An Overview of Grayscale Image Colorization Methods

Ivana Žeger, Sonja Grgić
University of Zagreb, Faculty of Electrical Engineering and Computing
Unska 3, 10 000 Zagreb
ivana.zeger@fer.hr

***Abstract*—Conversion of grayscale images to color images is a process of adding color to gray, monochrome images in a convincing, visually acceptable way. Nowadays, automated conversion is a challenging area that links machine and deep learning methods with art. Although many experts claim that grayscale images contain a special artistic value, lack of color can be considered as a loss of information. This paper presents an overview of methods and techniques that have been developed for grayscale image colorization. The paper provides a classification of relevant methods, explains the principles on which they are based and emphasizes their advantages and disadvantages. Special focus is put on methods that involve deep learning algorithms. The results show that deep learning colorization methods provide automated conversion and outperform other methods both in terms of quality and speed.**

***Keywords*—Colorization; Grayscale Image; Color Image; Scribble-based Methods; Example-based Methods; Deep Lerning Methods**

I. Introduction

Colorization is, in its essence, a process of imagining color where there is none. In a technical sense, it is a process of assigning three-dimensional RGB color information to every pixel with respect to intensity of a grayscale image in a visually acceptable, plausible way. For complexity reduction, conversion to a convenient luminance-chrominance color space (e.g. LAB, YUV) is used allowing the exploitation of the intensity information and the prediction of the two remaining color channels.

Historical black and white images, although with unattainable scene representations, are regarded as irreplaceable with an exceptional artistic value. However, color is a mighty tool of expression. The revival of black and white images strongly changes viewers' perspective. The time gap between the past and the present fades away while making the scene more conceivable. Early manual colorization techniques date back to the 19th century [1]. Well-known techniques from that period include coloring a daguerreotype with a mixture of gum arabic and pigments as well as photochrom process [2]. In 1970s, following the impact of the digital revolution, colorization is transferred to the computer domain. The term "colorization" was introduced by Wilson Markle to describe the computer-assisted process of adding color to black and white movies or TV programs [3]. Although many preceding techniques have been proprietary, it is known that the initial colorization attempts were characterized by low contrast and washed-out, pale resulting colors. In addition, considerable amount of human intervention was required in the process.

Technology development has brought more automated machine and deep learning techniques into focus. These techniques have demonstrated their effectiveness in various computer vision and image processing applications [4]. Both machine and deep learning handle enormous amounts of data extremely well while unfolding hidden patterns and producing satisfying approximations of the latent knowledge. Whilst machine learning endeavors to define a set of rules in data by extracting features regarding some form of *a priori* knowledge, its narrower field, deep learning, extracts regularities more independently utilizing a hierarchical level of artificial neural networks. This way, achieving exceptional colorization advantages has become possible [5-26].

II. Colorization methods

Research papers involving colorization vary substantially. The imagination and dissimilarity of problem-solving approaches make molding numerous methods into the appropriate categories an extremely hard task. Most existing papers classify colorization methods considering the amount of user involvement in problem solving and the way of retrieving the data required [7-11, 13, 14, 17]. Methods can be roughly divided into scribble-based and example-based. Source images for example-based methods may be obtained manually or automatically. However, in recent times, deep learning techniques have shown remarkable progress thereby indicating the need to elevate themselves from example-based category. That is why contemporary papers introduce using deep learning models with large quantity of training data as an additional method division criterion [19, 21]. Consequently, colorization methods are divided into scribble-based, example-based and deep learning methods.

A. Scribble-based methods

These methods find the inspiration in early digital colorization attempts dating to the end of the second half of the 20th century. Back in the day, the technicians used to choose by hand the convenient colors for every object contained in an image. In cases of a lack of the ground truth color information, they were engaged in speculation. Thus, scribble-based methods require annotating the grayscale image with marks of convenient colors, i.e. scribbles. They serve as a landmark for colorization. Scribbles are used-made and placed upon certain areas of the image. Color from the scribble is propagated across the image to the borders specified by intensity according to some optimization framework. An example of the colorization result made by applying scribbles is shown in Fig. 1.

978-1-7281-5973-7/20/$31.00 ©2020 IEEE

a) b) c)

Figure 1. a) Original photograph, b) scribbles applied on the grayscale version of the photograph, c) colorization result

The basic scribble-based method is presented in [6]. Spatial continuity is exploited on the assumption that neighboring pixels in space-time that have similar intensities should have similar color. By working in YUV color space, it is presumed that color is a linear function of intensity Y. The least squares optimization is used. The scribbles are formed as linear constraints of the optimization problem. The main advantage of this method is a fact that a variety of subsequent, enhanced methods have taken over its optimization function [7, 8] or its slightly revised version.

