HTCN: Harmonious Text Colorization Network for Visual-Textual Presentation Design

Xuyong Yang[1], Xiaobin Xu[2], Yaohong Huang[3], and Nenghai Yu[1(✉)]

[1] University of Science and Technology of China, Hefei, China
ynh@ustc.edu.cn
[2] Visnect Technology Co., Ltd., Shenzhen, China
[3] WeCar Technology Co., Ltd., Shenzhen, China
yaohong.huang@weicheche.cn

Abstract. The selection of text color is a time-consuming and important aspect in the designing of visual-textual presentation layout. In this paper, we propose a novel deep neural network architecture for predicting text color in the designing of visual-textual presentation layout. The proposed architecture consists of a text colorization network, a color harmony scoring network, and a text readability scoring network. The color harmony scoring network is learned by training with color theme data with aesthetic scores. The text readability scoring network is learned by training with design works. Finally, the text colorization network is designed to predict text colors by maximizing both color harmony and text readability, as well as learning from designer's choice of color. In addition, this paper conducts a comparison with other methods based on random generation, color theory rules or similar features search. Both quantitative and qualitative evaluation results demonstrate that the proposed method has better performance.

Keywords: Text colorization · Color harmonization · Text readability · Visual-textual presentation design

1 Introduction

In modern multimedia design, the proper combination of text and color can make the communication of information more diverse and efficient. Naturally, the need for text colorization is widespread in daily life, such as in the designing of advertisements, magazines, posters, signboards and webpages. However, it is often not easy to choose a suitable and aesthetically pleasing text color during the design process due to multiple reasons. Firstly, the number of colors is so large that it is difficult to enumerate all of them for the evaluation of coloring effectiveness. Secondly, to make the combination of selected text colors and scene colors achieve a harmonious and beautiful effect, the designer needs not only considerable professional experience, but also a certain number of trail-and-error coloring attempts. Thirdly, text colorization is different from image colorization, and it

H. Ma et al. (Eds.): PRCV 2021, LNCS 13020, pp. 560–571, 2021.
https://doi.org/10.1007/978-3-030-88007-1_46

is essential that the text is clear and easy to read in the position where it is placed. Therefore, the contrast between the text and its background should be considered. Finally, different colors can exert different effects on human psychology and emotions. Thus, in order to convey the right message, it is sometimes necessary to choose text colors that represent the right emotions and themes, such as energetic, sad, noble, etc.

The goal of this paper is to construct a model that automatically predicts harmonious, easy-to-read, and aesthetically pleasing text colors for overlapping visual-textual presentation design. Such a model makes text colorization easy, thus allowing professional designers to make a quick choice and the average person to obtain a result that satisfies basic aesthetics. Most traditional text coloring methods base on color-related design guidelines and theoretical models (e.g., [7,18]). However, their results are not optimal because theories and guidelines are often only general design guidance. Meanwhile, actual design work requires designers to make specific and detailed adjustments according to the situation.

In this paper, we propose a deep learning framework for text colorization. We model the color harmony between text and background images, the readability of text, as well as learn corresponding aesthetic knowledge from large scale design data. By considering the global color harmony of text, local text readability, and designer's choice of color, our proposed framework generates better text colorization results.

The main contributions of this paper are presented as follows:

1. We propose a global color harmony scoring network and a local text readability scoring network to measure the aesthetics of text color harmony and readability, respectively.
2. We put forward a deep learning framework HTCN (Harmonious Text Colorization Network) that integrates aesthetic color harmony and text readability, which can make good use of the inherent knowledge in big data for text colorization in visual-textual presentation design.
3. We construct a dataset called VTDSet (Visual-Textual Design Set) with 77,038 text colorization design samples and their corresponding background images, which can be capable of supporting more learning tasks.
4. Experiments demonstrate that the proposed network has better performance and can be easily applied to practical design tasks.

