INVERSE HALFTONE COLORIZATION: MAKING HALFTONE PRINTS COLOR PHOTOS

Yu-Ting Yen, Chia-Chi Cheng, Wei-Chen Chiu

Department of Computer Science, National Chiao Tung University

ABSTRACT

We propose to address the *Inverse Halftone Colorization* task, which tries to recover colorful images from black and white halftone prints, and can be treated as the joint problem of inverse halftone and colorization. Although the inverse halftone colorization seems to be achievable via applying inverse halftone followed by colorization, our proposed method advances from two perspectives: (1) we empirically discover that the orders of cascading inverse halftone and colorization (i.e first inverse halftone then colorization versus the reverse order) would lead to results with different properties, hence a fusion scheme is proposed to integrate their results; (2) we introduce several novel losses to encourage the realness, diversity, and the structural coherence of the colorization. Moreover, our model is flexible to support both exemplar-based and random colorization. We conduct extensive experiments to demonstrate the efficacy of our method as well as verify the contributions of our design choices.

Index Terms— Inverse halftone, Colorization

1. INTRODUCTION

As lots of photographs on printed matter from early years are based on the halftone printing, when we nowadays would like to digitally archive or further manipulate these historical photos obtained from the printed products, the "*inverse halftone*" process of recovering the continuous-tone images from halftone prints becomes a practical but difficult problem due to the loss of details in the halftone images. Numerous methods addressing the inverse halftone problem have been proposed, from the classicial approaches (e.g. based on filtering [1, 2], look-up-tables [3], or dictionary learning [4]) to recent deep-learning-based ones [5, 6, 7, 8].

Another widely discussed task on archiving historical or old images is "*colorization*", where the grayscale images are colorized for making them more visually appealing. Colorization is also challenging since its nature of being a ill-posed problem, in which a single gray-level image can be mapped into multiple colorful appearance. There already exists different methods of colorization, in which they can be roughly categorized into two groups by their ways of providing sources of color appearance: the passive ways (e.g. drawing color latent from random noise [9, 10, 11, 12]) and the active ones (e.g. user-guided scribbles [13, 14] or color reference images [15, 16, 17, 18]. When some of the passive methods suffer from the lack of diversity (e.g. [10, 11]) or are easily prone to produce visual artifacts such as color bleeding and color washout (e.g. [12]), the active ones however require users to carefully place the scribbles or choose a suitable reference image (e.g. with similar content as the input image) in order to achieve convincing results, thus making them impractical for large scale or general applications.

In this paper we propose to tackle the problem of jointly performing inverse halftone and colorization, as named as "*inverse halftone colorization*", which is a holistic procedure of generating colorful continuous-tone images from the given black-and-white halftone prints, as some examples shown in Figure 2. To the best of our knowledge, we are the first of its kind to discuss such joint problem, although there already exists quite some prior works on either inverse halftone or colorization. In particular, our proposed method is deep-learning-based and equipped with several novelties: Firstly, while we are able to obtain the primary results of inverse halftone colorization via straightforward combination of inverse halftone and colorization approaches, we empirically discover that the two opposite orders of cascading them (i.e. inverse halftone followed by colorization, and colorization followed by inverse halftone) will end up with distinct outputs (e.g. having artifacts on different image regions, see Figure 3). We therefore propose a fusion network to simultaneously take these two cascade orders into consideration for producing the fused results with better image quality; Secondly, with respect to previous colorization works, we not only make the colorization part of our proposed method to support different sources of color information (e.g. through random noise or color references) but also propose various objectives to better improve the quality of colorization, including the similarity loss for improving the colorization coherence among the pixels of the same semantic class, the histogram loss for encouraging the consistency between color distributions of the colorized output and its corresponding color reference image, as well as the mode seeking loss for ensuring the diversity on the colorization results. The details of our proposed method are provided in the following sections with the extensive experiments for showing our contributions and superior performance in comparison to several baselines. Our code is available at https://github.com/ccc870206/InverseHalftoneColorization.

