# Chapter 1

# A review of image colourisation

Bo Li, Yu-Kun Lai and Paul L. Rosin*
*School of Mathematics and Information Science,*
*Nanchang Hangkong University, Nanchang, China, 330063*
*School of Computer Science and Informatics, Cardiff University, Cardiff, UK*

This chapter reviews the recent development of image colourisation, which aims at adding colour to a given greyscale image. There are numerous applications involving image colourisation, such as converting black and white photos or movies to colour, restoring historic photographs to improve the aesthetics of the image, as well as colourising many other types of images lacking colour (e.g. medical images, infrared night time images). According to the source where the colours come from, the existing methods can be categorised into three classes: colourisation by reference, colourisation by scribbles and colourisation by deep learning. In this chapter, we introduce the basic idea and survey the typical algorithms of each type of method.

## 1. Introduction

The first monochrome (black and white) photography was captured in 1839, and until the mid-20th century the majority of photography remained monochrome. In order to produce more realistic images, photographers and artists attempted to add colours to the black and white images. In the mid-19th century to the mid-20th century, people hand-coloured monochrome photographs manually, such as shown in Fig. 1. However, hand-colouring of photographs requires expertise knowledge and is time consuming. In 1970, the term of *colourisation* was first introduced by Wilson Markle[1] to describe the computer-assisted process for adding colours to black and white movies. It arose from colouring the classic black and white photos and videos, and now has been applied in various fields, such as hyperspectral image visualisation, designing cartoons, and 3-dimensional data rendering, etc.

For human beings, the plausible colourisation images can be produced immediately through our brain. However, it is not so direct for computers to find a reasonable colourisation result since it requires predicting R, G and B colours for

*B. Li is with the School of Mathematics and Information Science, Nanchang Hangkong University, Nanchang, China, and also with the School of Educational Information Technology, Central China Normal University, Wuhan, China. e-mail: bolimath@gmail.com. Y.-K. Lai and P. L. Rosin are with the School of Computer Science and Informatics, Cardiff University, Cardiff, UK.

Fig. 1. A hand-coloured print from the same negative, hand-coloured by Stillfried & Andersen between 1875 and 1885. `https://en.wikipedia.org/wiki/Hand-colouring_of_photographs`

each pixel with the given intensity. Mathematically, image colourisation can be formulated as follows. Given a greyscale image $L \in \mathbb{R}^{m\times n}$, where $m$ and $n$ are the width and height of the image, image colourisation aims at finding a mapping function $f$ from the intensity image $L$ to its corresponding colour version $C \in \mathbb{R}^{m\times n\times 3}$,

$$f : L \in \mathbb{R}^{m\times n} \to C \in \mathbb{R}^{m\times n\times 3} \tag{1}$$

Image colourisation attempts to extrapolate the data from 1-dimension to 3-dimensions, and is a typical ill-posed problem which does not have a unique solution. In order to reduce the linearly dependent relationship between luminance and chrominance, CIELAB or CIEYUV colour space[2] is typically adopted rather than RGB colour space.

In order to produce plausible colourisation images, numerous methods have been studied. According to the source where the colours come from, the existing methods can be categorised into three classes: colourisation by reference, colourisation by scribbles and colourisation by deep learning.

Colourisation by reference refers to transferring colour from a colour example image to the target greyscale image. This type of method is fully automatic. Given a target greyscale image, the user is only required to provide a colour reference image which has similar content to the target image, and then the colour will be transferred from the reference image to the target image automatically. However, one of the main problems is that the spatial consistency of such colourisation results is often poor. Colourisation by scribbles aims at propagating the colour scribbles specified by the user to the whole image automatically. User interaction is required to produce the colour strokes. Scribble-based methods can produce smooth colour images via the diffusion process; however, it may produce colour bleeding effects around boundaries, and the performance is highly dependent on the accuracy and amount of user interactions. Benefitting from the development of artificial intelligence and neural networks, the colour components can be effectively learned from a large number of training images, and numerous deep learning based image colourisation methods have been proposed. Despite the powerful learning ability of deep neural networks, it is difficult to control the deep model to generate user desired colourful images due to its black-box property.

This chapter is organised as follows. We will introduce the basic idea and review the corresponding typical algorithms of each type of method respectively in sections 2–5, and finally draw conclusions in section 6.

## 2. Colourisation by reference image

Colourisation by reference image means that, given a target greyscale image and a colour reference image, the colour will be transferred automatically from the reference image to the grey image to produce a colourisation result. The basic pipeline of colourisation by reference image is shown in Fig. 2. Given a colour reference and greyscale destination pair of images, the first step is feature extraction for both images. Next, for each pixel in the destination image, the most similar pixel in the reference image will be found by feature matching, and then the chrominance information will be transferred to the destination image according to the matching results to form the initial colourisation image. Finally, a propagation process is performed to produce a smooth colour image.


Fig. 2. The pipeline of colourisation by reference.

The pioneering work of colourisation by reference image was proposed by Welsh et al.[3] Motivated by colour transfer, the method transfers the colour from the reference colour image to the target greyscale image based upon independent pixelwise matches. The algorithm proposed by Welsh et al.[3] is composed of three steps. First, both target image and reference image are converted into the CIELAB colour space;[2] then for each pixel in the target grey image, the best matching pixel in the reference image is selected according to a similarity measurement based on the intensity; finally, the colour will be transferred from the reference image to the target greyscale image according to the matching results. In order to allow more user interaction and improve the matching accuracy in the colour transfer procedure, some user-provided swatches are used to limit the feature matching.

The method[3] is user friendly and automatic, however, the produced colourisation results lack spatial consistency since each pixel in the target image is processed in isolation. Numerous neighbouring pixels with similar intensity can be mismatched to different colours.