2 Related Works

Color Harmony. Early studies of color harmony concentrated on hue templates, describing harmonies as a theory of fixed rotational categories of the color wheel. Hue templates were employed in much of the art and design work, and were a key early theory to support color design. Itten proposed in [6] that equidistant sets of 2, 3, 4, and 6 hues on the color wheel are harmonious, and the color wheel has rotational invariance. A shortcoming of the hue templates is that they are defined independently of the color space. The source of the harmonious aesthetics dataset used in this paper, the Kuler website [1], uses the BYR color

wheel, while other websites such as COLOURlovers [2] adopt the RGB color wheel, which suggest different colors despite using the same template rules. The color themes provided by Matsuda summarize a set of classic color patterns with 8 hue distributions and 10 tone distributions [11], and this color harmony model has been used in several computer vision and graphics projects [4,16,18].

Additionally, there are also many studies on color harmony which are not limited to hue templates [12,13]. Although different approaches have been empirically conducted in different color spaces, similar laws have been summarized. Such theories are fundamentally based on the fact that people visually expect smoother effects. For example, the Munsell system suggests that color combinations in HSV color space with fixed H and V but changes in saturation S are harmonious [12]. It has also been shown that colors are harmonious if they can be connected in a straight line in space, such as in the Ostwald system where it is considered harmonious when the colors have the same amount of white or black [11], because such colors can form a straight line in the color space.

There are also some studies evaluating the harmony of color themes through a data-driven approach. O'Donovan et al. [14] proposed a LASSO regression model to fit the scores of color themes (combinations of five colors). In the work of [17], Yang et al. constructed a framework that first performs maximum likelihood estimation for a color pair composed of two colors and then predicts the aesthetic score of the color theme by adopting a BPNN network.

Text Readability. There are quite a few studies concerning text readability. In the study [10], it was shown that the two forms of contrast, luminance and chromaticity, are processed independently within the visual system. Since visual acuity responds much better to changes in luminance than to changes in hue and saturation, the contrast between the luminance of the color and the background dominates when we make attempts to resolve fine details. Albers et al. [3] suggested that differences in perception of readability are difficult to measure mathematically. Zuffi et al. [21] presented an experiment on the readability of colored text on colored backgrounds, which was conducted through crowdsourcing on the web. Their goal was to contribute to the understanding of how easy it is to read text on a display under general viewing conditions. Their experiments showed that the design of the text display should be set with a 30-unit difference in brightness between the CIE text and the background. In addition, the study [20] suggested using a light and soft background for color selection, and light-colored text on a dark background is more challenging to read.

3 Proposed Method

3.1 Overall Framework

We propose a deep learning approach, Harmonious Text Colorization Network (HTCN), defining the loss function through a global harmony scoring network and a local readability scoring network, which can well introduce a large amount