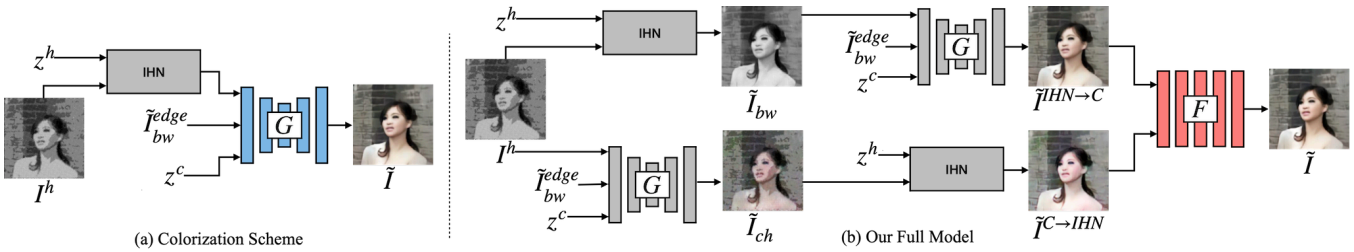

Fig. 1: Illustration of (a) the colorization scheme and (b) our full model with fusion network for the task of *inverse halftone colorization* task. Please note that we shade each sub-network in different colors, where the gray-shaded ones are pretrained and fixed, while others are learnable in our full model training procedure.

2. PROPOSED METHOD

Our goal is to transfer the bi-level (black and white) halftone image $I_h \in \mathbb{R}^{H\times W\times 1}$ into a colorful continuous-tone image $\tilde{I} \in \mathbb{R}^{H\times W\times 3}$, where the proposed method consists of three main components: the *inverse halftone network* (denoted as IHN), the *colorization network*, and the *fusion network*, as shown in Figure 1. We detail our model designs as follows.

2.1. Inverse Halftone Network (IHN)

Our inverse halftone network (IHN) aims to recover the continuous-tone image $\tilde{I}_{bw} \in \mathbb{R}^{H\times W\times 1}$ from a given halftone image I_h (we assume that I_h is obtained by first turning a color image I into grayscale I_{bw} then performing halftoning). Our IHN architecture is extended from the state-of-the-art PRL model [6] where its content aggregation for synthesizing the global tone is replaced by a 7-layers U-Net to better support the analog halftoning (i.e. the target type of halftoning in this paper and being used in most of the historical images). Moreover, the input of our content aggregation module now additionally takes a random noise map $z^h \in \mathbb{R}^{H\times W\times 8}$, which is drawn from $\mathcal{N}(0, I)$ and smoothed by a Gaussian kernel, for helping synthesize the information loss between the halftone image and its corresponding continuous-tone one. Our IHN achieves the better performance of inverse halftone comparing to the original PRL model [5] in terms of PSNR (25.77 versus 23.95 on our experiment datasets). The objective for training IHN includes the pixel loss $\mathcal{L}_{pixel}$ and the perceptual loss $\mathcal{L}_{percep}$:

$$\mathcal{L}_{pixel} = \sum \lambda_1 \|\tilde{I}'_{bw} - I_{bw}\|_2 + \lambda_2 |\tilde{I}_{bw} - I_{bw}|$$
$$\mathcal{L}_{percep} = \sum \| \Phi^{conv4\_4}(\tilde{I}_{bw}) - \Phi^{conv4\_4}(I_{bw})\|_2$$

where $\tilde{I}'_{bw}$ is the output of the content aggregation module, $\Phi^{conv4\_4}(\cdot)$ denotes the feature extracted from $conv4\_4$ layer of the ImageNet-pretrained VGG network, and λ_1 and λ_2 here are set to 5 and 1 respectively.

2.2. Colorization Network

Given the grayscale image $\tilde{I}_{bw}$, our colorization network G as shown in Figure 1(b) takes $\tilde{I}_{bw}$, the edge map $\tilde{I}^{edge}_{bw}$ of $\tilde{I}_{bw}$, and the color latent code z^c as input to produce the colorization result of $\tilde{I}_{bw}$. In particular, the edge map here is leveraged to help G better maintain the structure consistency. As our aiming to support both exemplar-based and random colorization, we propose to have three different training schemes:

Recovery Scheme. Ideally, when we colorize an inverse halftone image $\tilde{I}_{bw} = IHN(I_h, z^h)$ by the color appearance of its corresponding groundtruth color image I, the output $\tilde{I}_{rec}$ of G should well recover I, where the color appearance of I is extracted by an encoder E into a color latent code z^c_{rec} and $\tilde{I}_{rec} = G(\tilde{I}_{bw}, \tilde{I}^{edge}_{bw}, z^c_{rec})$.