Instead of relying on a series of independent pixel-level decisions, a new strategy accounting for the higher-level context of each pixel was proposed by Irony et al.[4]

The pipeline of the proposed method is shown in Fig. 3. Given a target greyscale image and a reference colour image, the method first segments the reference colour image by using a robust supervised classification scheme. Next, each pixel in the target image is mapped to one segment. As pixel classification can lead to a vast number of misclassified pixels, a voting postprocessing step is conducted to enhance locality consistency, then the pixels with a sufficiently high confidence level will be provided as colour strokes. Finally a colour propagation scheme is used to diffuse the colours from these strokes to the whole image. The work exploits higher level features which can discriminate between different regions rather than processing each pixel independently, and guarantees spatial consistency by adopting a voting process and global diffusion. However, the performance of this method is highly reliant upon the image segmentation stage.

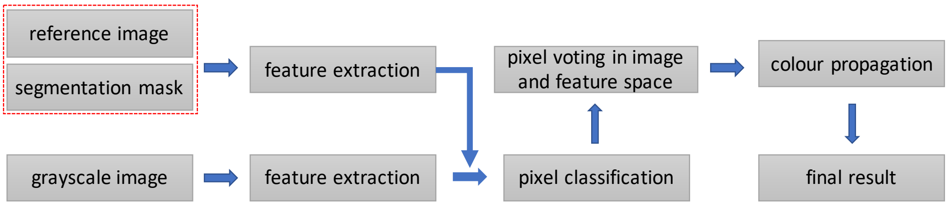


Fig. 3. Colourisation by example.[4]

A cascaded feature matching scheme was proposed in Gupta et al.[5] In this work, both the target image and reference image are first segmented into superpixels. On one hand, the use of a superpixel representation can reduce computational complexity. On the other hand, it can also enhance spatial consistency compared with procesing each pixel independently. Instead of using a combination of different kinds of features, a fast cascade feature matching scheme is adopted to find correspondences between superpixels of the reference and target images. To further enforce the spatial coherence of these initial colour assignments, an image space voting framework is used to correct invalid colour assignments.

An automatic feature selection and fusion based image colourisation method was proposed in Li et al.[6] Specifically, image regions can be generally classified as uniform background or non-uniform textures. Different regions have different characteristics and hence different features may work more effectively. Based on the above observation, the distribution of intensity deviation for uniform and non-uniform regions is learned, and the probability of a given region being assigned a uniform or non-uniform label is estimated by using Bayesian inference, which is then used for selecting suitable features. Instead of making individual decisions locally, a Markov Random Field (MRF) model is adopted to improve the labelling consistency which can be solved effectively by the graph cut algorithm (Fig. 4).

In order to enhance locality consistency, an image colourisation method based on sparse representation learning was proposed in Li et al.[7] In this work, the task

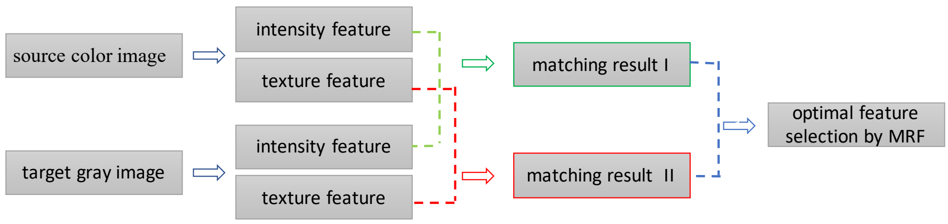


Fig. 4. Image colourisation by automatic feature selection and fusion.[6]

of colourisation is reformulated as a dictionary-based sparse reconstruction problem. Based on the assumption that superpixels with similar spatial location and/or feature representation are likely to match spatially close regions from the reference image, a new regularisation term was proposed to enhance the locality consistency. Although the locality consistent regularisation can improve the matching accuracy, the initial colourisation images obtained by feature matching often fail to preserve the edges. In order to improve colour coherence while preserving sharp boundaries, a new luminance guided joint filter was proposed. The joint filter attempts to ensure the edge structure of chrominance images is similar to the luminance channel, and the optimisation problem can be effectively solved by a typical screened Poisson equation.

Arbelot et al.[8] proposed a new edge-aware image texture descriptor by utilising spatial coherence around image structure. First, the region covariance matrices are computed to characterise the local texture. Since region covariances can only describe second-order statistics, and it is difficult to measure their similarity, the covariance matrices are then transformed into vectors by using Cholesky decomposition. Then a multiscale gradient descent is conducted to improve the features from being blurred. Finally, a luminance guided bilateral filter is utilised to improve the results while keeping sharp edges.

In the feature matching process of example-based image colourisation, the scale of features may vary between the reference image and the destination grey image. An automatic colourisation method based on a location-aware cross-scale matching algorithm and a simple combination of different scale features was proposed in Li et al.[9] First, the image pyramids are constructed for both the reference image and the destination image, then a cross-scale texture matching strategy is conducted, and the final fusion of the matching results is obtained by a global Markov Random Field optimisation. Since only low-level features are used to find the optimal matching, some unreasonable semantic errors in matching will appear. A novel up-down location aware semantic detector was proposed to automatically find these matching errors and correct the colourisation results. Finally, a nonlocal $\ell_1$ optimisation framework along with confidence weighting is used to suppress artefacts caused by wrong matchings while avoiding over-smoothing edges.

A variational colourisation model was proposed in Bugeau et al.[10] In this work, the optimal feature matching and colour propagation are simultaneously solved by a variational energy minimisation problem. First, for each pixel in the destination image, some candidate matching pixels in the reference image are selected by fast feature matching. Then a variational energy function is designed to choose the best candidate and produce a smooth colourisation result by minimising the colour variance in the interior region while keeping the edges as sharp as possible. However, the total variation regularisation term used in Bugeau et al.[10] is only composed of chrominance channels, which results in obvious halo effects around strong boundaries. Pierre et al.[11] proposed a new non-convex variational framework based on total variation defined on both luminance and chrominance channels to reduce the halo effects. With the regularisation of the luminance image, the method produces more spatially consistent results whilst preserving image contours. In addition, the authors prove the convergence of the proposed non-convex model.