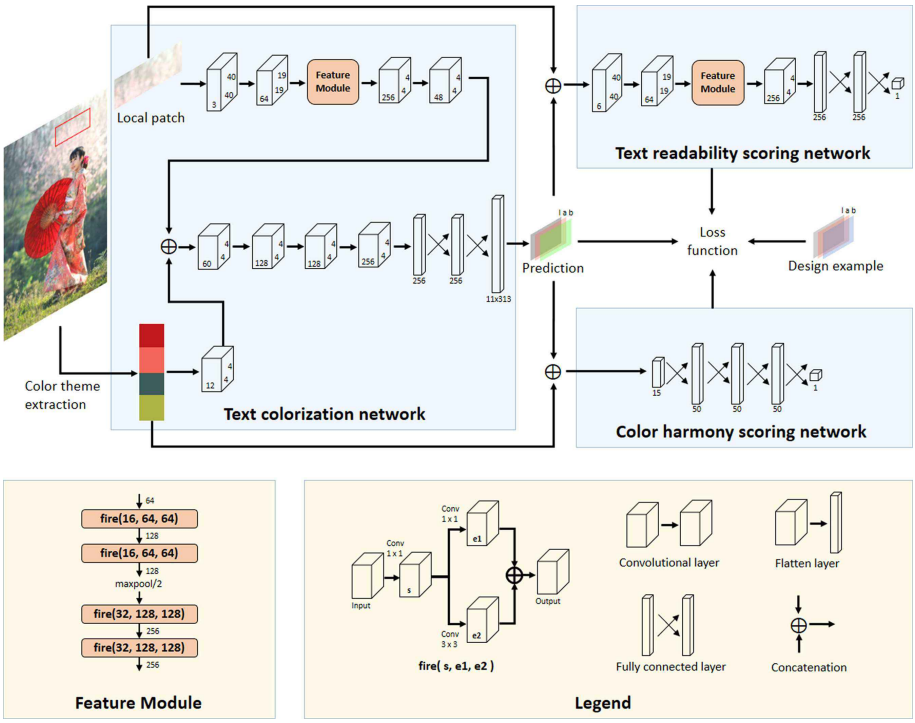


Fig. 1. HTCN network architecture.

of design experience and handle the multimodal problem of color prediction. The design of our network architecture is depicted in Fig. 1. The overall framework diagram shows that the proposed network framework provides an effective combination of global harmony and local readability as well as historical data experience. Besides, the whole network is end-to-end trainable.

In the proposed architecture, theme colors instead of the whole background image are employed as input for two reasons. One is that evaluating harmonious properties between text color and the background image is easier by abstracting the background image into theme colors and utilizing public theme color harmony datasets. The other reason is to make the HTCN architecture more generally applicable. For example, regarding signboard design or interior design, there is no well defined background image. However, theme colors can be extracted from the environment and our proposed architecture can be applied without compromise.

3.2 Text Colorization Network

The input of the text colorization network is the local background patch covered by the text and the global theme colors to be matched. The local background image is first passed through a local image feature encoding part to obtain the local features, which is concatenated with the theme color features and then

passed through the color prediction part to acquire the predicted text color. The local image patch is resized to 40×40, and the theme colors can be calculated using color theme extraction methods [8,9]. In this paper, the K-Means algorithm is employed to cluster 4 main theme colors for the sake of simplicity. The structure of the text colorization network is shown in Fig. 1. To reduce parameters while preserving accuracy, we adopt the Fire module proposed in the SqueezeNet architecture [5] in the feature module of the local image feature encoding part. Based on the multimodal nature of the color prediction problem, we choose to use classification rather than regression for colorization. We separate hue and lightness by transforming the color into Lab color space, and quantize the L component and ab component separately. The ab component is quantized into 313 categories similar to Zhang et al. [19]. In addition, the L component is uniformly devided into 11 categories.

3.3 Text Readability Scoring Network

The text readability scoring network aims to predict the readability score given certain text colorization. The input is the concatenation of the text color and the local background patch covered by the text. Besides, after the first fully connected layer, we employ the dropout regularization technique to avoid overfitting, with a dropout ratio of 50%.

The positive examples used for training include text colors extracted from design works and the corresponding local background images. We synthesize negative examples based on studies in association with readability. Specifically, we use the K-Means algorithm to cluster the ab values of the local background image and randomly select a cluster center as the ab value of the text color of negative examples. According to the results of text readability experiments by Zuffi et al. [21], we randomly selected values that differed from the median of the luminance of the local background image by 30 or less as the L values of the negative examples.

3.4 Color Harmony Scoring Network

The proposed color harmony scoring network is a fully connected network as shown in Fig. 1. The second and third fully connected layers are followed by a dropout layer with a ratio of 20% to prevent overfitting. The network takes the text color and the background theme colors as input and predicts harmony rating for the color combination.

3.5 Loss Function

The optimal text color C^* should satisfy the following constraints: (1) color harmony with the color theme of the whole image, which ensures a relatively high combined aesthetic score; (2) optimal contrast between the text color and the local image area, which ensures high readability of the text. This can be modeled

as maximizing an scoring function that includes two components, respectively, the global color harmony $\mathcal{H}(C,T)$, and local text readability $\mathcal{D}(C,I_{local})$.

$$C^* = \arg\max_C(\mathcal{D}(C,I_{local}) + \beta\,\mathcal{H}(C,T)) \tag{1}$$

where C is the text color, I_{local} is the local background image of the text block, and T is the global color theme.