Reference Scheme. For enabling G to produce more diverse colorization results, we advance to utilize different color reference images other than I. In other words, G colorizes $\tilde{I}_{bw}$ by the color appearance of an arbitrary color reference image I_{ref} to output the result $\tilde{I}_{ref} = G(\tilde{I}_{bw}, \tilde{I}^{edge}_{bw}, z^c_{ref})$, where $z^c_{ref} = E(I_{ref})$. As there is no groundtruth to evaluate $\tilde{I}_{ref}$, we take the colorization result $I_{example}$ produced by [15], the state-of-the-art exemplar-based colorization method, as our desired output for $\tilde{I}_{ref}$.

Random Scheme. In comparison to the recovery and reference schemes which need the color reference images, here we also aim to enable the random/automatic colorization. To this end, the latent color code z^c_{rand} now is randomly drawn from $\mathcal{N}(0, I)$, where G is supposed to provide different colorization results $\tilde{I}_{rand}$ with different latent codes.

We then propose to adopt various objectives for learning the colorization network G and the encoder E in the aforementioned three training schemes, as detailed below.

Reconstruction Loss $\mathcal{L}_{rec}$. We evaluate the reconstruction error in pixel-level on $\{\tilde{I}_{rec}, \tilde{I}_{ref}\}$ with respect to their corresponding groundtruths, i.e. $\{I, I_{example}\}$.

$$\mathcal{L}_{rec} = |\tilde{I}_{rec} - I| + \|\tilde{I}_{ref} - I_{example}\|_2$$

Structure Loss $\mathcal{L}_{struct}$. While there is no groundtruth for $\tilde{I}_{rand}$, we simply adopt the edge loss computed on the edge maps (where the color informtion is discarded) and the perceptual loss [19] (which is believed to be more related to the structural similarity) to encourage it to have the similar structure as I. With denoting ∇ for edge map computation,

$$\mathcal{L}_{struct} = |\nabla\tilde{I}_{rand} - \nabla I| + \|\Phi^{relu4\_1}(\tilde{I}_{rand}) - \Phi^{relu4\_1}(I)\|_2$$

Style Loss $\mathcal{L}_{style}$. We use the high-level style loss as in [9] to ensure that $\{\tilde{I}_{rec}, \tilde{I}_{ref}\}$ have the similar color appearance as

their corresponding color references $\{I, I_{ref}\}$.

$$\mathcal{L}_{style} = \sum \|g(\Phi^{relu4\_1}(\tilde{I}_{rec})) - g(\Phi^{relu4\_1}(I))\|_2 + \|g(\Phi^{relu4\_1}(\tilde{I}_{ref})) - g(\Phi^{relu4\_1}(I_{ref}))\|_2,$$

where g denotes the Gram matrix computation.

Latent Loss $\mathcal{L}_{latent}$. To enhance the joint training of E and G, we base on the constraint to define the latent loss that, the color latent codes extracted from $\{\tilde{I}_{ref}, \tilde{I}_{rand}\}$ by E should be identical to the ones used for their colorizations, therefore:

$$\mathcal{L}_{latent} = \sum \|z^c_{ref} - \hat{z}^c_{ref}\|_2 + \|z^c_{rand} - \hat{z}^c_{rand}\|_2.$$

where $\hat{z}^c_{ref} = E(\tilde{I}_{ref})$ and $\hat{z}^c_{rand} = E(\tilde{I}_{rand})$.

KL-Divergence Loss $\mathcal{L}_{KL}$. The latent distribution of color images encoded by E is encouraged to be close to a standard Gaussian $\mathcal{N}(0, I)$ by KL-divergence $\mathcal{D}_{KL}$, in order to enable the sampling of color latent code for random colorization.

$$\mathcal{L}_{KL} = \mathbb{E}[\mathcal{D}_{KL}(\{E(I), E(I_{ref})\}||\mathcal{N}(0, I))]$$

Adversarial Loss $\mathcal{L}_{adv}$. We utilize the adversarial learning on the colorization results to improve their realness, where the discriminator D considers the joint distribution of the color images and their corresponding color latent codes. We skip the detailed formulation of $\mathcal{L}_{adv}$ here as for the simplicity.