Instead of local pixelwise prediction, Charpiat et al.[12] tried to solve the problem by learning multimodal probability distributions of colours, and finally a global graph cut is used for automatic colour assignment. In this paper, image colourisation is stated as a global optimisation problem with an explicit energy. First, the probability of every possible colour at each pixel is predicted, which can be seen as a conditional probability of colours given the intensity feature. Then a spatial coherence criterion is learned, and finally a global graph cut algorithm is used to find the optimal colourisation result. The method performs at the global level, and performs more robustly to texture noise and local prediction errors with the help of graph cuts.

A global histogram regression based colourisation method was proposed in Liu et al.[13] The basic assumption is that the final colourisation image and the reference image should have similar colour distributions. First, a locally weighted linear regression on the luminance histograms of both source and target images is performed. Next, zero-points (i.e., local maxima and minima) of the approximated histogram can be detected and adjusted to match between target and reference image. Then, the luminance-colour correspondence for the target image can be solved by calculating the average colour from the source image. Finally, the colourisation result is achieved by directly mapping this luminance-colour correspondence with the target image. However, due to the fact that the method does not take the structural information of the target image into account, it may produce many colour bleeding effects around boundaries.

In order to remove the influences of illumination, Liu et al.[14] proposed an illumination-independent intrinsic image colourisation algorithm (Fig. 5). First, both reference images and destination image are decomposed into reflectance (albedo) components and illumination (shading) components. In order to obtain robust intrinsic decomposition, multiple reference images containing a similar scene to the destination image are collected from the Internet. Then the colours from the reference reflectance image will be transferred to the pixels of the grey destination

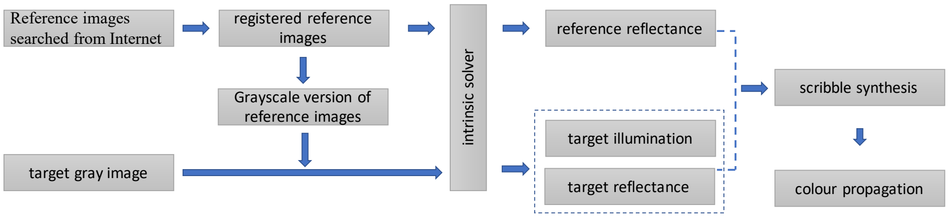


Fig. 5. Intrinsic colourisation.[14]

image with high confidence in the reference decomposition result. An optimisation model is conducted to propagate the colour through the whole destination reflectance image to enhance spatial consistency. Finally, the illumination component of the destination image is put back to produce the final colourisation result.

By leveraging the rich image content on the internet, a semantic colourisation using Internet images was proposed in Chia et al.[15] First, the user needs to provide semantic labels and segmentation clues for the foreground objects in the destination image, then for each foreground object, numerous similar colour images are collected from the Internet. In order to find the suitable candidates, an image filtering algorithm based on the spatial distribution of local and regional features is used to refine the collected images. Finally, the colours are transferred from the reference images to the destination image by using a graph-based optimisation algorithm. The method can produce multiple plausible colourisation results, although user effort is needed to assist in segmentation and label specification.

## 3. Colourisation by scribbles

Given a destination greyscale image with some pre-scribbled colour strokes, colourisation by scribbles attempts to propagate the colour from the desired colour strokes to the whole image automatically based on the assumption that neighbouring pixels with similar intensity features should have similar colours. The performance of the colourisation is dependent on the construction of the affinity matrix, and how to reduce the colour bleeding effects around boundaries is another crucial problem for the scribble-based colourisation.

The first colourisation model by scribbles was proposed in Levin et al..[16] They assume that neighbouring pixels that have similar intensity features should have similar colours. Based on the above basic assumption, the colourisation is resolved by an optimisation process. The algorithm is composed of three steps. First, the user must paint some colour scribbles in the interior of various regions, such as shown in Fig. 6. Then an affinity matrix $W$ is constructed, where each element of the affinity matrix $\omega_{r,s}$ measures the similarity between pixels $r$ and $s$. Finally, the colours from the scribbles will be propagated to the whole image by minimising the following quadratic energy function which measures the difference between the


Fig. 6. Colourisation using optimisation.[16]

colour $u_r$ at pixel $r$ and the weighted average of the colours at neighbouring pixels,

$$\min_u \sum_r (u_r - \sum_{s \in \mathcal{N}_r} \omega_{r,s} u_s)^2, \ s.t. \ u_r = u_{0,r}, r \in \Omega \tag{2}$$

where $u$ means a chrominance channel, and $\Omega$ denotes the group of user scribbles. As problem (2) is a smooth convex optimisation, it can be solved effectively by traditional methods.

However, the performance of Levin et al.'s method[16] is highly dependent on the accuracy and amount of user scribbles. For images with complex textures, a very large number of strokes are required to guarantee high quality colourisation results. In addition, there are obvious colour bleeding effects around boundaries due to the characteristic of isotropic diffusion defined in (2).

In order to reduce the burden of users, an efficient interactive colourisation algorithm was proposed in Luan et al.[17] Compared with Levin et al.'s method,[16] only a small number of colour strokes are required. The algorithm consists of two stages, colour labelling stage and colour mapping stage. In the first stage, the image will be segmented into coherent regions according to the intensity similarity to a small number of user-provided colour strokes. Instead of propagating colours from strokes to the neighbourhood pixels directly, this method first groups pixels with similar texture features which should have similar colours. The amount of colour strokes is reduced dramatically by this strategy. In the colour mapping stage, the user needs to assign the colour for a few pixels with significant luminance in each coherent region, and then the colour of the rest of the pixels will be produced by a simple linear blending by piece-wise linear mapping in the luminance channel.

Xu et al.[18] proposed to reduce the colour strokes from a feature space perspective. The method adaptively determines the impact of each colour stroke in the feature space composed of spatial location, image structure, and spatial distance.