Since there is no exact mathematical expression for color harmony and text readability, we design a color harmony scoring network and a text readability scoring network to model them in the energy function. Before training the text colorization network, we first trained the color harmony scoring network and the text readability scoring network. Then, we fix the parameters of these two networks before the overall training of the text colorization network. It can be noticed that when the local background is complex, it is difficult to find text colors with high harmony as well as high readability ratings. In this situation, the color C^{ref} used by the designer in the actual design work is an proper choice with comprehensive aesthetic considerations. As a result, we design the network to learn from this choice. Combining the above considerations, we design the scoring function as follows:

$$\mathcal{R}(C,C^{ref},I_{local},T) = \mathcal{D}(C,I_{local}) + \beta\,\mathcal{H}(C,T) + \alpha\ \mathrm{G}_\sigma\left(\|C - C^{ref}\|_2\right) \tag{2}$$

where α and β are weights for controlling the balance among different factors and G_σ is a Gaussian smoothing function with standard deviation σ, $\mathrm{G}_\sigma(x) = \exp(-\frac{x^2}{2\sigma^2})$.

To solve the problem of gradient transfer and use the calculation results of the loss function in order to effectively update the parameters of the text colorization network, we propose to adopt the entire predicted color classification probability distribution for loss calculation. Let the color distribution best reflecting the combined ranking of Eq. (2) on that text block be noted as p. We want the predicted distribution $\hat{p}$ to approximate p, and thus we minimize their cross entropy. The total loss is defined as follows:

$$\mathcal{L} = -\sum_i \mathrm{Softmax}(\mathcal{R}(C_i,C^{ref},I_{local},T))\ \log(\hat{p}_i) \tag{3}$$

where the softmax function maps the overall score $\mathcal{R}(C_i,C^{ref},I_{local},T)$ to probability of text color, $\mathrm{Softmax}(x_i) = \frac{e^{x_i}}{\sum_i e^{x_i}}$.

4 Experiments

4.1 Datasets

Color Combination Aesthetics Score Dataset. We obtained the MTurk public dataset from [14], which consists of 10,743 carefully selected color themes created by users on Adobe Kuler [1], covering a wide range of highly and poorly

rated color themes, each of which rated by at least 3 random users with ratings between 1 and 5. The MTurk dataset uses Amazon Mechanical Turk[1] to collect more user ratings for the selected topics, making each topic rated by 40 users. Finally, the average score for each topic was taken as the final score.

Visual-Textual Design Works Dataset. We constructed a visual-textual design dataset called VTDSet (Visual-Textual Design Set) where 10 designers selected text colors in 5 to 7 areas on each of the total 1226 images, resulting in 77,038 designed text colors and their corresponding information. We randomly selected 10,000 design results associated with 1000 background images from the dataset as the training dataset, and 2260 design results associated with the remaining 226 background images as the testing dataset.

