$\gamma$=1. The value range of SSIM is [0–1]. The larger the value, the smaller the gap between the predicted image and the real image, which means that the colorized image is more consistent with human visual perception. Therefore, it is widely used in the performance evaluation of the image colorization model.

**(d) MS-SSIM.** Multi-scale SSIM (MS-SSIM) is an extended version of SSIM, which is more flexible in terms of changes in image resolution or angle. Compared with SSIM, MS-SSIM is closer to the subjective quality evaluation method, and its calculation equation is as follows.

$$MS\text{-}SSIM = \left[l_M(x,y)\right]^{\alpha_M} \times \prod_{j=1}^{M}\left[c_j(x,y)\right]^{\beta_j} \times \left[s_j(x,y)\right]^{\gamma_j} \tag{26}$$

The original image scale is 1, the highest scale is $M$, the width and height are reduced by $2M-1$ as the factor, the similarity and contrast

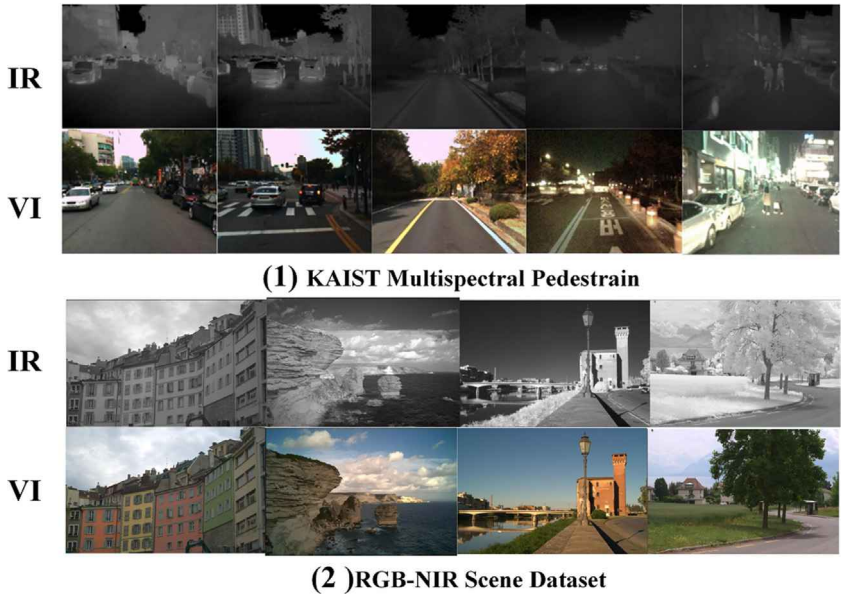



**Fig. 18.** Examples of Different Near Infrared Datasets. (1) KAIST Multispectral Pedestrian Dataset (2) RGB-NIR Scene Dataset.

are compared on $[1-M]$ scales, and only the brightness is compared at the scale $M$. On the $j$th scale, $\alpha_j=\beta_j=\gamma_j$, and the sum of these three parameters in $M$ scales is 1, i.e., $\sum_{j=1}^{M}\alpha_j=\sum_{j=1}^{M}\beta_j=\sum_{j=1}^{M}\gamma_j=1$.

#### 6.2.3. Non-reference methods

Compared with the reference methods, the evaluation method without reference images is strongly flexible. In the actual application of image colorization, it is difficult to obtain the real color image, so the non-reference metrics are more suitable for evaluating the colorized image quality. Next, we will introduce two common no-reference metrics: inception score (IS) and image entropy (EN).

**(a) IS.** Inception score (IS) indicator (Salimans et al., 2016) is calculated through the pre-training network InceptionNet-V3, as the Eq. (27). Firstly, the pre-trained image classification model Inception is applied to generated images to obtain conditional label distribution. Then, IS is obtained by calculating the KL divergence between the conditional label distribution and the edge label distribution. This metric is often used to evaluate the quality and diversity of the generated image, especially the rationality and diversity of the colorized image obtained by the image colorization. However, IS also has some limitations. Such as, the existence of multiple objects in the image may lead to the increase of conditional distribution entropy, which is not related to the quality of the colorized image, but may mislead IS measurement.

$$\mathrm{IS}=\exp\left(\mathrm{E}_{x\sim p_g}\left[KL\left(p(y|x)\parallel p(y)\right)\right]\right) \tag{27}$$

where, image $x$ is sampled from the distribution $p_g$ of generated image, and $KL$ stands for KL divergence. $p(y|x)$ represents the classification vector of Inception network output, and $p(y)$ is the marginal distribution of the generated image on all categories.

**(b) Image Entropy.** Image entropy is a statistical form of feature, which reflects the average amount of information in the image. The one-dimensional entropy of an image represents the information contained in the aggregation feature of the gray distribution in the image, which can be calculated by

$$H=-\sum_{i=0}^{255}p_i\log p_i \tag{28}$$

where $p_i$ is the probability that a certain gray level appears in the image, which can be obtained by the gray level histogram.

#### 6.2.4. Reduced-reference methods

The reduced-reference method only extracts part of the image information as reference, which has a wide range of applications. This method evaluates the image quality by comparing the feature error between the ground truth and colorized image.

**(a) FID.** Fréchet inception distance score (FID) (Heusel et al., 2017) is an improvement on IS, which is also based on InceptionNet-V3. IS directly evaluates the generated images, and the larger the value, the better the image quality. The FID score is obtained by comparing the generated image with the real image. The smaller the metric value is, the better the image quality is. FID is less sensitive to noise and more in line with human evaluation. FID score calculated relies on a pre-trained Inception to extract features of the colorized image and the ground truth, these features are considered as the statistics of the image. This metric is often applied to the colorization performance of GANs-based models, which can calculate by

$$\begin{aligned}FID&=\left\|\mu_{x1}-\mu_{x2}\right\|_2^2\\&\quad+Tr\left(\sum x1+\sum x2-2\left(\sum x1\sum x2\right)^{1/2}\right)\end{aligned} \tag{29}$$

where $\mu_{x1}$ and $\mu_{x2}$ represent the mean values of eigenvectors of generated images $x1$ and real images $x2$ respectively. $\sum x1$ and $\sum x2$ are the covariance matrices of the eigenvectors, and $Tr$ represents the trace of the matrix.

**(b) LPIPS.** Learned perceptual image patch similarity metric (LPIPS) is proposed by Zhang et al. (2018), which compares the perceptual differences between depth features extracted by CNNs. Compared with PSNR, SSIM and other metrics, LPIPS is closer to human perception and can be used to evaluate the diversity of generated sample quality, so this metric is widely used in the colorization task. Given ground truth and colorized image, the perception similarity measurement equation is shown as follows.

$$d\left(x,x_0\right)=\sum_l\frac{1}{H_lW_l}\sum_{h,w}\left\|w_l\odot\left(\hat{Y}^l_{hw}-\hat{Y}^l_{0hw}\right)\right\|_2^2 \tag{30}$$

where $\hat{Y}^l,\hat{Y}^l_0\in\mathbb{R}^{H_l\times W_l\times C_l}$ represent the output of the $l$th layer.

Although the above methods show better performance in capturing human visual perception, what perceptual quality (more natural, or more consistent with ground truth, or the color more reasonable) we need is still a question to be explored. Therefore, evaluation indicators such as PSNR and SSIM are still mainstream at present.