General advantages of scribble-based methods include globality – there is no need for explicit segmentation due to color propagation limiting with intensity values. Furthermore, there is no need for searching an adequate, possibly unreachable referent image. The user may allocate the scribbles strategically or even add more scribbles if needed. Besides, the user has the potential of changing a desirable color. On the other hand, scribble-based methods demand significant human effort considering the necessary time as well as experience and a sense of aesthetics. A careful selection of palette colors is a prerequisite.

B. Example-based methods

The idea of the example-based methods is transferring color information from a colored source image to the matching regions of the target grayscale image. The improvement of internet search engines and the appearance of centralized and indexable image databases enabled the transfer of the general "atmosphere" between images [5]. The goal is to locate pixels with corresponding luminance values in both target and reference images by means of neighboring pixels' statistical information and similar texture recognition. However, the usage of local information is conspicuous. While local information is crucial for disjoining the objects' boundaries within the image correctly, broadening the horizon with global information highly increases the probability of accurate color assignment.

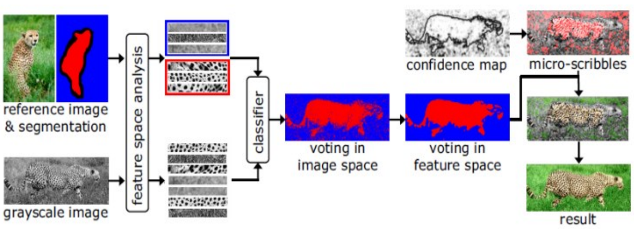

Figure 2. Structure of Irony et al. example-based method [7]

Since 2005, mentioned methods have been reviewed, improved and combined. Image processing and the usage of machine learning intensify. Image segmentation is used for more efficient color assignment as shown in Fig. 2 [7]. Suitable feature space for region differentiation using k-Nearest Neighbors (k-NN) algorithm is constructed [7]. The influence of the shadows and light reflection as well as changing lighting conditions is asserted [8]. The emphasis is switched to global problem formulation with a statistical approach to its solving [9].

Since 2010 more attention has been given to qualitative and quantitative method comparisons. Additionally, an increasing number of reference images is used while searching for corresponding color of segmented image objects. Mainly local features and their influence on color transfer is explored [10, 11]. For feature studies, groups of pixels like patches and superpixels are used while exploiting spatial consistency. Grouping of segments assigned with similar color values is carried out with k-means algorithm gaining reliability of color assessment. More complicated mathematical formulations of loss functions with manually adjustable parameters are used alongside with more advanced optimization methods [12].

The limitations of example-based methods include possible non-existence of a single suitable reference image. Besides, the selection of the appropriate reference image is often done manually. In addition, the quality of the result is highly dependent on the quality of the image found. The target and the reference images need to indicate visual similarity. Also, algorithm overfitting is risked by the usage of a single reference image (or even a small number of images). However, example-based methods are characterized by simplicity and speed.

C. Deep learning methods

The expectation that several reference images could contain enough color information for appealing colorization results is ingenuous. The evolution of deep learning has enabled user activity reduction by training an artificial neural network with plenty of source images. Methods yield better results by adding more layers to the network and more images to the training set. Deep neural networks have been introduced to the colorization problem in 2015 [13]. Neural networks automatically extract regularities within data by minimizing the corresponding loss function in the training phase. Model learns a mapping function (between pixels' features in a grayscale image and color values of the source image) automatically.

Basic deep neural networks remove the spatial form of an image. The decomposition of an image into one-dimensional vector of numbers occurs at the input, before propagation to the subsequent layers. Convolutional neural networks (CNN) conserve the spatial information. CNNs are widespread means for image processing. Even in colorization problem, CNNs enjoy enormous popularity [14-17].

Frequently adverted CNN regression colorization method is proposed in [14]. Regression methods use Euclidean distance between the predicted and the ground truth color as the loss function. This type of loss is exceedingly esteemed for its simplicity in terms of reasoning and computation. Believable results very different from the ground truth are severely punished. Also, unfamiliar objects are assigned desaturated reddish and brownish color values. In [14] the importance of global features (scene semantics) is emphasized in reducing solution ambiguities. Nevertheless, local information is not diminished. The network model consists of four main components: a low-level features network, a mid-level features network, a global features network and a colorization network, as shown in Fig. 3. Local and global information are fused together enabling total automaticity. The model is trained *end-to-end* on a large dataset for scene recognition with a joint colorization and classification loss. Classification of the scene in one of predetermined categories significantly improves the result.