4.2 Implementation Details

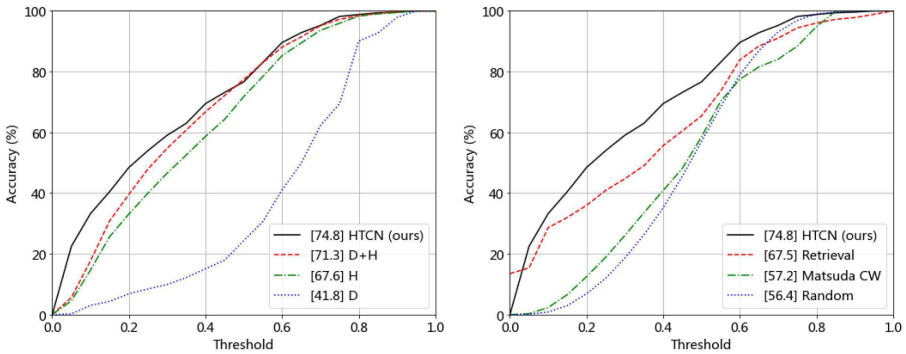
Regarding the color harmony scoring network, we transformed the colors in the color theme into the normalized Lab color space and randomly selected 70% of the samples in the aesthetic score dataset for training with the remaining 30% as the test set. We used the Adam algorithm for optimization with a learning rate of 0.001, and $\beta_1 = 0.9$, $\beta_2 = 0.999$, and L2 regularization coefficient of 0.0005. An early termination strategy was used to avoid overfitting, and a total of 300 epochs were trained with batch size of 64. For the text readability scoring network, we employed the SGD algorithm for parameter updating with momentum of 0.9 and L2 regularization coefficient of 0.0005. We adopted a multi-step learning rate with an initial learning rate of 0.001. At both the 600th and 800th epochs, the learning rate decays by 0.1. In addition, we used an early termination strategy to avoid overfitting, and trained for a total of 1000 cycles with batch size of 8.

When training the text colorization network, we fix the parameters of the color harmony scoring network and the text readability scoring network. We set $\alpha = 0.5$, $\beta = 1$ and $\sigma = 4$. The theme color of the background image and the text color and corresponding local patches of all 5 to 7 text regions in the same design image are extracted from the design results and combined into a batch for training. We adopted the Adam algorithm for parameter updating with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and the L2 regularization coefficient of 0.0005. We used a learning rate of 0.001 and an early termination strategy to avoid overfitting, and trained for a total of 30 epochs.

4.3 Effectiveness of Network Design

In order to verify the effectiveness of the proposed HTCN framework, we disassembled the proposed network structure and used only a part of it for colorization. The results obtained on the test set are shown in Fig. 2. The accuracy in the figure is obtained by calculating the Euclidean distance between the predicted color and the color chosen by the designer. To facilitate the calculation,

[1] https://www.mturk.com.



(a) Different architectures (b) Different methods

Fig. 2. Comparison of accuracy of text colorization (a) D denotes using only the local readability evaluation network, H denotes using only the global harmony evaluation network, and D+H denotes using both the local readability and global harmony evaluation networks. (b) Comparison with other methods.

we used the RGB color space and normalized the distance. As presented in the legend, we report the average accuracy of thresholds ranging from 0.05 to 1, with an interval of 0.05. According to the figure, the organic combination of local readability and global harmony can effectively improve the performance of colorization, which is in line with the designer's guidelines for selecting colors when designing, i.e., making adjustments to ensure that text is clear and easy to read while ensuring that the text color is in harmony with the background image. In addition, the complete HTCN network obtains the best colorization results with an average accuracy of 74.8% because it also learns the aesthetics implicit in the designer's design choices in difficult scenarios under the condition of ensuring local readability and global harmony.

4.4 Comparison with Other Methods

We compare the text colorization network HTCN proposed in this paper with the following three approaches:

Random Text Colorization ("Random"). A random value is selected in the RGB color space, and this baseline is used to check whether the color design of the text in the generation of the visual-textual presentation layout is arbitrary.

Text Colorization Based on Matsuda Color Wheel Theory ("Matsuda CW"). This text colorization method bases on the color wheel theory, which is also adopted in the work of Yang et al. [18]. We reproduce the method by first performing principal component analysis on the image to obtain the color theme, taking the color with the largest proportion as the base color C_d of the image, and then calculating the minimum harmonic color wheel distance between the base color C_d and the aesthetic template color set according to the constraint

defined by Matsuda to obtain the optimal hue value of the text color C_r. Finally, the color mean $\mu_{h,s,v}$ of the image covered by the text area is calculated, and the optimal text color is obtained by reasonably maximizing the distance between $\mu_{h,s,v}$ and C_r in the (s, v) saturation and luminance space.