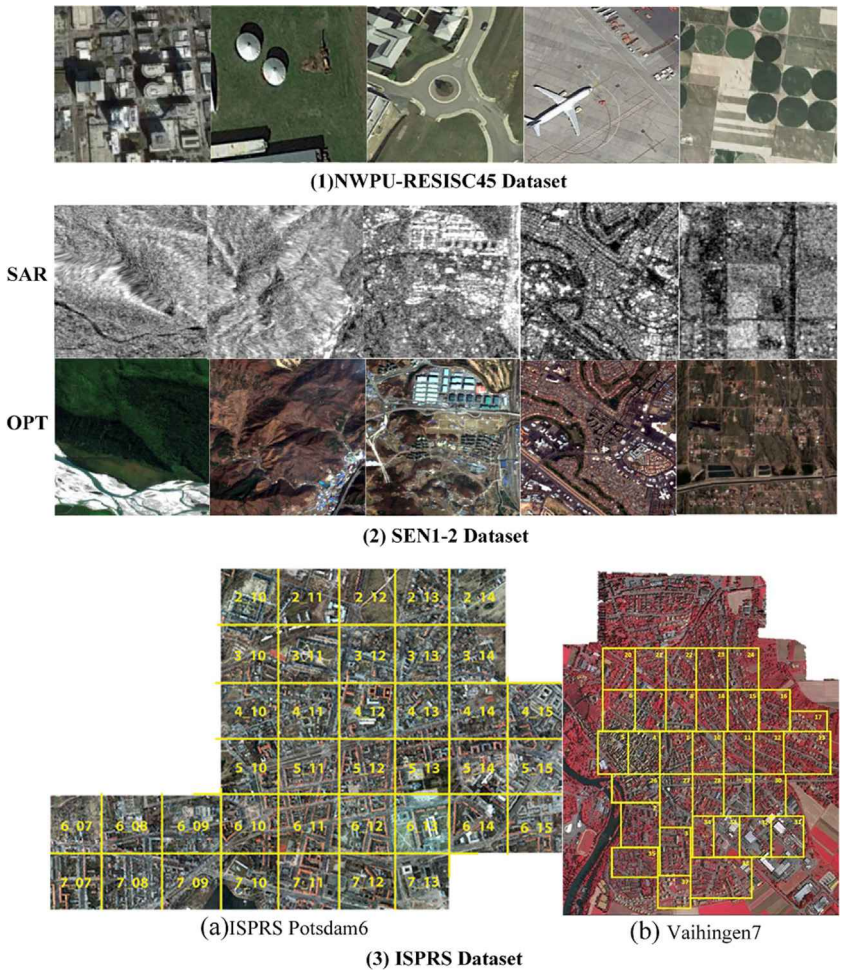



**Fig. 19.** Examples of Different Remote Sensing Datasets. (1) NWPU-RESISC45 Dataset (2) SEN1-2 Dataset (3(a)) ISPRS Potsdam6 (3(b)) Vaihingen7.

*6.2.5. Other methods*

In addition to the commonly used evaluation metrics mentioned above, there are also many other image quality evaluation metrics. For example, Chybicki et al. (2019) adopted an evaluation metric that combined the natural image statistical model, image distortion model, and HVS model, namely visual information fidelity (VIF). This metric has a high consistency with subjective vision. Kuang et al. (2020) used noise quality measure (NQM) to evaluate the performance of the colorization model. Besides, in Varga and Sziranyi (2016), Min et al. (2020), Quaternion SSIM (QSSIM) was used to evaluate the colorized image quality. Because simple FID scores do not provide a fair comparison in the colorization field, thus Lee and Lee (2020) proposed an image quality evaluation metric, namely colorization-FID, which is based on the segmentation suggestion constraint. In general, there is no unified image quality evaluation system in image colorization, which is an urgent problem to be solved.

## 7. Experiments

In this section, the existing advanced image colorization methods are experimentally analyzed according to the application fields, including natural image colorization methods, remote sensing image colorization methods, and infrared image colorization methods. The experimental results are shown in Figs. 20–22. In addition, we also evaluated the performance of the natural image colorization model on different datasets by objective image quality evaluation metrics, as shown in Table 4.

For natural image colorization methods, we conducted comparative experiments of five representative methods on the oxford flower dataset. By observing Fig. 20, we can find that although these methods can obtain good colorization results on some specific grayscale images, most of the existing methods still suffer from color overflow, uneven coloring, and unsaturated tones. For the remote sensing image colorization methods, we compared five advanced image colorization methods on remote sensing images obtained by three different satellites(WorldView-II, QuickBird, GF-2), and most of the methods were able to obtain good colorization results due to the relatively single color of remote sensing images (mostly green, blue, gray, ect.). However, careful observation can still reveals that colored images obtained by some methods suffer from color loss, as shown in the first row of Fig. 21, where all methods perform poorly except for column (d) (i.e., Vitoria et al. (2019)), which enables accurate coloration on the roof. There is also the problem of low color saturation, as shown in the last two rows of Fig. 21 in (e) (f).

Analogously, for the infrared image colorization method, we conducted a series of comparative experiments on the KAIST Multispectral Pedestrian dataset (Hwang et al., 2015), and the results are shown in Fig. 22. The experimental results show that most of the infrared image colorization methods are able to give good color to the images, such as sky, trees, roads, etc., but the retention of detail information is poor, especially for cars, road signs, buildings, etc. In addition, most of the existing infrared image colorization methods require pairs of infrared–visible images, however, it is practically difficult to obtain strictly aligned image pairs, which largely limits the application of such methods. There are some methods that can preserve the original information of infrared images well without the use of strictly aligned visible images, such as Zhu et al. (2017) Li et al. (2021a), as shown