Each stroke is confined to control a subset of pixels via a Laplacian weighted global optimisation. Numerous regularisation terms can be incorporated with the global optimisation to enhance edge-preserving property.

In Ding et al.,[19] an automatic scribble generation and colourisation method was proposed. Instead of assigning colour strokes by users, the authors propose to generate scribbles automatically by distinguishing the pixels where the spatial distribution entropy achieves locally extreme values. Then the colourisation will be conducted by computing quaternion wavelet phases along equal-phase lines, and a contour strength model is also established in scale space to guide the colour propagation while preserving the edge structure.

In order to fix the artefacts of colour bleeding around boundaries, an adaptive edge detection based colourisation algorithm was proposed in Huang et al.[20] First the reliable edge information is extracted from the greyscale image, and then the similar propagation method to Levin et al.'s method[16] is conducted with the assistance of the edge structure. In the work by Anagnostopoulos et al.,[21] salient contours are utilised to improve the colour bleeding artefacts caused by weak object boundaries. Their method is composed of two stages. In the first stage, the user-provided scribble image is enhanced with the assistance of salient contours automatically detected in the destination greyscale image. Meanwhile, the image will be segmented into homogeneous colour regions of high confidence and critical attention-needing regions. For pixels in the homogeneous regions, the colour will be diffused by the model proposed in Levin et al.'s method,[16] while for the pixels in attention-needing regions, a second edge-preserving diffusion stage will be performed with the guidance of salient contours.

In order to reduce the complexity of optimisation-based colour propagation, a fast image colourisation algorithm using chrominance blending was proposed in Yatziv et al.[22] Based on the basic observation that most of the time is spent on the iterative solution of the optimisation model defined in Levin et al.,[16] a non-iterative method was proposed in this paper. The proposed scheme is based on the concept of weighted colour blending derived from the geodesic distance between different pixels computed in the luminance channel. The method is fast and permits the user to interactively get the desired results promptly after providing a reduced set of chrominance scribbles.

Scribble-based image colourisation can also be solved by sparse representation learning.[23] First, an over-complete dictionary in chrominance space is trained on numerous sample colour patches to explore the low-dimensional subspace manifold structure. Given a greyscale image with a small subset of colour strokes, the image is first segmented into overlapping patches, and then the sparse coefficients of each patch on the pretrained dictionary can be learned using a sparse representation based on the luminance and the given chrominance within the patch. Once the sparse coefficients are solved, the colour of each patch can be generated by a sparse linear combination of the colour dictionary. A large dataset which can cover the variation of target images is required to train the dictionary, and each patch is processed independently without considering the locality consistency.

Benefitting from the strong theories and tractable computations of matrix recovery, Wang et al.[24] made the first attempt to reformulate the task of image colourisation as a matrix completion problem. Each chrominance image can be seen as a corrupted matrix with reliable values only on the locations of scribbles, then the task of image colourisation is formulated to complete the chrominance matrix with a semi-supervised learning method. Based on the basic assumption that any natural image can be effectively approximated by a low-rank matrix plus a sparse matrix, a low-rank subspace learning method which can be effectively solved by the augmented Lagrange multiplier algorithm is utilised to complete the colour matrix.

However, the image matrix cannot guarantee low-rank for images with complex textures. In this case, a colourisation algorithm by patch-based local low-rank matrix completion was proposed in Yao et al.[25] Instead of assuming that the whole image matrix is low-rank, the method first divides the image into small patches, and assumes that the subspace consists of all patches which have a low-rank structure. Then a local low-rank matrix factorisation algorithm was proposed to complete the colour image, and an efficient optimisation algorithm based on alternating direction method of multipliers was proposed.

An image colourisation method based on colour propagation and low-rank minimisation was proposed in Ling et al.[26] Given a greyscale image and a few colour strokes, the paper first propagates the colour from the colour strokes to the neighbourhood pixels according to the Chi-square distance of local texture features, and meantime a confidence map is computed. The initial colourisation result computed by propagation is not accurate enough, and so a rank optimisation constrained by the previous computed confidence map was proposed to improve the performance.

## 4. Colourisation by deep learning

Deep learning methods[27] have achieved breakthroughs in numerous research fields, such as image classification,[28–30] image segmentation[31–33] and speech recognition,[34,35] etc. Deep learning models are good at learning or approximating a non-linear mapping between different domains by training the parameters on a large dataset. For the task of image colourisation, the deep learning methods attempt to learn a mapping from the luminance channel to the chrominance channel. The input of the network is the luminance channel and the output is the chrominance channel, which when concatenated with the input luminance image produces the colourised image. We note that any colour image can be separated into its luminance and colour components, and in this manner, we can collect as many training samples as we want in order to train a neural network for image colourisation. Although we have enough training data, this learning problem is less straightforward than one may expect.

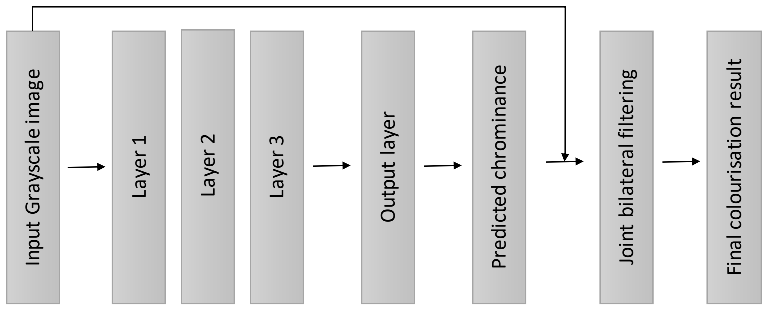


Fig. 7. Deep colourisation.[36]

The first deep learning based image colourisation method was proposed by Cheng et al.[36] In this paper, image colourisation is reformulated as a regression problem and is solved by a regular fully-connected deep neural network (Fig. 7). Finally, a joint bilateral filtering is utilised as a post-processing step to reduce the artefacts. The model utilised in this paper is a three-layer fully connected neural network. Three levels of features are utilised, including the raw image patch, DAISY features, and semantic features. Given the features as the input, the output of the network is the prediction of the corresponding chrominance. 2344 training images are used to train the network. In addition, the model requires handcrafted features as the input of the network rather than learning features solely from the input images themselves.