Text Colorization Based on Image Feature Retrieval ("Retrieval"). Retrieval-based strategy is frequently used in design, i.e., seeking reference among solutions of similar problems. For the text colorization problem, the original designer's color can become the recommended color when the background image and the text area are similar. As a result, we concatenate the global features of the image and the local image features of the text-covered region to obtain the K nearest neighbor recommendations for the current text coloring by the cosine distance. We used the VGG-16 network [15] pretrained on the ImageNet dataset, and selected the output of the fc6 layer as the image features. The combined feature of the text region image I_{text} on the global image I is $f = <VGG_I, VGG_{I_{text}}>$. The text color corresponding to the feature with greatest similarity in the design library is selected for colorization.

We implemented these three methods along with HTCN and compared the results on the test set of the dataset VTDSet. Figure 2 shows the accuracy curves of the HTCN network and the three methods which are compared. They are calculated in the same way as in Sect. 4.3. It can be seen that the HTCN text colorization network has obvious superiority over other methods. Typical results of each baseline method and the algorithm in this paper are shown in Fig. 3. We can intuitively observe the difference between various methods. The results of the random method are the most unsecured, while the HTCN text colorization network proposed in this paper can obtain some results that are different from the designer's work but still have good performance. In the color theory-based method, it can be found that all the text colors in the same design are close to a similar hue value, which has a very large limitation of color selection. The retrieval based approach gives historical design results. However, it is not sufficient for creativity and novelty, especially its quality depends on the quality of design examples in the database.

4.5 User Study

We performed user study for qualitative evaluation of the proposed method. Totally 20 persons aged 21 to 35 were recruited, among which eleven were females, and twelve had design-related working experience. Text colorization results of 20 design works randomly selected from the test dataset were used in the evaluation. In the readability evaluation, we let the user read the letters on the picture and kept the user's eyes 60 cm away from the screen. The images displayed on the screen were resized to the height of 15 cm.

(1) Global color harmony evaluation: participants assess the harmony of the text color with overall image in each text box, scoring from 1 to 5, respectively, as ugly, discordant, average, harmonious, and very aesthetically pleasing.

Fig. 3. Comparison of the actual effect of text colorization under various algorithms: (a) random generation of text colors, (b) method based on the Matsuda color wheel theory, (c) retrieval-based method that directly obtains corresponding color recommendations from historically similar design examples, (d) the HTCN network proposed in this paper, and (e) is the designer's original work.

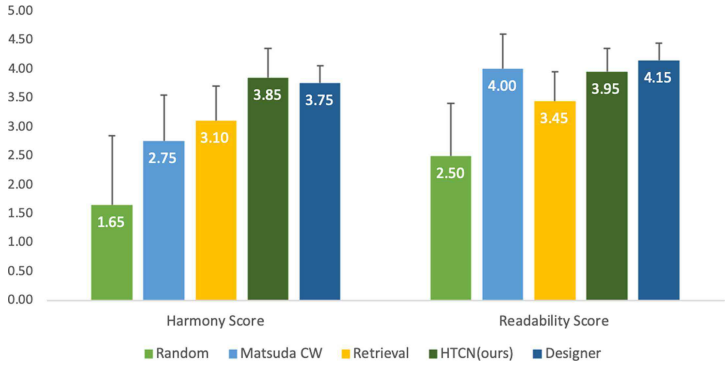


Fig. 4. Harmony and readability scores of text color with different methods

(2) Text readability evaluation: participants assess the readability of the text for each text box, scoring from 1 to 5, which are completely unreadable, partially readable, readable requiring concentration, readable with relative ease, and readable at a glance.

Figure 4 shows that the proposed method can obtain similar aesthetic harmony as the designer. In some cases, the harmony score is better than the designer, because the text color design problem is multimodal and there may exist not only one optimal color combination. By learning from a large number of design works and modeling the aesthetic metric of color themes, our method can acquire design knowledge that are not available in traditional methods, and obtain more harmonious and design-oriented text colorization results.

Obviously, the randomly generated results are most likely to produce crossover with the background because there are no basic aesthetic constraints, making it the worst readability choice and fully illustrating the necessity of other methods. The interesting finding is that the readability of the theoretically derived text color is little better than our method. The reason is that in guaranteeing readability, by explicitly stretching the saturation and the brightness difference between image and background color, a better readability can be obtained through the rule constraint while the overall color harmony is relatively lacking.