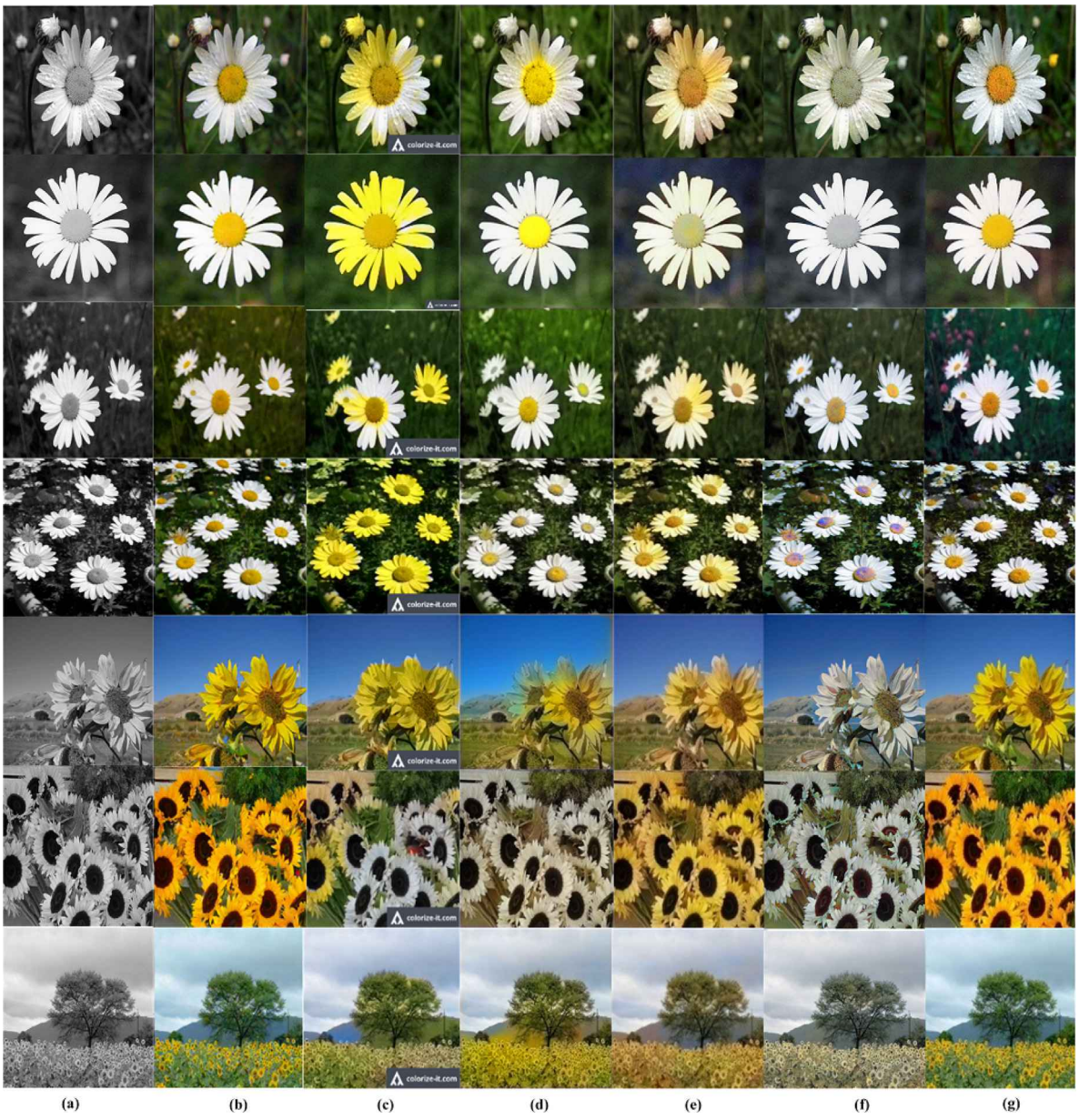



**Fig. 20.** Experimental comparison of different image colorization methods in oxford flower datasets. (a) Gray image. (b) Zhu's (Zhu et al., 2017). (c) Zhang's (Zhang et al., 2016). (d) Larsson's (Larsson et al., 2016). (e) Lizuka's (Iizuka et al., 2016). (f) Antic's (Antic, 2022). (g) Ground Truth.

in Fig. 22(e) (f). However, there is a problem of missing color in areas where there is a large difference in brightness between infrared images and visible images (such as zebra lines). In general, deep learning-based image colorization methods still have a big space for improvement and are worthy of our in-depth study.

## 8. Challenges and discussion

Although DLIC models have made critical progress, they still have limitations in practical application and need further improvement and research.

**(1) Limited Training Data and Open Source Codes.** Many existing deep learning-based colorization methods do not have open source code for further research or use, which greatly limits the research progress in this field. In addition, Numerous methods are trained on the real-world image datasets, and their applications are mostly limited to the colorization of historical photographs. But there are no large-scale public datasets available in many application fields, such as line art images, infrared images, remote sensing images, and medical images. Besides, the performance of the most current models seriously relies on training datasets. In practice, it is difficult or impossible for users to collect enough samples to train a model for the colorization of all natural scenes or objects. Therefore, how to make the DLIC models more practical in the case of limited data is worth further research. This problem may be solved by transfer learning and few shot learning.

**(2) Lack of Color Saturation and Diversity.** Most of the existing colorization models ignore the rare instances in the data, and often select the most frequent color to generalize the data. Although this approach succeeds in producing naturally believable colorized images, each image loses its color characteristics. For example, in anime images, different characters are often distinguished by their hair color, skin color, or pupil color. In addition, for different target objects, it can actually show a variety of colors, which has been neglected by most studies. Especially for man-made scenes/objects, taking cars as an example, although they are of the same brand and series, they may have different colors, such as flamboyant red, calm black, or even bright

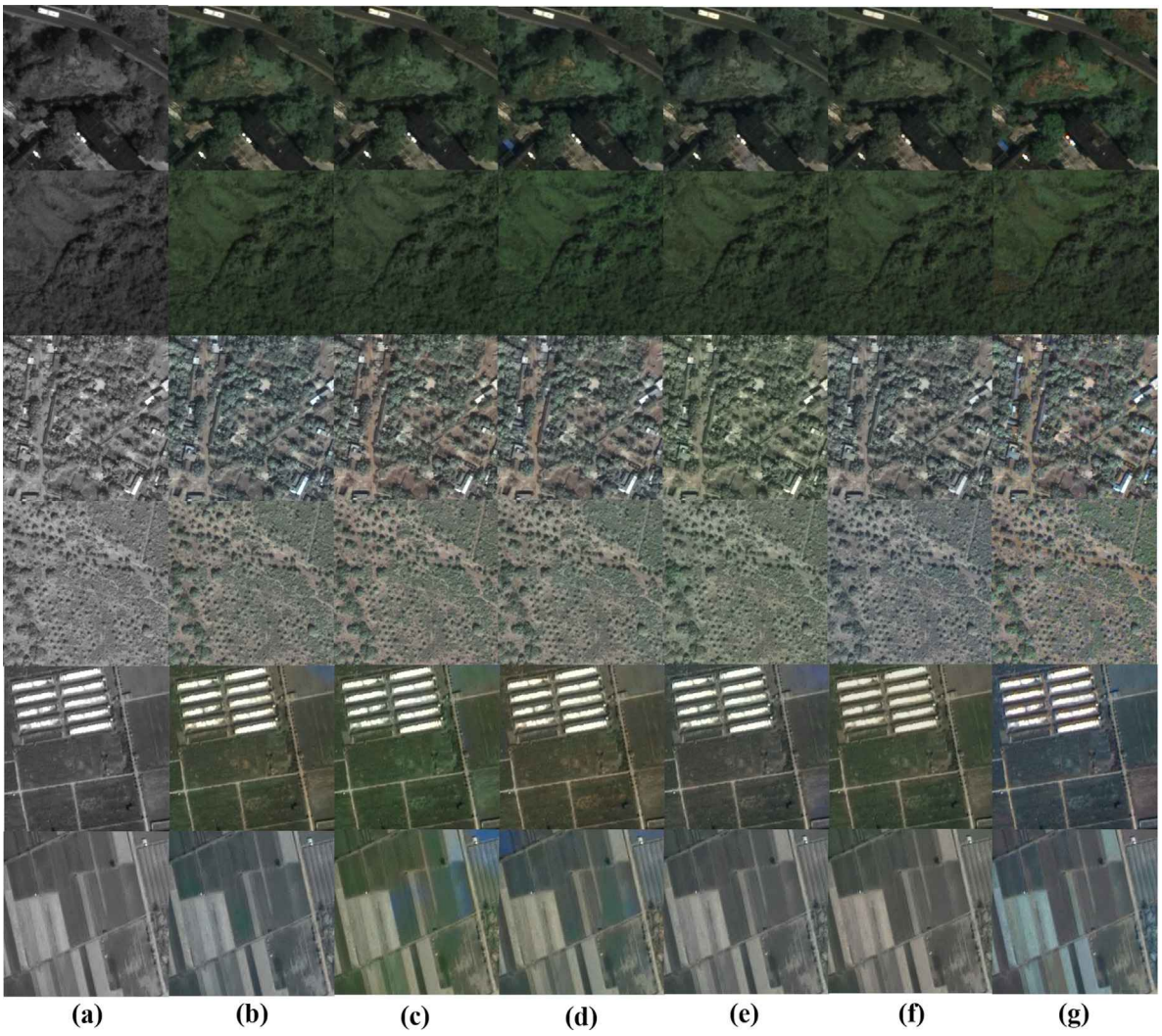



**Fig. 21.** Comparison of different remote sensing image colorization methods. (a) Gray image. (b)Lizuka's (Iizuka et al., 2016). (c) Zhang's (Zhang et al., 2016). (d) Vitoria's (Vitoria et al., 2019). (e) Yoo's (Yoo et al., 2019). (f) Isola's (Isola et al., 2017). (g) Ground Truth. The first and second rows show the experimental results on the dataset obtained from the WorldView-II satellite. The third and fourth rows show the experimental results on the dataset obtained from the QuickBird satellite. The fifth and sixth rows are the experimental results on the dataset obtained from the GF-2 satellite.

yellow. Therefore, how to realize the diversification of the color scheme for the same object is also a major research point in image colorization. For GAN-based methods, perhaps we can add additional noise vectors to the generator to increase the color diversity and saturation of the generated image. In addition, by exploring the potential space of GAN, we can find the direction to control the color of the object, which is also a possible solution to realize the color diversity of the generated image.