Instead of relying on hand-crafted features, a full automatic end-to-end image colourisation model was proposed by Larsson et al.[37] A pretrained VGG network is utilised to generate features of different scales. For each pixel, a hypercolumn feature is extracted by concatenating the features at its spatial location in all layers, which incorporates the semantic information and localisation property. Taking into account that some objects (such as clothing) may be drawn from many suitable colours, this paper treats colour prediction as a histogram estimation task rather than as regression, and a KL-divergence based loss function is designed to measure the prediction accuracy.

Due to the underlying uncertainty of image colourisation, regression based learning methods often result in desaturated colourisation. Zhang et al.[38] proposed a novel classification based colourisation network. In order to model the multimodal nature of image colourisation, the authors attempt to predict a distribution of possible colours for each pixel rather than a fixed colour. Due to the distribution of chrominance values in natural images being strongly biased, a class rebalancing process is utilised to emphasise rare colours. Finally, vibrant and realistic colourisation results are produced by taking the "annealed mean" of the distribution. The main contribution of this work is designing an appropriate objective function that handles the multimodal uncertainty of the colourisation problem and captures a wide diversity of colours. In addition, this paper proposes a novel framework for testing colourisation results.

A novel end-to-end framework which combines both global priors and local image features was proposed in Iizuka et al.[39] The proposed architecture can extract local, mid-level and global features jointly from an image, which can then be fused for predicting the final colourisation. In addition, a global semantic class label is utilised during the training process to learn more discriminative global features. The proposed model is composed of four main components: a low-level feature network, a mid-level feature network, a global feature network and a colourisation network. First, a 6-layer convolutional neural network is used to learn the low-level features from the image, and then mid-level and high-level features are learned based on the shared low-level features. Next, a fusion layer is designed to incorporate the global features into local mid-level features, and then the fused features are processed by a set of convolutions and upsampling layers to generate the final colourisation results. In order to incorporate global semantic priors, a global classification branch is added to help learn the global context of the image. In addition, the model can directly transfer the style of an image into the colourisation of another.

Colourisation is an ambiguous problem, with multiple plausible colourisation results being possible for a single grey-level image. For example, a tree can be green, yellow, brown or red. However, the above end-to-end deep learning methods can only produce a single colourisation.

A user-guided deep image colourisation method was proposed in Zhang et al.[40] Compared with traditional optimisation-based interactive colourisation methods,[16] the proposed deep neural network propagates user edits by fusing low-level cues along with high-level semantic information, learned from a million images rather than using hand-defined rules. The proposed network learns how to propagate sparse user hints by training a deep network to directly predict the mapping from a greyscale image and randomly generated user colour hints to a full colour image on a large dataset. In addition, a data-driven colour palette is designed to suggest colours for each pixel.

Another user-guided deep colourisation method was proposed in He et al.[41] In this paper, a reference colour image is used to guide the output of the deep colour model rather than using user-provided local colour hints as in Zhang et al.[40] It is the first deep learning approach for exemplar-based local colourisation. The proposed network is composed of two sub-networks, a similarity sub-net and a colourisation sub-net. The similarity sub-net can be seen as a pre-processing step for the colourisation sub-net. It computes the semantic similarities between the reference and the target image by using a pretrained VGG-19 network, and generates a dense bidirectional mapping function by using a deep image analogy technique. Then the greyscale target image, the colour reference image and the learned bidirectional mapping functions are fed into the colourisation sub-net. The architecture of the colourisation sub-net is a typical multi-task learning framework consisting of two branches. The first branch is used to measure the chrominance loss, which encourages the propagated colourisation results to be as close as possible to the ground truth chrominance. A high-level perceptual loss is introduced in the second branch, which is used to predict perceptually plausible colours even without a proper reference. In addition, a novel image retrieval algorithm was proposed to automatically recommend good references to the user.

Another way to produce diverse colourisation results for a grey-scale target image is to learn a conditional probability model in a low dimensional embedding of the colour fields. Given a target grey-scale image $G$, a conditional probability $P(C|G)$ for the chrominance field $C$ is learned from a large scale dataset. Then diverse colourisation results can be generated by drawing samples from the learned model, $\{C_k\}_{k=1}^{N} \sim P(C|G)$. However, it is difficult to learn such a conditional distribution in high-dimensional colour spaces.

A variational autoencoder model which aims to learn a low dimensional embedding of colour spaces was proposed in Deshpande et al.[42] Instead of learning the conditional distribution in the original high dimensional colour spaces, the method attempts to find a low dimensional feature representation of colour fields which is useful to build such a prediction model. Loss functions consist of specificity, colourfulness and gradient terms which are designed to avoid the over-smoothing and washed out colour effects. Finally, the samples from the learned conditional model result in diverse colourisation results.

A probabilistic image colourisation model was proposed in Royer et al.[43] The proposed network consists of two sub-nets. The first feed-forward network learns a low-dimensional embedding encoding information about plausible image colours based on the high-level features learned from the greyscale image. Then the embedding is fed to an autoregressive PixelCNN network[44] to predict a proper distribution of the image chromaticity conditioned on the greyscale input. In order to enhance the structural consistency, a conditional random field based variational auto-encoder formulation was proposed in Messaoud et al.[45] The method attempts to produce diverse colourisation results while taking structural consistency into account. Moreover, the method also incorporates external constraints from diverse sources including a user interface.