5 Conclusion

To conclude, this study introduces a deep learning framework to learn inherent aesthetic design knowledge beyond design rules from color annotation data. We propose two scoring networks that can learn well the harmony metric of color and readability metric of text in aesthetic design. By exploring the text colorization problem, we put forward a deep neural network architecture HTCN for automatic text colorization in designing of visual-textual presentation layout. Moreover, the experiments reveal that our algorithm can perform better than other methods.

References

1. Adobe Kuler. https://color.adobe.com. Accessed 19 Mar 2021
2. Colourlovers. http://www.colourlovers.com. Accessed 21 Mar 2021
3. Albers, J.: Interaction of Color. Yale University Press, London (2013)
4. Cohen-Or, D., Sorkine, O., Gal, R., Leyvand, T., Xu, Y.Q.: Color harmonization. In: ACM SIGGRAPH 2006 Papers, pp. 624–630 (2006)
5. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K.: SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. arXiv preprint arXiv:1602.07360 (2016)
6. Itten, J.: The Art of Color: The Subjective Experience and Objective Rationale of Color. Translated by Ernst Van Haagen. Van Nostrand Reinhold (1973)
7. Jahanian, A., et al.: Recommendation system for automatic design of magazine covers. In: Proceedings of the 2013 International Conference on Intelligent User Interfaces, pp. 95–106 (2013)
8. Jahanian, A., Vishwanathan, S., Allebach, J.P.: Autonomous color theme extraction from images using saliency. In: Imaging and Multimedia Analytics in a Web and Mobile World 2015, vol. 9408, p. 940807 (2015)
9. Lin, S., Hanrahan, P.: Modeling how people extract color themes from images. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 3101–3110 (2013)
10. MacIntyre, B.: A constraint-based approach to dynamic colour management for windowing interfaces. Master's thesis, University of Waterloo (1991)
11. Matsuda, Y.: Color design. Asakura Shoten **2**(4), 10 (1995)
12. Munsell, A.H., Cleland, T.M.: A grammar of color: arrangements of Strathmore papers in a variety of printed color combinations according to the Munsell color system. Strathmore Paper Company (1921)
13. Nemcsics, A.: Coloroid colour system. Hungarian Electronic Journal of Sciences, HEJ Manuscript no.: ARC-030520-A (2003)
14. O'Donovan, P., Agarwala, A., Hertzmann, A.: Color compatibility from large datasets. In: ACM SIGGRAPH 2011 papers, pp. 1–12 (2011)
15. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
16. Tokumaru, M., Muranaka, N., Imanishi, S.: Color design support system considering color harmony. In: 2002 IEEE World Congress on Computational Intelligence. 2002 IEEE International Conference on Fuzzy Systems. FUZZ-IEEE 2002. Proceedings (Cat. No. 02CH37291), vol. 1, pp. 378–383. IEEE (2002)
17. Yang, B., et al.: A color-pair based approach for accurate color harmony estimation. In: Computer Graphics Forum, vol. 38, pp. 481–490. Wiley Online Library (2019)
18. Yang, X., Mei, T., Xu, Y.Q., Rui, Y., Li, S.: Automatic generation of visual-textual presentation layout. ACM Trans. Multimed. Comput. Commun. Appl. (TOMM) **12**(2), 1–22 (2016)
19. Zhang, R., Isola, P., Efros, A.A.: Colorful image colorization. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 649–666. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46487-9_40
20. Zuffi, S., Brambilla, C., Beretta, G., Scala, P.: Human computer interaction: legibility and contrast. In: 14th International Conference on Image Analysis and Processing (ICIAP 2007), pp. 241–246. IEEE (2007)
21. Zuffi, S., Brambilla, C., Beretta, G.B., Scala, P.: Understanding the readability of colored text by crowd-sourcing on the web. HP Laboratories (2009)