**(3) Precise Colorization.** For the application scenes of image colorization, most studies only focus on some simple natural scene images, such as grassland, ocean, buildings or some simple indoor scenes. In fact, deep neural networks with a minimum semantic interpretation can solve it. However, the research of high-resolution complex scene colorization is insufficient with a great challenge. Imagine a busy street with lots of people, different buildings, and even cars. It is obviously difficult to achieve precise colorization for this scene, this is because each color in this complex scene is special and there is no valid ground truth, and any color combination can work if some special harmony is followed. Therefore, we may first carry out scene recognition for complex scenes, and then coloring all kinds of target objects respectively to achieve the precise colorization of the whole scene.

**(4) Reasonable Colorization.** How to realize the reasonable image colorization is also an urgent problem to be solved. In natural scenes, we need to consider the relationship between various objects in a scene. Particularly, we should consider the relationship between light and shadow. The common situation is, when the blue sky, white clouds and surrounding trees are reflected on the calm river, their reflections should be given the same color as their own. However, the existing colorization models often give them average but unreasonable colors only according to the brightness information of the scene. Similar situation is more common in video colorization. As the shots change, the characters or scenes may show different colors due to the lighting changes, but in fact, most researches have not considered this factor. In addition, most studies have focused on the colorization of a single video scene, while very few studies have explored the colorization of a video under different scene transitions, such as walking from indoor to outdoor. Therefore, how to guarantee the quality of the color image and give consideration to the rationality of the image/video color scheme is a problem that needs to be fully explored.

**(5) Insufficient Generalization.** For different scenes, the colorization models in most existing researches often need to be retrained because the color difference between two real application scenarios are great and the colors of the same object in different scenes may be various, which causes the waste of computing resources and time largely. Besides, the requirement of training dataset is also difficult for a specific scenario, and the colorization effect achieved in practical

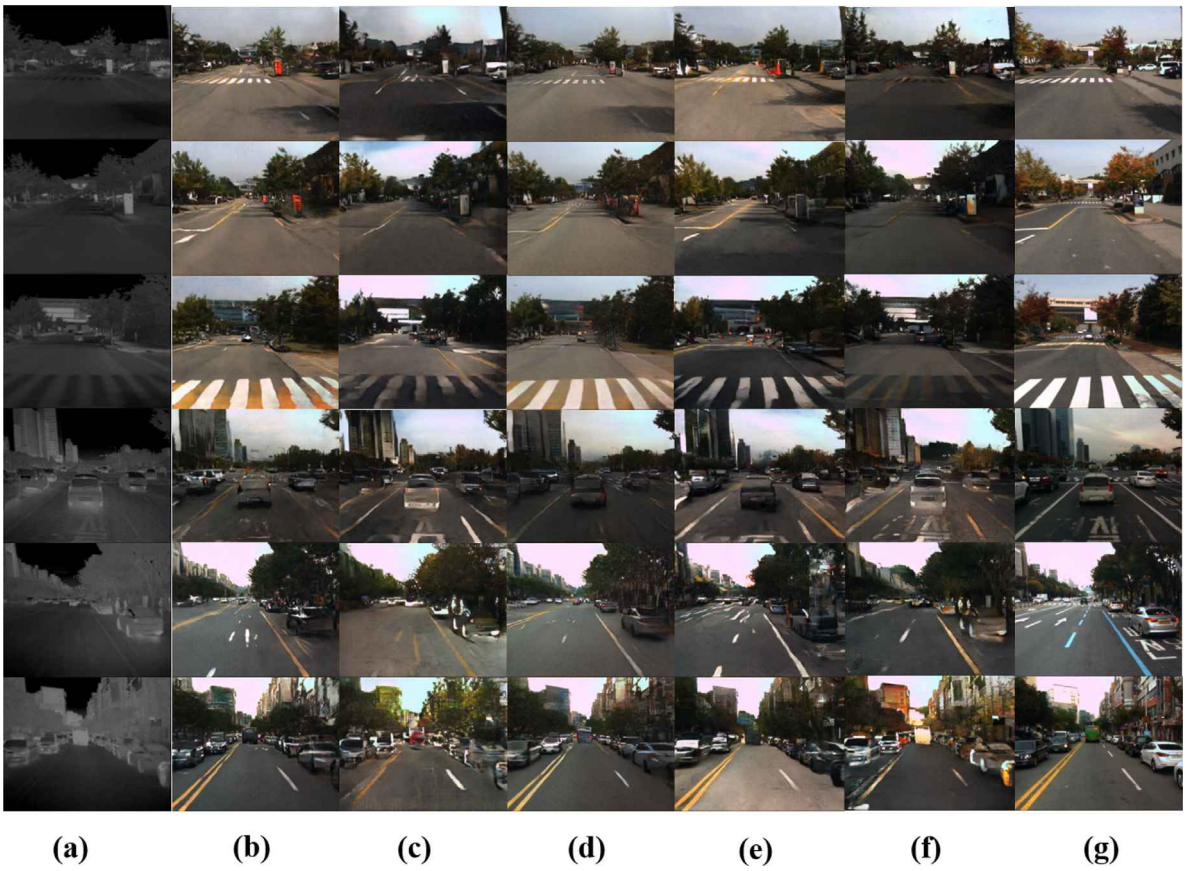



**Fig. 22.** Comparison of different infrared image colorization methods. (a) Infrared image. (b) Isola's (Isola et al., 2017). (c) Zhu's (Zhu et al., 2017). (d) Kuang's (Kuang et al., 2020). (e) Bansal's (Bansal et al., 2018). (f) Li's (Li et al., 2021a). (g) Visible image.

**Table 4**
Experimental comparison of several representative image colorization methods.

| Methods | Publication | Datasets | SSIM↑ | PSNR↑ |
|---|---|---|---|---|
| Zhu's (Zhu et al., 2017) | ICCV/2017 | Oxford Flower | 0.5616 | 20.3554 |
| Isola's (Isola et al., 2017) | CVPR/2017 | Oxford Flower | 0.9631 | 30.3054 |
| Zhang's (Zhang et al., 2016) | ECCV/2016 | Oxford Flower | 0.3699 | 14.0588 |
| | | ImageNet | 0.8921 | 21.7912 |
| | | COCO | 0.8952 | 21.8383 |
| | | Places205 | 0.9214 | 22.5811 |
| Lasson's (Larsson et al., 2016) | ECCV/2016 | Oxford Flower | 0.6913 | 19.741 |
| | | ImageNet | 0.9271 | 25.1074 |
| | | COCO | 0.9302 | 25.0618 |
| | | Places205 | 0.9518 | 25.7224 |
| Lizuka's (Iizuka et al., 2016) | TOG/2016 | Oxford Flower | 0.6903 | 20.722 |
| | | ImageNet | 0.9177 | 23.6365 |
| | | COCO | 0.9225 | 23.8635 |
| | | Places205 | 0.9502 | 25.5817 |
| Antic's (Antic, 2022) | Online/2022 | Oxford Flower | 0.6396 | 28.5172 |
| | | ImageNet | 0.9145 | 23.5374 |
| | | COCO | 0.9207 | 23.6928 |
| | | Places205 | 0.9396 | 23.9836 |
| Su's (Su et al., 2020) | CVPR/2020 | ImageNet | 0.9332 | 26.9805 |
| | | COCO | 0.9401 | 27.777 |
| | | Places205 | 0.9546 | 27.1675 |
| Wu's (Wu et al., 2021b) | Arxiv/2021 | ImageNet | 0.8825 | 21.8162 |
| Kim (Kim et al., 2021) | Arixiv/2021 | ImageNet | 0.9316 | 26.2615 |
| | | COCO | 0.9424 | 26.2337 |
| | | Places205 | 0.9537 | 27.4834 |
| Li's (Li et al., 2021c) | TIP/2021 | Places205 | 0.8925 | 28.9862 |

application is not ideal and needs further study. Thus, how to improve the performance of model generalization is serious for its application. We think that the model generalization should be considered in its designing and training. Besides, the testing technique of model generalization should be explored as well.