A pixel recursive colourisation method was proposed in Guadarrama et al.[46] Based on the observation that image colourisation is robust to the scale of the chrominance image, a conditional PixelCNN[44] is first trained to generate a low resolution colour image for a given greyscale image with relatively higher resolution. Then, a second convolutional neural network is trained to generate a high-resolution colourisation with the input of the original greyscale image and the low resolution colour image learned in the first stage.

Image colourisation was solved by a conditional generative adversarial network in Isola et al.[47] Given an input grey-scale image, a colourisation will be generated conditioned on the input by a generative adversarial network. Instead of designing an effective loss function manually with convolutional neural network based methods, a high-level semantic loss function which can make the output indistinguishable from reality is learned automatically by a generative adversarial network, which is then used to train the network to learn the mapping from the input image to the output image. In addition, the proposed architecture can learn a loss that adapts to the data, which avoids designing different loss functions for specific tasks.

The image-to-image translation network has to be trained on aligned image pairs in Isola et al.,[47] however, paired training data is not available for many tasks. A novel unpaired image-to-image translation framework was proposed in Zhu et al.[48] by using cycle-consistent adversarial networks. The proposed cycle generative model first learns a mapping from the source domain to the target domain, and then an inverse mapping is introduced to learn the mapping from the target domain to the input domain coupled with a cycle consistency loss function.

Some experimental results of deep learning based colourisation methods are shown in Fig. 8. For popular scenes, such as shown in the first row, most of the algorithms can produce plausible results. However, when the texture of different objects look similar, many semantically wrong colours will be generated, such as shown in the second and the third row. In addition, the colourisation results generated by most of the existing methods are not colourful enough yet, such as shown in the last three rows.

## 5. Other related work

Interactive outline colourisation is a special case of image colourisation. Outline colourisation aims to generate a colour and shaded image from a black and white outline, such as shown in Fig. 9. Different from greyscale images, outline images do not have greyscale information. Therefore, most of the existing image colourisation methods cannot achieve desirable performance on outline images.

One of the first manga colourisation methods was proposed in Qu et al.[49] The proposed method propagates colour strokes with the constraints of both pattern-continuity and intensity-continuity. The algorithm is composed of two steps: segmentation and colour-filling. In the first step, local and statistical pattern features are extracted by Gabor wavelet filters, which are then used to guide the boundary segmentation based on a level set propagation. After the segmentation, various colour propagation techniques can be employed in the second stage for filling colours.

A deep learning based outline colourisation method was proposed in Frans.[50] The model is composed of two distinct convolutional networks in tandem. The first sub-net attempts to predict the colour based only on outlines, and the second network is adopted to generate shadings conditioned on both outlines and a colour scheme. Both networks are trained in tandem to produce a final colourisation.

Based on the observation that the background colours of comics are often consistent but random, a consistent comic colourisation with pixel-wise background classification was proposed in Kang et al.[51] A conditional generative adversarial network based outline colourisation was proposed in Liu et al.[52] Given a black-

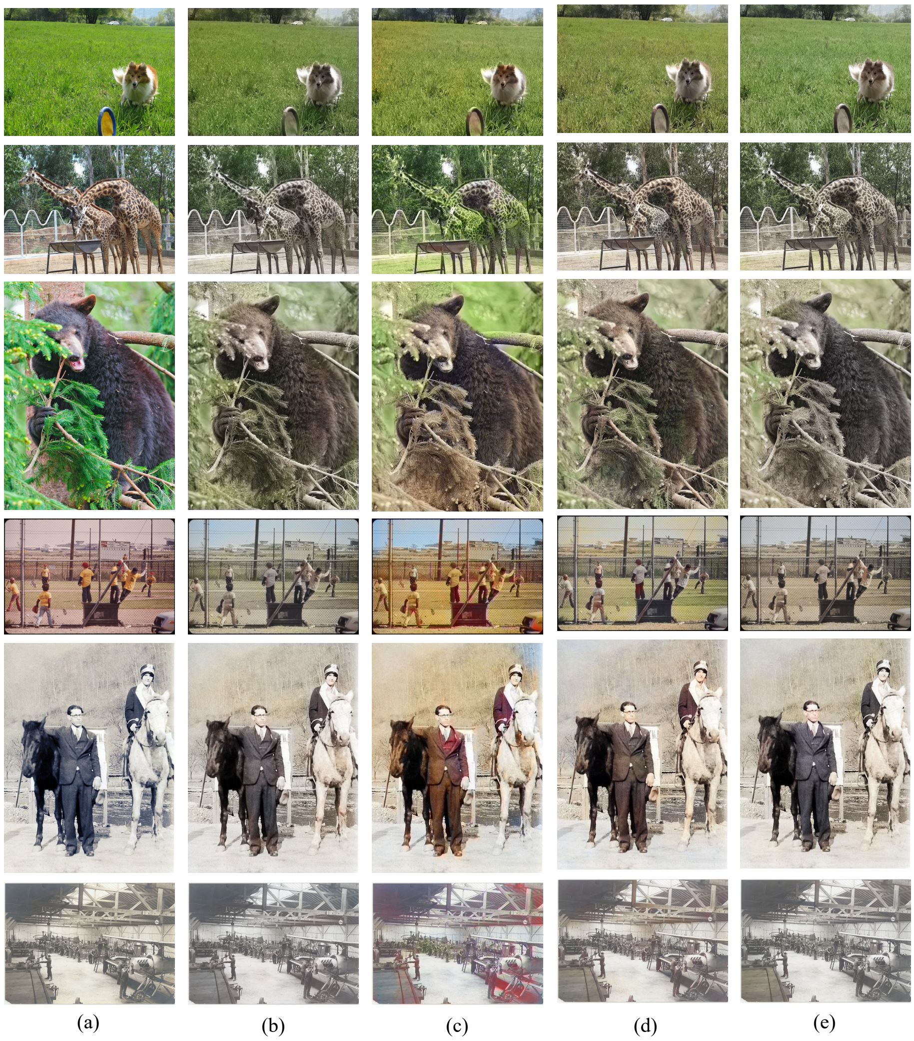

Fig. 8. Colourisation results by deep learning models. From (a) to (e) are respectively the ground truth and results of methods: Larson et al.,[37] Zhang et al.,[38] Iizuka et al.[39] and Zhang et al.[40]

and-white outline image, an auto-painter module was proposed to automatically generate plausible colours adapted with colour control. The Wasserstein distance is adopted to assist the training of the generative network. A user-guided deep anime line art colourisation method with conditional adversarial networks was proposed in Ci et al.[53]