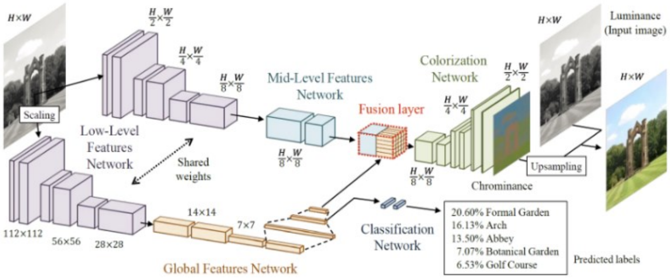

Figure 3. Architecture of Iizuka et al. [14]

The most well-known colorization method is introduced in [15] and shown in Fig. 4. It is based on multinomial classification of pixels according to color and class rebalancing for increasing diversity of resulting colors. The distribution of possible colors is predicted for every pixel. The classification of pixels is determined by probabilities of belonging into one of 313 segments of discretized and quantized ab-plane of LAB color space. The major contribution of the method is the observation that the number of pixels in natural images at desaturated values is orders of magnitude higher than for saturated values. Without taking this into account, the loss function is dominated by desaturated ab values. Hence, the adjustment of the loss function by reweighting the loss at training time is used based on the pixel color rarity. The need for modifying the loss function, in opposition to the widely accepted Euclidean distance loss is noticed even in [16]. Color histogram prediction in [17] skillfully resolves ambiguities.

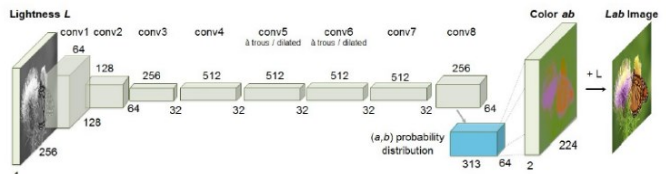

Figure 4. Architecture of Zhang et al. [15]

Since 2017, Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) have come under the spotlight. The main advantage of GANs includes the fact that the rivalry between the two neural networks, generator and discriminator, creates the corresponding loss function while producing vivid results [19]. Systems that avoid regression averaging of possible colors are created with modifications of ordinary architecture and loss function while producing manifold of realistic results [18, 20].

Furthermore, GANs join forces with semantic information exploitation [21]. Unsupervised learning techniques are heavily investigated [20, 21, 23]. Language-based colorization is explored [25]. Colorization methods are merged with challenging computer graphics problems; 3D modeling and image synthesis from 2D and 3D models [22]. Methods adjusted to coloring comic books, cartoons, icons, fashion sketches are being developed [22, 23].

III. COMPARION OF COLORIZATION METHODS

Colorization methods differ in the amount of necessary human intervention, the ways of retrieving reference images, whether they use deep learning, etc. Furthermore, time complexity, computational complexity and the ways of hardware usage are indispensable indicators of method divergence. Nevertheless, the direct comparison of certain methods is difficult to perform mainly because of the great diversity of the main ideas in the problem-solving approaches. Although many methods share approach similarity, especially in deep learning category, some are trained with different datasets under different terms, mainly regarding training time.

In this paper methods [6, 14, 15, 26] are compared. The method presented in [6] uses color scribbles. The methods proposed in [14, 15] are deep learning methods trained on ImageNet dataset. An adaptation of the method [26] was implemented in Keras and trained on an improvised dataset. The accent of creation was on the comprehension of the main concept of colorization featuring neural networks. Moreover, the incorporation of the basic steps needed for successful problem solving was crucial. Necessary training parameters were investigated and tuned.

For method comparison the author's photography collection was used. Image dataset contains 100 images with resolution of 320x240. Dataset includes 35 images of nature, 10 images of people, 15 images of buildings, 6 images of text, 5 images of animals, 7 taken during night and 22 images of various objects in interior and exterior. The original photographs, their grayscale versions and the colorization method comparison are shown in Fig. 5-7. More results can be seen in [27]. In Fig. 5 is shown that all evaluated methods are able to construct an appealing solution. Regression methods [14, 26] tend to demonstrate desaturation as shown in Fig. 6. An abstract scene is shown in Fig. 7. Deep learning methods show their weakness there because of an inappropriate dataset. On the other hand, the dominance of the scribble-based method is highlighted because of adding *a priori* knowledge.