Similarity Loss $\mathcal{L}_{sim}$. We introduce the similarity loss $\mathcal{L}_{sim}$ which compares the difference between the colorization results and the corresponding real image in terms of their patch-wise self-similarity matrix $\mathcal{S}$ (computed by the cosine similarity between each pair of image patches), where such objective highly encourages the better coherence on the appearance for pixels of the same semantic class and more realistic results.

$$\mathcal{L}_{sim} = \|\mathcal{S}(\tilde{I}_{rec}) - \mathcal{S}(I)\|_2 + \|\mathcal{S}(\tilde{I}_{ref}) - \mathcal{S}(I)\|_2 + \|\mathcal{S}(\tilde{I}_{rand}) - \mathcal{S}(I)\|_2.$$

where N set to 4096 is the number of patches in an image.

Histogram Loss $\mathcal{L}_{hist}$. In addition to the high-level texture similarity as the style loss, we introduce the histogram loss $\mathcal{L}_{hist}$ to consider the statistical similarity in terms of color histograms between the colorization results and their color references. We explicitly propose a histogram estimator $H(I)$ which is pretrained to regress from the image I to the probability value of each bin in the color histogram. With denoting BCE for binary cross-entropy computation,

$$\mathcal{L}_{hist} = BCE(H(I), H(\tilde{I}_{rec})) + BCE(H(I_{ref}), H(\tilde{I}_{ref}))$$

Mode-Seeking Loss $\mathcal{L}_{ms}$. In order to prevent the model training from mode collapse and enrich the diversity in the colorization results, the mode seeking loss [20] is used in our setting, which encourages a minor change in the color latent code to have a significant impact on the colorization results:

$$\mathcal{L}_{ms} = \max_G \left(\frac{d_g(G(\tilde{I}_{bw}, \tilde{I}^{edge}_{bw}, z^c_{r1}), G(\tilde{I}_{bw}, \tilde{I}^{edge}_{bw}, z^c_{r2}))}{d_z(z^c_{r1}, z^c_{r2})}\right)$$

where z^c_{r1} and z^c_{r2} are two different noises $\sim \mathcal{N}(0, I)$; d_z denotes the average L_1 distance; and d_g computes the average L_1 distance between the Gram matrices of image features.

The weighted-sum of the aforementioned objectives (where the weights as hyper-parameters simply help to balance the numerical ranges of each loss) is used to jointly train the colorization network G and the color encoder E.

2.3. Fusion Network

After having IHN and colorization network well-trained, ideally the inverse halftone colorization can be achieved by either sequentially applying IHN and colorization network (denoted as path $IHN \rightarrow C$) or the reverse order (denoted as path $C \rightarrow IHN$, where the IHN is applied on each color channel). However, we observe that the colorization results obtained by these two orders, i.e. $\tilde{I}^{IHN \rightarrow C}$ and $\tilde{I}^{C \rightarrow IHN}$, have distinct properties, where the former shows vivid colors but with artifacts such as ripples, while the latter looks like an oil painting but retains obvious borders, as shown in Figure 3. We therefore design a fusion network F composed of multiple residual blocks to integrate these two results to get the final output $\tilde{I}$ of inverse halftone colorization, where F helps enhance the image details, remove artifacts, and correct the color tone. The implementation details are skipped here due to the space limit, we will release all the source codes and the trained model after paper acceptance.

3. EXPERIMENTS

Datasets and Metrics. We adopt Helen dataset [22] with 2000 and 330 images for training and testing respectively, where each image I can be transformed into the bi-level halftone one I_h by applying analog halftoning. The color reference images are based on the training set of Helen and other 6000 images from 100 categories of ImageNet. Each training image is paired with 6 color reference images to produce $I_{example}$ via [15] for training the reference scheme. For evaluation, we adopt the PSNR and SSIM metrics for the recovery scheme, the diversity metric [9] based on the average LPIPS distance between 1900 pairs of colorizations from random scheme, and the Fréchet Inception Distance (FID) score [23] for the overall performance of inverse halftone colorization (between 2970 results produced from all three colorization schemes and the real imges).