A two-stage sketch colourisation method was proposed in Zhang et al.[54] In the first stage, a convolutional neural network is trained to determine the colour composition and predict a coarse colourisation with a rich variety of colours over

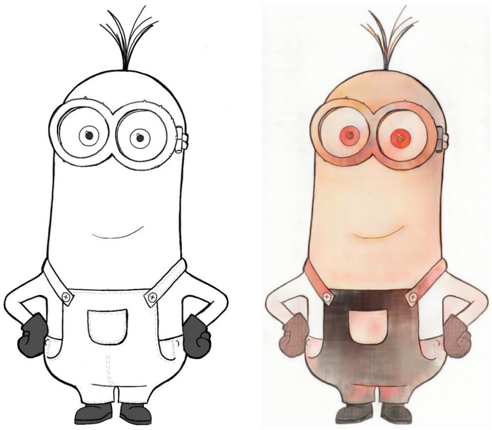

Fig. 9. Outline colourization through tandem adversarial networks.[50]

the outlines. Then in the second stage, incorrect colour regions are detected and refined with an additional set of user hints. Each stage is learned independently in the training phase, and in the test phase they are concatenated to produce the final colourisation.

## 6. Conclusion

Image colourisation is an important and difficult research topic in computer vision. It is a typical ill-posed problem since image colourisation attempts to extrapolate the data from 1-dimensional greyscale images to 3-dimensional colour images. The history of the development of image colourisation is introduced in this chapter, and most of the existing methods are reviewed and discussed. Compared with traditional reference-based or scribble-based methods, deep learning methods can leverage large-scale data to learn high level features, and produce robust and meaningful colourisation results. However, due to the black-box property of deep learning models, the colourisation results produced by deep learning methods are more difficult to control. Moreover, most of the existing work evaluates the colourisation results by visual inspection, as no quantitative metric has been developed to measure the quality of the colourisation results. How to develop a meaningful metric is an interesting and important research topic for future work.

## References

1. G. C. Burns, Museum of broadcast communications: Encyclopedia of television, *http://www.museum.tv/archives/etv/index.html.* (1970).
2. T. Smith and J. Guild, The CIE colorimetric standards and their use, *Transactions of the Optical Society.* **33**(3), 73, (1931).

3. T. Welsh, M. Ashikhmin, and K. Mueller, Transferring color to greyscale images, *ACM Transactions on Graphics.* **21**(3), 277–280, (2002).
4. R. Irony, D. Cohen-Or, and D. Lischinski. Colorization by example. In *Eurographics Conference on Rendering Techniques*, pp. 201–210, (2005).
5. R. K. Gupta, A. Y.-S. Chia, D. Rajan, E. S. Ng, and H. Zhiyong. Image colorization using similar images. In *ACM International Conference on Multimedia*, pp. 369–378, (2012).
6. B. Li, Y.-K. Lai, and P. L. Rosin, Example-based image colorization via automatic feature selection and fusion, *Neurocomputing.* **266**, 687–698, (2017).
7. B. Li, F. Zhao, Z. Su, X. Liang, Y.-K. Lai, and P. L. Rosin, Example-based image colorization using locality consistent sparse representation, *IEEE Transactions on Image Processing.* **26**(11), 516–525, (2017).
8. B. Arbelot, R. Vergne, T. Hurtut, and J. Thollot. Automatic texture guided color transfer and colorization. In *Expressive*, (2016).
9. B. Li, Y.-K. Lai, M. John, and P. L. Rosin, Automatic example-based image colourisation using location-aware cross-scale matching, *IEEE Transactions on Image Processing.* (2019).
10. A. Bugeau, V.-T. Ta, and N. Papadakis, Variational exemplar-based image colorization, *IEEE Transactions on Image Processing.* **23**(1), 298–307, (2014).
11. F. Pierre, J.-F. Aujol, A. Bugeau, N. Papadakis, and V.-T. Ta, Luminance-chrominance model for image colorization, *SIAM J. Imaging Sciences.* **8**(1), 536–563, (2015).
12. G. Charpiat, M. Hofmann, and B. Schölkopf. Automatic image colorization via multimodal predictions. In *European Conference on Computer Vision*, pp. 126–139. (2008).
13. S. Liu and X. Zhang, Automatic grayscale image colorization using histogram regression, *Pattern Recognition Letters.* **33**(13), 1673–1681, (2012).
14. X. Liu, L. Wan, Y. Qu, T.-T. Wong, S. Lin, C.-S. Leung, and P.-A. Heng, Intrinsic colorization, *ACM Trans. Graph.* **27**(5), 152, (2008).
15. A. Y.-S. Chia, S. Zhuo, R. K. Gupta, Y.-W. Tai, S.-Y. Cho, P. Tan, and S. Lin, Semantic colorization with internet images, *ACM Trans. Graph.* **30**(6), 156, (2011).
16. A. Levin, D. Lischinski, and Y. Weiss, Colorization using optimization, *ACM Transactions on Graphics.* **23**(3), 689–694, (2004).
17. Q. Luan, F. Wen, D. Cohen-Or, L. Liang, Y.-Q. Xu, and H.-Y. Shum. Natural image colorization. In *Eurographics Conference on Rendering Techniques*, pp. 309–320. Eurographics Association, (2007).
18. L. Xu, Q. Yan, and J. Jia, A sparse control model for image and video editing, *ACM Transactions on Graphics (TOG).* **32**(6), 197, (2013).
19. X. Ding, Y. Xu, L. Deng, and X. Yang, Colorization using quaternion algebra with automatic scribble generation, *Advances in Multimedia Modeling.* pp. 103–114, (2012).
20. Y.-C. Huang, Y.-S. Tung, J.-C. Chen, S.-W. Wang, and J.-L. Wu. An adaptive edge detection based colorization algorithm and its applications. In *ACM Multimedia*, pp. 351–354, (2005).
21. N. Anagnostopoulos, C. Iakovidou, A. Amanatiadis, Y. Boutalis, and S. A. Chatzichristofis. Two-staged image colorization based on salient contours. In *IEEE International Conference on Imaging Systems and Techniques*, pp. 381–385, (2014).
22. L. Yatziv and G. Sapiro, Fast image and video colorization using chrominance blending, *IEEE Transactions on Image Processing.* **15**(5), 1120–1129, (2006).
23. J. Pang, O. C. Au, K. Tang, and Y. Guo. Image colorization using sparse representation. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1578–1582, (2013).