**(6) Lack of Uniform Colorization Performance Evaluation System.** For image colorization, the existing image quality evaluation metrics are not targeted, which cannot comprehensively evaluate the color quality of the colorized image. In addition, it is necessary to explore a new evaluation metric for image colorization rationality, which

can be used to evaluate whether the color scheme of colorized image is reasonable or consistent with common sense, so as to avoid ignoring the rationality of image while considering the color diversity of image. Especially for natural scene images, the sky may show different colors at different times of the same day or in different seasons, but it is impossible to be green even though the color of the sky is changeable. We consider that the human perception should be took into account in colorization evaluation system. Therefore, it is necessary to explore a unified and pointed image colorization quality evaluation system to evaluate model performance, comprehensively.

Furthermore, among the existing methods, the automatic method often fails to realize the controllable image colorization, while the semi-automatic method (such as reference-based methods, scribble-based methods) can make up for this defect, but it largely relies on human intervention. Therefore, how to achieve a fully automatic and controllable image colorization, which is still a challenging and interesting research. This problem may be realized by analyzing the latent space or activation space of GANs to find the direction or variable that controls the color of different object categories in the gray image.

## 9. Conclusion

In this paper, we comprehensively review image colorization techniques based on deep learning. Firstly, we describe the problem definition of DLIC, and introduce the commonly used color space and loss function, and on this basis, the DLIC models are classified. Then, we provide new ideas for the classification of DLIC methods from three different perspectives: network structure, degree of automation, and application domain. Next, we cover a contemporary study of popular public datasets and evaluation criteria, and experimentally compare different image colorization methods in three application domains, as well as fully analyze the performance of different colorization methods using a comprehensive image quality evaluation system. Finally, we discuss several open issues and challenges of colorization in the deep learning era, and identify some potentially productive directions forward.

All in all, image colorization methods have made remarkable progress in deep learning era. To achieve more effective model designing, training, and reasoning, there are still many problems to be explored for academic researches and practical applications. We hope that this review will provide an effective way to understand the current state of art and provide insights into the future of DLIC.

## CRediT authorship contribution statement

**Shanshan Huang:** Conceptualization, Methodology, Investigation, Writing – original draft. **Xin Jin:** Conceptualization, Supervision, Project administration. **Qian Jiang:** Funding acquisition, Writing – reviewing and editing. **Li Liu:** Supervision, Project administration, Funding acquisition.

## Declaration of competing 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.

## Acknowledgments

This study is supported by the National Major Science and Technology Projects of China (No. 2018AAA0100703), National Natural Science Foundation of China (Nos. 62101481, 62002313, 61977012), Yunnan Fundamental Research Projects, China (Nos. 202201AU070033, 202201AT070112), Key Laboratory in Software Engineering of Yunnan Province, China (No. 2020SE408), the Central Universities Project in China for the Digital Cinema Art Theory and Technology Lab at Chongqing University (Nos. 2017CDJSK06PT10, 2020CDJSK06PT14), and the funding of Chongqing Institute of Cultural Relics and Archaeology, China (No. 00001).

## References

Abu-El-Haija, S., Kothari, N., Lee, J., et al., 2016. Youtube-8M: A large-scale video classification benchmark. arXiv:1609.08675.

Aizawa, M., Sei, Y., Tahara, Y., 2019. Do you like sclera? Sclera-region detection and colorization for anime character line drawings. Int. J. Netw. Distrib. Comput. 7 (3).

Akimoto, N., Hayakawa, A., Shin, A., et al., 2020. Reference-based video colorization with spatiotemporal correspondence. https://arxiv.org/abs/2011.12528.

An, J., Kpeyiton, K.G., Shi, Q., 2020. Grayscale images colorization with convolutional neural networks. Soft Comput. 24 (3).

Antic, Jason, 2022. Jantic/deoldify: A deep learning based project for colorizing and restoring old images (and video!). https://github.com/jantic/DeOldify.Online; (Accessed: 3 February 2022).

Anwar, Saeed., Tahir, Muhammad., Li, Chongyi., et al., 2022. Image colorization: A survey and dataset. Arxiv.

Arjovsky, M., Chintala, S., Bottou, L., 2017. Wasserstein gan. arXiv preprint arXiv:1701.07875.

Bahng, H., Yoo, S., Cho, W., et al., 2018. Coloring with words: Guiding image colorization through text-based palette generation. In: ECCV.

Bansal, A., Ma, S., Ramanan, D., Sheikh, Y., 2018. Recycle-GAN: Unsupervised video retargeting. In: ECCV.

Berg, A., Ahlberg, J., Felsberg, M., 2018. Generating visible spectrum images from thermal infrared. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 1143–1152.

Bian, Y., Jiang, Y., Huang, Y., et al., 2021. Deep learning virtual colorization overcoming chromatic aberrations in singlet lens microscopy. APL Photon..

Cao, R.Z., Mo, H.R., Gao, C.Y., 2021. Line art colorization based on explicit region segmentation. Comput. Graph. Forum. 40 (7), 1–10.

Cao, Y., Zhou, Z., Zhang, W., et al., 2017. Unsupervised diverse colorization via generative adversarial networks. In: Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2017, In: Lecture Notes in Computer Science, vol. 10534.

Carion, N., Massa, F., Synnaeve, G., et al., 2020. End to End Object Detection with Transformer. Paris Dauphine University, Facebook AI.

Casey, Evan, Pérez, Víctor, Li, Zhuoru, et al., 2021. The animation transformer: Visual correspondence via segment matching. Arxiv.

Chakraborty, S., 2019. Image colourisation using deep feature-guided image retrieval. IET Image Process. 13 (7), 1130–1137, (2019).

Chen, W., Hays, J., 2018. SketchyGAN: Towards diverse and realistic sketch to image synthesis. In: CVPR.

Chen, Q., Koltun, V., 2017. Photographic image synthesis with cascaded refinement networks. In: ICCV, Vol. 1.

Chen, D., Liao, J., Yuan, L., et al., 2017. Coherent online video style transfer. In: ICCV.

Chen, J., Shen, Y., Gao, J., et al., 2018a. Language-based image editing with recurrent attentive models. In: CVPR. pp. 8721–8729.

Chen, Z., Wu, R., Lin, Y., Li, C., Chen, S., Yuan, Z., Chen, S., Zou, X., 2022. Plant disease recognition model based on improved YOLOv5. Agronomy 12, 365.

Chen, C., Xu, Y., Yang, X., 2019. User tailored colorization using automatic scribbles and hierarchical features. Digit. Signal Process. 87, 155–165.

Chen, D., Yuan, L., Liao, J., et al., 2018b. Stereoscopic neural style transfer. In: CVPR.

Chen, S., Zhang, J., Gao, L., et al., 2020. Active colorization for cartoon line drawings. IEEE Trans. Visual. Comput. Graph..

Cheng, Z., Yang, Q., Sheng, B., 2015. Deep colorization. In: ICCV. pp. 415–423.

Cheng, Z., Yang, Q., Sheng, B., 2017. Colorization using neural network ensemble. In: IEEE Trans. Image Process.. 11 (26), 5491–5505.

Chybicki, M., Kozakiewicz, W., Sielski, D., et al., 2019. Deep cartoon colorizer: An automatic approach for colorization of vintage cartoons. Eng. Appl. Artif. Intell. 37–46.

Ci, Y., Ma, X., Wang, Z., et al., 2018. User-guided deep anime line art colorization with conditional adversarial networks. In: Proceedings of ACM Multimedia.

Cordonnier, J.B., Mahendran, A., Dosovitskiy, A., et al., 2021. Differentiable patch selection for image recognition. In: CVPR.

Dabas, C., Jain, S., Bansal, A., et al., 2020. Implementation of image colorization with convolutional neural network. Int. J. Syst. Assur. Eng. Manag. 11 (1).

Dai, Z., Cai, B., Lin, Y., et al., 2021. UP-DETR: Unsupervised pre-training for object detection with transformers. In: CVPR.