Qualitative Results. First, we provide in Figure 2 some example results of the inverse halftone colorization produced by our proposed method. Second, as there exists no prior work of inverse halftone colorization, we take the direct cascade of our IHN with some diverse colorization methods (e.g. [9, 11, 12, 21]) as the baselines, with qualitative results shown in the left-hand part of Figure 4. We observe that these baselines could produce artifacts (IHN+[9]), scattered color blocks (IHN+[11]), uneven color and diffused contour (IHN+[12]), and low diversity with grayish color (IHN+[21]). In comparison, our proposed method outper-

I_h I_{ref-1} $\tilde{I}_{ref-1}$ I_{ref-2} $\tilde{I}_{ref-2}$ $\tilde{I}_{rand-1}$ $\tilde{I}_{rand-2}$

Fig. 2: Given black and white halftone prints I_h, our method is able to generate colorful continuous-tone images, in which it supports exemplar-based colorization ($\tilde{I}_{ref-1}$ and $\tilde{I}_{ref-2}$) based on the color references (I_{ref-1} and I_{ref-2} respectively) as well as the random colorization ($\tilde{I}_{rand-1}$ and $\tilde{I}_{rand-2}$).

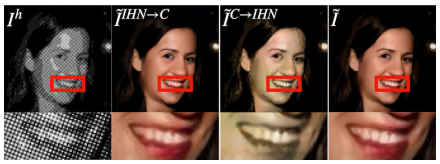

Fig. 3: Our fusion network F combines the different properties of both paths $IHN \to C$ and $C \to IHN$ to achieve better results of inverse halftone colorization (best viewed on screen with zoom and by color).

I_h IHN+[9] IHN+[11] IHN+[12] IHN+[21] Ours

Method	Diversity↑	FID↓
IHN+[9]	0.134	103.00
IHN+[11]	**0.162**	166.95
IHN+[12]	0.150	102.57
IHN+[21]	0.051	95.04
Ours	0.120	**82.40**

Fig. 4: The qualitative results (left-hand part) and the quantitative comparison (right-hand part) of diverse colorization.

forms these baselines as well as generates both diverse and realistic results of inverse halftone colorization. Please note that here we skip the comparison with respect to the baseline of cascade of IHN with the exemplar-based colorization method (e.g. [15]) as it typically needs the color reference image which has similar patches as the ones in the input image for achieving good colorization results, where such requirement is a crucial issue and brings additional cost. Our method instead encodes the whole color reference into the color latent code therefore does not suffer from such requirement.

Quantitative Results. The table on the right-hand part of Figure 4 shows quantitative results of the baselines. We use both the diversity and FID scores as our evaluation metrics here. Please note that for our model the random colorization is adopted for FID evaluation in order to have fair comparison with respect to the baselines. Overall, our proposed method achieves the best performance in terms of FID, which is much lower than the most competitive baseline IHN+[21]. When it comes to the diversity score, our work is superior to IHN+[21], and slightly worse than IHN+[12]/[11]/[9]. However, as the trade-off between the realness and diversity, we see from qualitative samples of Figure 4 that IHN+[12]/[11]/[9] generates unrealistic image with bad quality. Overall, our proposed method results in better balance between the realness and the diversity of the inverse halftone colorization.

Ablation Study. We conduct an ablation study here: the experiment start from the *Base* model, which only contains IHN, the recovery and random scheme, without having $\mathcal{L}_{ms}$, $\mathcal{L}_{sim}$, and $\mathcal{L}_{hist}$. And we sequentially include the reference scheme, $\mathcal{L}_{ms}$, $\mathcal{L}_{sim}$ and $\mathcal{L}_{hist}$ back to the Base model for studying their respective contributions. The results shown in Table 1, where we observe that: the Base model almost degenerates to have no diverse colorization (i.e. diversity-score close to zero), while adding the reference scheme is able to produce diverse results (large improvement in diversity). Moreover, our proposed objective functions and the fusion network in the full model also clearly benefit to various metrics, with having better balance between the realness and the diversity.

Method	PSNR↑	SSIM↑	Diversity↑	FID↓
Base w/o ref	21.5255	0.7495	0.0099	91.44
Base w/ ref	22.5375	0.7527	0.0786	86.87
$+\mathcal{L}_{ms}$	22.6942	0.7446	**0.1297**	85.98
$+\mathcal{L}_{sim}$	22.6206	0.7461	0.1160	87.84
$+\mathcal{L}_{hist}$	22.6537	0.7543	0.1255	82.14
Full	**22.7014**	**0.7609**	0.1199	**80.91**

Table 1: Ablation study.

4. CONCLUSION

We propose a holistic framework to transfer a bi-level halftone image into a colorful continuous-tone image. Our model can either take a color reference image as the colorization guidance or automatically generate reasonable and diverse colorization results. Furthermore, our method utilizes a fusion network to efficiently integrate the results obtained from different orders of cascading the inverse halftone and colorization module, and finally produces high quality images.