24. S. Wang and Z. Zhang. Colorization by matrix completion. In *AAAI Conference on Artificial Intelligence*, pp. 1169–1175, (2012).
25. Q. Yao and J. T. Kwok. Colorization by patch-based local low-rank matrix completion. In *AAAI Conference on Artificial Intelligence*, pp. 1959–1965, (2015).
26. Y. Ling, O. C. Au, J. Pang, J. Zeng, Y. Yuan, and A. Zheng. Image colorization via color propagation and rank minimization. In *IEEE International Conference on Image Processing*, pp. 4228–4232, (2015).
27. Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature*. **521**(7553), 436, (2015).
28. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: a large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, (2009).
29. A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pp. 1097–1105, (2012).
30. M. Tan and Q. V. Le, EfficientNet: rethinking model scaling for convolutional neural networks, *arXiv preprint arXiv:1905.11946*. (2019).
31. R. Girshick. Fast R-CNN. In *IEEE International Conference on Computer Vision*, pp. 1440–1448, (2015).
32. J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, (2015).
33. K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask R-CNN. In *IEEE International Conference on Computer Vision*, pp. 2961–2969, (2017).
34. G. Hinton, L. Deng, D. Yu, G. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, B. Kingsbury, et al., Deep neural networks for acoustic modeling in speech recognition, *IEEE Signal Processing Magazine*. **29**, (2012).
35. A. Graves, A.-R. Mohamed, and G. Hinton. Speech recognition with deep recurrent neural networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 6645–6649, (2013).
36. Z. Cheng, Q. Yang, and B. Sheng. Deep colorization. In *IEEE International Conference on Computer Vision*, pp. 415–423, (2015).
37. G. Larsson, M. Maire, and G. Shakhnarovich. Learning representations for automatic colorization. In *European Conference on Computer Vision*, pp. 577–593. Springer, (2016).
38. R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization. In *European Conference on Computer Vision*, pp. 649–666. Springer, (2016).
39. S. Iizuka, E. Simo-Serra, and H. Ishikawa, Let there be color!: Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification, *ACM Transactions on Graphics (TOG)*. **35**(4), 110, (2016).
40. R. Zhang, J.-Y. Zhu, P. Isola, X. Geng, A. S. Lin, T. Yu, and A. A. Efros, Real-time user-guided image colorization with learned deep priors, *ACM Transactions on Graphics (TOG)*. **36**(4), 119, (2017).
41. M. He, D. Chen, J. Liao, P. V. Sander, and L. Yuan, Deep exemplar-based colorization, *ACM Transactions on Graphics (TOG)*. **37**(4), 47, (2018).
42. A. Deshpande, J. Lu, M.-C. Yeh, M. Jin Chong, and D. Forsyth. Learning diverse image colorization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6837–6845, (2017).
43. A. Royer, A. Kolesnikov, and C. H. Lampert, Probabilistic image colorization, *CoRR*. **abs/1705.04258**, (2017).
44. A. Van den Oord, N. Kalchbrenner, L. Espeholt, O. Vinyals, A. Graves, et al. Condi-

tional image generation with pixelCNN decoders. In *Advances in Neural Information Processing Systems*, pp. 4790–4798, (2016).
45. S. Messaoud, D. Forsyth, and A. G. Schwing. Structural consistency and controllability for diverse colorization. In *Proceedings of the European Conference on Computer Vision*, pp. 596–612, (2018).
46. S. Guadarrama, R. Dahl, D. Bieber, M. Norouzi, J. Shlens, and K. Murphy. Pixcolor: Pixel recursive colorization. In *BMVC*, (2017).
47. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134, (2017).
48. J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE International Conference on Computer Vision*, pp. 2223–2232, (2017).
49. Y. Qu, T.-T. Wong, and P.-A. Heng. Manga colorization. In *ACM Transactions on Graphics (TOG)*, vol. 25, pp. 1214–1220, (2006).
50. K. Frans, Outline colorization through tandem adversarial networks, *arXiv preprint arXiv:1704.08834*. (2017).
51. S. Kang, J. Choo, and J. Chang. Consistent comic colorization with pixel-wise background classification. In *NIPS'17 Workshop on Machine Learning for Creativity and Design*, (2017).
52. Y. Liu, Z. Qin, T. Wan, and Z. Luo, Auto-painter: Cartoon image generation from sketch by using conditional wasserstein generative adversarial networks, *Neurocomputing*. **311**, 78–87, (2018).
53. Y. Ci, X. Ma, Z. Wang, H. Li, and Z. Luo. User-guided deep anime line art colorization with conditional adversarial networks. In *2018 ACM Multimedia Conference on Multimedia Conference*, pp. 1536–1544, (2018).
54. L. Zhang, C. Li, T.-T. Wong, Y. Ji, and C. Liu. Two-stage sketch colorization. In *SIGGRAPH Asia 2018 Technical Papers*, p. 261. ACM, (2018).