DanbooruCommunity, 2018. Danbooru2017: A large-scale crowdsourced and tagged anime illustration dataset.

Deng, X., Zhang, Y., Xu, M., et al., 2021. Deep coupled feedback network for joint exposure fusion and image super-resolution. IEEE Trans. Image Process. (99), 1.

Deshpande, A., Lu, J., Yeh, M., et al., 2017. Learning diverse image colorization. In: CVPR. pp. 2877–2885.

Dias, M., Monteiro, J., Estima, J., et al., 2020. Semantic segmentation and colorization of grayscale aerial imagery with W-net models. Expert Syst..

Dogan, P., Aydın, T.O., Stefanoski, N., et al., 2015. Key- frame based spatiotemporal scribble propagation. In: Proceedings of the Eurographics Workshop on Intelligent Cinematography and Editing. pp. 13–20.

Doi, K., Sakurada, K., Onishi, M., et al., 2020. GAN-based SAR-to-optical image translation with region information. In: 2020 IEEE International Geoscience and Remote Sensing Symposium. pp. 2069–2072.

Dong, Z., Kamata, S., Breckon, T.P., 2018. Infrared image colorization using S-shape network. In: 25th IEEE International Conference on Image Processing. ICIP, pp. 2242–2246.
Dong, X., Li, W., Hu, X., et al., 2022. A colorization framework for monochrome-color dual-lens systems using a deep convolutional network. In: IEEE Transactions on Visualization and Computer Graphics.
Du, K., Liu, C., Cao, L., et al., 2021. Double-channel guided generative adversarial network for image colorization. IEEE Access PP (99), 1, (2021).
Endo, R., Kawai, Y., Mchizuki, T., 2021. A practical monochrome video colorization framework for broadcast program production. IEEE Trans. Broadcast. 67 (1), 1–13.
Everingham, M., Eslami, S., Gool, L.Van., et al., 2015. The pascal visual object classes challenge: A retrospective. Int. J. Comput. Vis. 111 (1), 98–136.
Furusawa, C., Hiroshiba, K., Ogaki, K., et al., 2017. Comicolorization: semi-automatic manga colorization. In: SIGGRAPH Asia 2017 Technical Briefs. pp. 1–4.
Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al., 2014. Generative adversarial networks. In: Advances in Neural Information Processing Systems, Vol. 3. pp. 2672–2680.
Gravey, M., Rasera, L.G., Mariethoz, G., 2019. Analogue-based colorization of remote sensing images using textural information. ISPRS J. Photogramm. Remote Sens. 147 (JAN), 242–254, (2019).
Guo, J., Chen, J., Lu, C., Huang, H., 2021. Medical image enhancement for lesion detection based on class-aware attention and deep colorization. In: 2021 IEEE 18th International Symposium on Biomedical Imaging. ISBI, pp. 1746–1750.
He, M., Chen, D., Liao, D., et al., 2018. Deep exemplar-based colorization. ACM Trans. Graph. 37, 47.1–47.16.
Hensman, P., Aizawa, K., 2017. cGAN-based manga colorization using a single training image. In: 2017 14th IAPR International Conference on Document Analysis and Recognition. ICDAR, pp. 72–77.
Heusel, M., Ramsauer, H., Unterthiner, T., et al., 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: Advances in Neural Information Processing Systems. pp. 6626–6637.
Hou, J., Zhao, B., Ansari, N., et al., 2019. Range image based point cloud colorization using conditional generative model. In: 2019 IEEE International Conference on Image Processing. ICIP.
Huang, S., Jin, X., Jiang, Q., et al., 2021. A fully-automatic image colorization scheme using improved CycleGAN with skip connections. Multimedia Tools Appl. 1 (2021), 1–28.
Huang, Z., Karpathy, A., Khosla, A., et al., 2015. Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. 3 (2015), 211–252.
Hwang, S., Park, J., Kim, N., et al., 2015. Multispectral pedestrian detection: Benchmark dataset and baseline. In: CVPR. pp. 1037–1045.
Iizuka, S., Simo-Serra, E., 2019. DeepRemaster: Temporal source reference attention networks for comprehensive video enhancement. ACM Trans. Graph. 38 (6).
Iizuka, S., Simo-Serra, E., Ishikawa, H., 2016. 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), 110.1–110.11.
Ishaan, G., Faruk, A., Martin, A., et al., 2017. Improved training of Wasserstein GANs. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17, pp. 5769–5779.
Isola, P., Zhu, J., Zhou, T., et al., 2017. Image-to-image translation with conditional adversarial networks. In: CVPR.
Jampani, V., Gadde, R., Gehler, P.V., 2017. Video propagation networks. In: CVPR. pp. 3154–3164.
Ji, G., Wang, Z., Zhou, L., et al., 2020. SAR image colorization using multidomain cycle-consistency generative adversarial network. IEEE Geosci. Remote Sens. Lett. PP (99), 1–5, (2020).
Jin, X., Huang, S., Jiang, Q., et al., 2021a. Semi-supervised remote sensing image fusion using multi-scale conditional generative adversarial network with siamese structure. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. (99), 1.
Jin, X., Li, Z., Liu, K., et al., 2021b. Focusing on persons: Colorizing old images learning from modern historical movies. Arxiv.
Jin, Z., Liu, B., Chu, Q., et al., 2021c. ISNet: Integrate image-level and semantic-level context for semantic segmentation. In: ICCV.
Johari, M., Behroozi, H., 2020a. Context-aware colorization of gray-scale images utilizing a cycle-consistent generative adversarial network architecture. Neurocomputing 407, 94–104.
Johari, M., Behroozi, H., 2020b. Gray-scale image colorization using cycle-consistent generative adversarial networks with residual structure enhancer. In: 2020 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP.
Justin, J., Alexandre, A., Li, F., 2016. Perceptual losses for real-time style transfer and super-resolution. In: ECCV. Springer, pp. 694–711.
Karras, T., Aittala, M., Laine, S., et al., 2021. Alias-free generative adversarial networks. In: NeurIPS.