CONCLUSION

In this paper, a survey of colorization methods was conducted with emphasis on [6, 14, 15, 26]. Based on the Keras implementation of [26] it is inferred that a big training set is a prerequisite for this process. The speed of the long training could be increased by using stronger hardware and GPU acceleration. Parameter tuning is exceptionally important for avoiding overfitting. Dataset content could cause bias in colorization results. Although the best results have been obtained by deep learning methods, especially with the discrete, multimodal method [15], the potential upgrade of the results can be accomplished. Even though the essence of machine learning is generalization, it is not possible to expect it to work with this type of problem. The creation of a huge dataset that would contain natural images, rare scenes and abstract representations would not yield to a better colorization, but to averaging the possibilities of assigning a certain color resulting in desaturation. The proposition is to use different type of methods depending on the desired outcome.

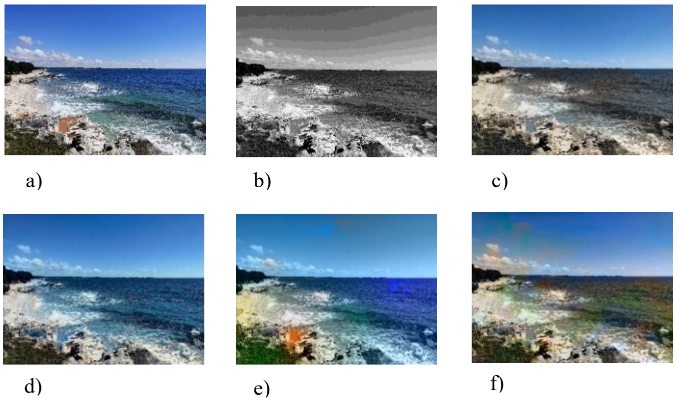

Figure 5. a) Original photograph, b) grayscale version, c) result of [14], d) result of [15], e) result of [6], f) result of [26]

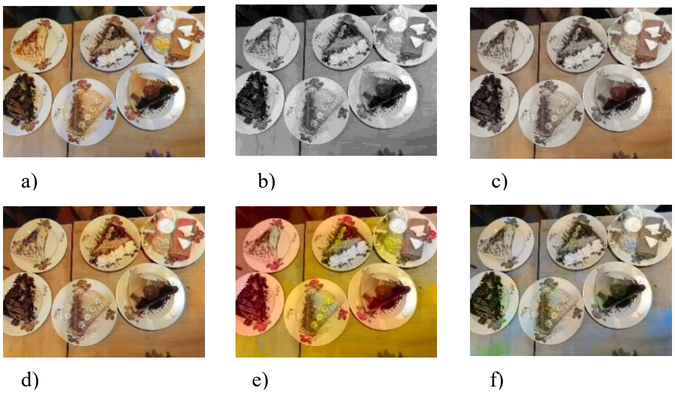

Figure 6. a) Original photograph, b) grayscale version, c) result of [14], d) result of [15], e) result of [6], f) result of [26]

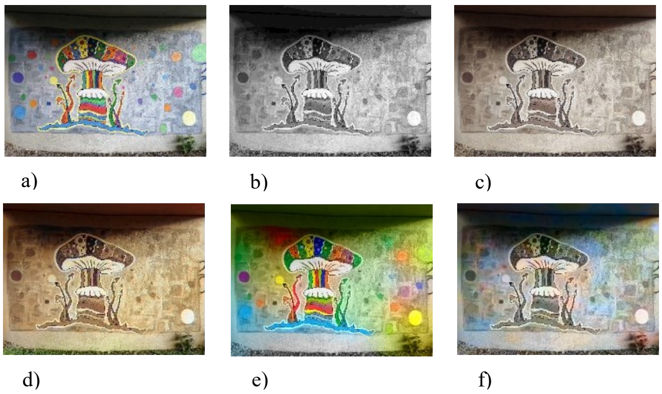

Figure 7. a) Original photograph, b) grayscale version, c) result of [14], d) result of [15], e) result of [6], f) result of [26]

References

[1] A Weltz, "The basics of hand-coloring black-and-white prints," https://www.bhphotovideo.com/explora/photography/tips-and-solutions/the-basics-of-hand-coloring-black-and-white-prints, 2020

[2] G. Gene, "Scanning around with Gene: the miracle of photochrom," https://creativepro.com/scanning-around-gene-miracle-photochrom/, 2009

[3] E. Trex, "How (and why) are black and white films colorized?," https://www.mentalfloss.com/article/26956/how-and-why-are-black-and-white-films-colorized, 2011

[4] S. Moraes, "Wonders in image processing with machine learning," https://medium.com/@ODSC/wonders-in-image-processing-with-machine-learning-9c6f2e070e99, 2019