5. REFERENCES

[1] Ping Wah Wong, "Inverse halftoning and kernel estimation for error diffusion," *IEEE Transactions on Image Processing (TIP)*, 1995.

[2] Li-Ming Chen and Hsueh-Ming Hang, "An adaptive inverse halftoning algorithm," *IEEE Transactions on Image Processing (TIP)*, 1997.

[3] Kuo-Liang Chung and Shih-Tung Wu, "Inverse halftoning algorithm using edge-based lookup table approach," *IEEE Transactions on Image Processing (TIP)*, 2005.

[4] Pedro G Freitas, Mylène CQ Farias, and Aletéia PF Araújo, "Enhancing inverse halftoning via coupled dictionary training," *Signal Processing: Image Communication*, 2016.

[5] Yi Xiao, Chao Pan, Xianyi Zhu, Hai Jiang, and Yan Zheng, "Deep neural inversexiao2017deep halftoning," in *International Conference on Virtual Reality and Visualization (ICVRV)*, 2017.

[6] Menghan Xia and Tien-Tsin Wong, "Deep inverse halftoning via progressively residual learning," in *Asian Conference on Computer Vision (ACCV)*, 2018.

[7] Qifan Gao, Xiao Shu, and Xiaolin Wu, "Deep restoration of vintage photographs from scanned halftone prints," in *IEEE International Conference on Computer Vision (ICCV)*, 2019.

[8] Chang-Hwan Son, "Inverse halftoning through structure-aware deep convolutional neural networks," *Signal Processing*, 2020.

[9] Jun-Yan Zhu, Richard Zhang, Deepak Pathak, Trevor Darrell, Alexei A Efros, Oliver Wang, and Eli Shechtman, "Toward multimodal image-to-image translation," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[10] Sergio Guadarrama, Ryan Dahl, David Bieber, Mohammad Norouzi, Jonathon Shlens, and Kevin Murphy, "Pixcolor: Pixel recursive colorization," in *British Machine Vision Conference (BMVC)*, 2017.

[11] Yun Cao, Zhiming Zhou, Weinan Zhang, and Yong Yu, "Unsupervised diverse colorization via generative adversarial networks," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2017.

[12] Aditya Deshpande, Jiajun Lu, Mao-Chuang Yeh, Min Jin Chong, and David Forsyth, "Learning diverse image colorization," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[13] Richard Zhang, Jun-Yan Zhu, Phillip Isola, Xinyang Geng, Angela S Lin, Tianhe Yu, and Alexei A Efros, "Real-time user-guided image colorization with learned deep priors," *ACM Transactions on Graphics (TOG)*, 2017.

[14] Yi Xiao, Peiyao Zhou, Yan Zheng, and Chi-Sing Leung, "Interactive deep colorization using simultaneous global and local inputs," in *IEEE International Conference on Acoustics, Speech, & Signal Processing (ICASSP)*, 2019.

[15] Mingming He, Dongdong Chen, Jing Liao, Pedro V Sander, and Lu Yuan, "Deep exemplar-based colorization," *ACM Transactions on Graphics (TOG)*, 2018.

[16] Faming Fang, Tingting Wang, Tieyong Zeng, and Guixu Zhang, "A superpixel-based variational model for image colorization," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 2019.

[17] Chufeng Xiao, Chu Han, Zhuming Zhang, Jing Qin, Tien-Tsin Wong, Guoqiang Han, and Shengfeng He, "Example-based colourization via dense encoding pyramids," in *Computer Graphics Forum*, 2020.

[18] Bo Li, Yu-Kun Lai, Matthew John, and Paul L Rosin, "Automatic example-based image colorization using location-aware cross-scale matching," *IEEE Transactions on Image Processing (TIP)*, 2019.

[19] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision (ECCV)*, 2016.

[20] Qi Mao, Hsin-Ying Lee, Hung-Yu Tseng, Siwei Ma, and Ming-Hsuan Yang, "Mode seeking generative adversarial networks for diverse image synthesis," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[21] Chenyang Lei and Qifeng Chen, "Fully automatic video colorization with self-regularization and diversity," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[22] Vuong Le, Jonathan Brandt, Zhe Lin, Lubomir Bourdev, and Thomas S Huang, "Interactive facial feature localization," in *European Conference on Computer Vision (ECCV)*, 2012.

[23] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.