Khanolkar, P.M., Mccomb, C.C., Basu, S., 2021. Predicting elastic strain fields in defective microstructures using image colorization algorithms. Comput. Mater. Sci. 186, 110068.
Kiani, L., Saeed, M., Nezamabadi-Pour, H., 2020. Image colorization using a deep transfer learning. In: 2020 8th Iranian Joint Congress on Fuzzy and Intelligent Systems. CFIS.
Kim, H., Jhoo, H., Park, E., et al., 2019. Tag2Pix: Line art colorization using text tag with secat and changing loss. In: ICCV.
Kim, E., Lee, S., Park, J., et al., 2021. Deep edge-aware interactive colorization against color-bleeding effects. Arxiv.
Klein, V., Kurth, P., Keinert, et al., 2020. Proxy painting: Digital colorization of real-world objects. J. Comput. Cult Herit. 13 (3), 1–20.
Kong, G., Tian, H., Duan, X., et al., 2021. Adversarial edge-aware image colorization with semantic segmentation. IEEE Access PP (99), 1, (2021).
Kuang, X., Zhu, J., Sui, X., et al., 2020. Thermal infrared colorization via conditional generative adversarial network. Infrared Phys. Technol. 107.
Lai, W.-S., Huang, J.-B., Wang, O., et al., 2018. Learning blind video temporal consistency. In: ECCV.
Larsson, G., Maire, M., Shakhnarovich, G., 2016. Learning representations for automatic colorization. In: ECCV.
Larsson, G., Maire, M., Shakhnarovich, G., 2017a. Colorization as a proxy task for visual understanding. In: CVPR. pp. 840–849.
Larsson, G., Maire, M., Shakhnarovich, G., 2017b. Colorization as a proxy task for visual understanding. IEEE Comput. Soc..
Lee, Y., Cho, S., 2020. Design of semantic-based colorization of graphical user interface through conditional generative adversarial nets. Int. J. Hum.-Comput. Interact. 36 (8), 699–708.
Lee, J., Kim, E., Lee, Y., et al., 2020. Reference-based sketch image colorization using augmented-self reference and dense semantic correspondence. In: CVPR.
Lee, Y., Lee, S., 2020. Automatic colorization of anime style illustrations using a two-stage generator. Appl. Sci. 10 (23), 8699, (2020).
Lei, C., Chen, Q., 2019. Fully automatic video colorization with self-regularization and diversity. In: CVPR.
Li, M., Gou, Y., Gong, B., et al., 2020. Gan based AI drawing board for image generation and colorization. In: Special Interest Group on Computer Graphics and Interactive Techniques Conference Posters. SIGGRAPH '20 Posters.
Li, S., Han, B., Yu, Z., et al., 2021a. I2V-GAN: Unpaired infrared-to-visible video translation. In: Proceedings of the 29th ACM International Conference on Multimedia. MM'21.
Li, X., Li, H., Wang, C., et al., 2021b. Visual-attention GAN for interior sketch colourisation. Image Process., IET 15 (4), 997–1007.
Li, F., Ma, L., Cai, J., 2018. Multi-discriminator generative adversarial network for high resolution gray-scale satellite image colorization. In: 2018 IEEE International Geoscience and Remote Sensing Symposium. pp. 3489–3492.
Li, H., Sheng, B., Li, P., et al., 2021c. Globally and locally semantic colorization via exemplar-based broad-GAN. IEEE Trans. Image Process. 30, 8526–8539.
Liang, Y., Lee, D., Li, Y., et al., 2021. Unpaired medical image colorization using generative adversarial network. Multimedia Tools Appl. (2021), 1–15.
Liang, X., Su, Z., Xiao, Y., et al., 2016. Deep patch-wise colorization model for grayscale images. In: SIGGRAPH ASIA 2016 Technical Briefs. 13, pp. 1–4.
Limmer, M., Lensch, H.P.A., 2016. Infrared colorization using deep convolutional neural networks. In: 2016 15th IEEE International Conference on Machine Learning and Applications. ICMLA, pp. 61–68.
Liu, Y., Qin, Z., Wan, T., et al., 2017. Auto-painter: Cartoon image generation from sketch by using conditional generative adversarial networks. Neurocomputing S0925231218306209, 2017.
Liu, X., Wang, Y., Liu, Q., 2018a. PSGAN: a generative adversarial network for remote sensing image pan-sharpening. In: Proceedings of the IEEE International Conference on Image Processing. pp. 873–877.
Liu, S., Zhong, G., Mello, S., et al., 2018b. Switchable temporal propagation network. In: ECCV. pp. 89–104.
Ma, J., Yu, W., Chen, C., et al., 2020. Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion. Inf. Fusion 62, 110–120.
Maejima, A., Kubo, H., Funatomi, T., et al., 2019. Graph matching based anime colorization with multiple references. In: ACM SIGGRAPH, no. 13. pp. 1–2.
Manjunatha, V., Iyyer, M., Boyd-Graber, J., et al., 2018. Learning to color from language conference of the North American chapter of the association for computational linguistics: Human language technologies.
Manoj, K., Dirk, W., Nal, K., 2021. Colorization transformer. In: ICLR.
Mao, X., Li, Q., Xie, H., et al., 2017. Least squares generative adversarial networks. In: ICCV.
Masi, G., Cozzolino, D., Verdoliva, L., Scarpa, G., 2016. Pansharpening by convolutional neural networks. Remote Sens. 7 (8), 594.
Mathur, A.N., Khattar, A., Sharma, O., 2021. 2D to 3D medical image colorization. In: 2021 IEEE Winter Conference on Applications of Computer Vision. WACV, pp. 2846–2855.
Matsui, Y., Ito, K., Aramaki, Y., et al., 2017. Sketch-based manga retrieval using manga109 dataset. Multimedia Tools Appl. 20, 21811–21838.
Meyer, S., Cornillere, V., Djelouah, A., et al., 2018. Gross. Deep video color propagation. In: BMVC.
M.H. Baig, L., 2017b. Multiple hypothesis colorization and its application to image compression. Torresani. Comput. Vis. Image Underst. 164, 111–123, (2017).
Min, L., Li, Z., Jin, Z., et al., 2020. Color edge preserving image colorization with a coupled natural vectorial total variation. Comput. Vis. Image Underst..
Morra, L., Piano, L., Lamberti, F., et al., 2021. Bridging the gap between natural and medical images through deep colorization. In: 2020 25th International Conference on Pattern Recognition. ICPR, pp. 835–842.