[5] T. Welsh, M. Ashikhmin, K. Mueller, "Transferring color to grayscale images," ACM Trans. on Graphics , vol. 21, no. 3, pp. 277-280, 2002

[6] A. Levin, D. Lischinski, Y. Weiss, "Colorization using optimization," ACM Trans. on Graphics, vol. 23, no. 3., pp. 689-694, 2004

[7] R. Irony, D. Cohen-Or, D. Lischinski, "Colorization by example," Proc. Eurographics Symposium on Rendering, pp. 201-210, 2005

[8] X. Liu, L. Wan, Y. Qu, T. T. Wong, S. Lin, C. S. Leung, P. A. Heng, "Intrinsic colorization," ACM Trans.on Graphics, vol. 27, no. 5, 2008

[9] G. Charpiat, M. Hofmann, B. Schölkopf, "Automatic image colorization via multimodal predictions," Computer Vision-ECCV 2008. Springer, pp. 126-139, 2008

[10] A. Y. S. Chia, S. Zhuo, R. K. Gupta, Y. W. Tai, S. Y. Cho, P. Tan, S. Lin, "Semantic colorization with internet images," ACM Trans. on Graphics, vol. 30, no. 6, 2011

[11] R. K. Gupta, A. Y.-S. Chia, D. Rajan, E. S. Ng, H. Zhiyong, "Image colorization using similar images," Proceedings of the 20th ACM international conference on Multimedia, ACM, pp. 369-378, 2012

[12] A. Deshpande, J. Rock, D. Forsyth, "Learning large-scale automatic image colorization," Proceedings of the IEEE International Conference on Computer Vision, pp. 567-575, 2015

[13] Z. Cheng, Q. Yang, B. Sheng, "Deep colorization," Proceedings of the IEEE International Conference on Computer Vision. pp. 415-423, 2015

[14] S. Iizuka, E. Simo-Serra, 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 (Proc. Of SIGGRAPH 2016), vol. 35, no. 4., 2016

[15] R. Zhang, P. Isola, A. A. Efros, "Colorful image colorization," European Conference on Computer Vision, pp. 649-666, 2016

[16] R. Dahl, "Automatic colorization," https://tinyclouds.org/colorize/, 2016

[17] G. Larsson, M. Maire, G. Shakhnarovich, "Learning representations for automatic colorization," ECCV, pp. 577-593, 2016

[18] A. Deshpande, J. Lu, M.-C. Yeh, M. J. Chong, D. Forsyth, "Learning diverse image colorization," ECCV, 2017

[19] S. Koo, "Automatic colorization with deep convolutional generative adversarial networks," Stanford University, Stanford, California, 2016

[20] Y. Cao, Z. Zhou, W. Zhang, Y. Yu, "Unsupervised diverse colorization via generative adversarial networks," Machine Learning and Knowledge Discovery in Databases. Lecture Notes in Computer Science, vol. 10534, Springer, Cham, pp. 151-166., 2017

[21] P. Vitoria, L. Raad, C. Ballester, "ChromaGAN: adversarial picture colorization with semantic class distribution," IEEE Winter Conference on Applications of Computer Vision, 2020

[22] W. Xian, P. Sangkloy, V. Agrawal, A. Raj, J. Lu, C. Fang, F. Yu, J. Hays, "TextureGAN: controlling deep image synthesis with texture patches," IEEE/CVF Conf. on Comp. Vision and Pattern Rec., 2018

[23] S. Yoo, H. Bahng, S. Chung, J. Lee, J. Chang, J. Choo, "Coloring with limited data: few-shot colorization via memory augmented networks," IEEE/CVF Conf. on Comp. Vision and Pattern Rec., 2019

[24] T.-H. Sun, C-H. Lai, S-K. Wong, Y.-S. Wang, "Adversarial colorization of icons based on structure and color conditions," Proceedings of the 27th ACM International Conference on Multimedia, pp. 683-691, 2019

[25] V. Manjunatha, M. Iyyer, J. Boyd-Graber, L. Davis, "Learning to color from language," Proceedings of the 2018 Conference of NAACL: Human Language Technologies, vol. 2, pp. 764-769, 2018

[26] E. Wallner, "Colorizing B&W photos with neural networks, https://blog.floydhub.com/colorizing-b-w-photos-with-neural-networks/, 2017

[27] I Žeger, Conversion of black and white pictures to coloured images using machine learning techniques, MSc thesis, Faculty of Electrical Engineering and Computing, Zagreb, 2020