Mourchid, Y., Donias, M., Berthoumieu, Y., 2021. Automatic image colorization based on multi-discriminators generative adversarial networks. In: 2020 28th European Signal Processing Conference. EUSIPCO.
Nyberg, A., Eldesokey, A., Bergstrm, D., et al., 2019. Unpaired thermal to visible spectrum transfer using adversarial training. In: ECCV Workshop.
Ozcelik, F., Alganci, U., Sertel, E., et al., 2020. Rethinking CNN-based pansharpening: Guided colorization of panchromatic images via GANs. IEEE Trans. Geosci. Remote Sens. (99), 1–16.
Paul, S., Bhattacharya, S., Gupta, S., 2017. Spatiotemporal colorization of video using 3d steerable pyramids. IEEE Trans. Circuits Syst. Video Technol. 27 (8), 1605–1619.
Perazzi, F., Pont-Tuset, J., McWilliams, L., et al., 2016. A benchmark dataset and evaluation methodology for video object segmentation. In: CVPR.
Poterek, Q., Herrault, P.-A., Skupinski, G., et al., 2020. Deep learning for automatic colorization of legacy grayscale aerial photographs. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. (13), 2899–2915.
Qayyum, U., Ahsan, Q., Mahmood, Z., et al., 2018. Thermal colorization using deep neural network. In: International Bhurban Conference on Applied Sciences & Technology. pp. 325–329.
Ramassamy, S., Kubo, H., Taku, 2019. Robust image colorization using self attention based progressive generative adversarial network. In: CVPRW.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. Springer International Publishing.
Rudin, L.I., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. Physica D 60.
Sajjadi, M., Scholkopf, B., Hirsch, M., 2017. Enhancenet: Single image super-resolution through automated texture synthesis. In: CVPR. pp. 4491–4500.
Salimans, T., Goodfellow, I., Zaremba, W., et al., 2016. Improved techniques for training gans. In: Advances in Neural Information Processing Systems. pp. 2234–2242.
Scarpa, G., Vitale, S., Cozzolino, D., 2018. Target-adaptive CNN-based pansharpening. IEEE Trans. Geosci. Remote Sens. 9 (56), 5443–5457.
Schmitt, M., Hughes, L.H., Zhu, X.X., 2018. The SEN1-2 dataset for deep learning in SAR-optical data fusion. arXiv:1807.01569.
Seo, C., Seo, Y., 2021. Seg2pix: Few shot training line art colorization with segmented image data. Appl. Sci. 11 (1464).
Shao, Z., Lu, Z., Ran, M., Fang, L., Zhou, J., Zhang, Y., 2019. Residual encoder–decoder conditional generative adversarial network for pansharpening. IEEE Geosci. Remote Sens. Lett..
Shi, M., Zhang, J., Chen, S., et al., 2020. Deep line art video colorization with a few references. https://arxiv.org/abs/2003.10685v2.
Silva, F., Castro, P., Marujo, E., et al., 2019. Mangan: Assisting colorization of manga characters concept art using conditional GAN. In: 2019 IEEE International Conference on Image Processing. ICIP.
Song, Q., Xu, F., Jin, Y., 2017. Radar image colorization: Converting single-polarization to fully polarimetric using deep neural networks. IEEE Access.
Su, J., Chu, H., Huang, J., 2020. Instance-aware image colorization. In: CVPR. pp. 7965–7974.
Su, A., Liang, X., Guo, J., et al., 2018. An edge-refined vectorized deep colorization model for grayscale-to-color images. Neurocomputing 311 (2018), 305–315.
Suarez, P., Sappa, A., Vintimilla, B., et al., 2018. Near InfraRed imagery colorization. In: 25th IEEE International Conference on Image Processing. ICIP.
Suarez, P.L., Sappa, A.D., 2017. Infrared image colorization based on a triplet DCGAN architecture. In: CVPRW.
Sun, T., Jung, C., Fu, Q., et al., 2019a. NIR to RGB domain translation using asymmetric cycle generative adversarial networks. In: IEEE Access. 7, 112459-112469.
Sun, T., Lai, C., Wong, S., et al., 2019b. Adversarial colorization of icons based on contour and color conditions. In: Proceedings of the 27th ACM International Conference on Multimedia. MM '19, pp. 683–691.
Tang, Y., Zhu, M., Chen, Z., Wu, C., Chen, B., Li, C., Li, L., 2022. Seismic performance evaluation of recycled aggregate concrete-filled steel tubular columns with field strain detected via a novel mark-free vision method. Structures 37, 426–441.
Teng, X., Li, Z., Liu, Q., et al., 2021. Subjective evaluation of colourized images with different colorization models. Color Res Appl. 46, 319–331.
Thasarathan, H., Ebrahimi, M., 2019. Artist-guided semiautomatic animation colorization. In: ICCV Workshop. ICCVW.
Valanarasu, J., Oza, P., Hacihaliloglu, I., et al., 2021. Medical transformer: Gated axial-attention for medical image segmentation. In: MICCAI.
Varga, D., Sziranyi, T., 2016. Fully automatic image colorization based on convolutional neural network. In: 23rd International Conference on Pattern Recognition. ICPR.
Vitoria, P., Raad, L., Ballester, C., 2019. ChromaGAN: Adversarial picture colorization with semantic class distribution..
Vondrick, C., Shrivastava, A., Fathi, A., et al., 2018. Tracking emerges by colorizing videos. In: ECCV. p. 11217.
Wan, S., Xia, Y., Qi, L., et al., 2020. Automated colorization of a grayscale image with seed points propagation. IEEE Trans. Multimed. 22 (7), 1756–1768.
Wu, F., Duan, J., Chen, S., Ye, Y., Ai, P., Yang, Z., 2021a. Multi-target recognition of bananas and automatic positioning for the inflorescence axis cutting point. Front. Plant Sci. 12, 705021.
Wu, M., Jin, X., Jiang, Q., et al., 2020. Remote sensing image colorization using symmetrical multi-scale DCGAN in YUV color space. Vis. Comput. (2).
Wu, Y., Wang, X., Li, Y., et al., 2021b. Towards vivid and diverse image colorization with generative color prior. Arxiv.
Xian, W., Sangkloy, P., Agrawal, V., et al., 2018. Texturegan: Controlling deep image synthesis with texture patches. In: CVPR. pp. 8456–8465.
Xiao, J., Hays, J., Ehinger, K., et al., 2010. SUN database: Large-scale scene recognition from abbey to zoo. In: CVPR. pp. 3485–3492.
Xiao, Y., Jiang, A., Liu, C., Wang, M., 2019a. Single image colorization via modified cyclegan. In: 2019 IEEE International Conference on Image Processing. ICIP, pp. 3247–3251.
Xiao, Y., Zhou, P., Zheng, Y., Leung, C., 2019b. Interactive deep colorization using simultaneous global and local inputs. In: ICASSP. pp. 1887–189.
Xie*, M., Li*, C., Liu, X., et al., 2020. Manga filling style conversion with screentone variational autoencoder. ACM Trans. Graph. 39 (6).
Xu, J., Lu, K., Shi, X., et al., 2021. A DenseUnet generative adversarial network for near-infrared face image colorization. Signal Process. 11, 108007.
Xu, Z., Wang, T., Fang, F., et al., 2020. Stylization-based architecture for fast deep exemplar colorization. In: CVPR. pp. 9360–9369.
Xuan, A.D., Wl, B., Xw, A., 2021. Pyramid convolutional network for colorization in monochrome-color multi-lens camera system. Neurocomputing 450 (2021), 129–142.
Yang, X., Chen, J., Yang, Z., et al., 2022. Attention-guided NIR image colorization via adaptive fusion of semantic and texture clues. https://arxiv.org/abs/2107.09237.
Yang, J., Fu, X., Hu, Y., Huang, Y., Ding, X., Paisley, J., 2017. Pannet: a deep network architecture for pan-sharpening. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1753–1761.
Yin, Wang., Lu, Peng., Zhao, Zhaoran., et al., 2021. Yes, attention is all you need, for exemplar based colorization. MM.
Yoo, S., Bahng, H., Chung, S., et al., 2019. Coloring with limited data: Few-shot colorization via memory-augmented networks. In: CVPR.
Yu, F., Zhang, Y., Song, S., et al., 2015. LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop. CoRR.
Yu, L., et al., 2020. A practical PET/CT data visualization method with dual-threshold PET colorization and image fusion. Comput. Biol. Med..
Zbulak, G., 2020. Image colorization by capsule networks. In: CVPR Workshops.
Žeger, I., Grgic, S., Vuković, J., Šišul, G., 2021. Grayscale image colorization methods: Overview and evaluation. IEEE Access 9, 113326–113346. http://dx.doi.org/10.1109/ACCESS.2021.3104515.
Zhang, B., He, M., Liao, J., et al., 2019. Deep exemplar-based video colorization. In: CVPR. pp. 8044–8053.
Zhang, R., Isola, P., Efros, A.A., 2016. Colorful image colorization. In: ECCV.
Zhang, R., Isola, P., Efros, A., et al., 2018. The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR.
Zhang, L., Ji, Y., Lin, X., Liu, C., 2017. Style transfer for anime sketches with enhanced residual U-net and auxiliary classifier GAN. In: 2017 4th IAPR Asian Conference on Pattern Recognition. ACPR, pp. 506–511.
Zhang, L., Li, C., Simo-Serra, Edgar, et al., 2021a. User-guided line art flat filling with split filling mechanism. In: CVPR.
Zhang*, L., Li*, C., Wong, T., et al., 2018. Two-stage sketch colorization. ACM Trans. Graph. 37 (6), 261.
Zhang, Q., Wang, B., Wen, W., Li, H., Liu, J., 2021. Line art correlation matching feature transfer network for automatic animation colorization. In: 2021 IEEE Winter Conference on Applications of Computer Vision. WACV, pp. 3871–3880.
Zhang*, R., Zhu*, J., Isola, P., et al., 2017. Real-time user-guided image colorization with learned deep priors. ACM Trans. Graph. 36 (4), 1–11.
Zhao, J., Han, J., Shao, L., et al., 2020. Pixelated semantic colorization. Int. J. Comput. Vis. 128, 818–834. http://dx.doi.org/10.1007/s11263-019-01271-4.
Zhao, G., Huang, X., Taini, M., et al., 2011. Facial expression recognition from near-infrared videos. Image Vis. Comput. 29 (9), 607–619.
Zhong, X., Lu, T., Huang, W., et al., 2020. Visible-infrared person re-identification via colorization-based siamese generative adversarial network. In: International Conference on Multimedia Retrieval.
Zhou, J., Hong, K., Deng, T., et al., 2020. Progressive colorization via iterative generative models. IEEE Signal Process. Lett. 27 (2020), 2054–2058.
Zhou, B., Lapedriza, A., Xiao, J., et al., 2014. Learning deep features for scene recognition using places database. In: NIPS.
Zhu, J., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE International Conference on Computer Vision. ICCV, pp. 2242–2251.
Zou, C., Mo, H., Gao, C., et al., 2019. Language-based colorization of scene sketches. ACM Trans. Graph. 38 (6), 233. 1-16.
Zou, C., Yu, Q., Du, R., et al., 2018. SketchyScene: Richly-annotated scene sketches. In: ECCV. pp. 438–454.