

Emerging Applications of Reversible Data Hiding

Dongdong Hou
University of Science and
Technology of China
Hefei, China
houdd@mail.ustc.edu.cn

Weiming Zhang
University of Science and
Technology of China
Hefei, China
zhangwm@ustc.edu.cn

Jiayang Liu
University of Science and
Technology of China
Hefei, China
ljyljy@mail.ustc.edu.cn

Siyan Zhou
East China Normal University
Shanghai, China
51184506084@stu.ecnu.edu.cn

Dongdong Chen
University of Science and
Technology of China
Hefei, China
cd722522@mail.ustc.edu.cn

Nenghai Yu
University of Science and
Technology of China
Hefei, China
ynh@ustc.edu.cn

ABSTRACT

Reversible data hiding (RDH) is one special type of information hiding, by which the host sequence as well as the embedded data can be both restored from the marked sequence without loss. Beside media annotation and integrity authentication, recently some scholars begin to apply RDH in many other fields innovatively. In this paper, we summarize these emerging applications, including steganography, adversarial example, visual transformation, image processing, and give out the general frameworks to make these operations reversible. As far as we are concerned, this is the first paper to summarize the extended applications of RDH.

Keywords

reversible steganography, reversible adversarial example, reversible visual transformation, reversible image processing, reversible data hiding.

1. INTRODUCTION

Data hiding embeds messages into digital multimedia such as image, audio, video through an imperceptible way, which is mainly used for copyright protection, integrity authentication, covert communication. Some special signals such as medical imagery, military imagery and law forensics are so precious that cannot be damaged. To protect these signals, reversible data hiding (RDH) [1] is developed. Taking image as example (see Fig. 1), by RDH after embedding messages into the host image the generated marked image is visually invariant, and at the same

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ICIGP 2019, February 23–25, 2019, Singapore, Singapore

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6092-0/19/02...\$15.00

<https://doi.org/10.1145/3313950.3313952>

time we can losslessly restore host image after extracting the embedded messages.

RDH algorithms are well established, even the schemes [9], [10] approaching theoretical optimum have also appeared. As for RDH, the first step is to generate a host sequence with the small entropy such as prediction errors (PEs) [2]-[6], and then the users reversibly embed messages into the host sequence by modifying its histogram [7]-[9].

RDH is mainly used for media annotation and integrity authentication, but its application is now extended by scholars. With RDH we can restore both embedded messages and host image, this make the host image like storage disk which can be erasable. However, marked image generated from RDH is hard to resist detection. If we endow RDH with undetectability, and then such RDH algorithms called reversible steganography can be applied for convert storage [11]-[13]. Besides convert storage, we can also regard RDH as one tool to do many reversible image operations. To be detailed, after operating image to the desired target, we can explore the auxiliary parameters for restoring the original image from the target image, and then reversibly embed the parameters into the target image to get the reversible operated image. At the receiver's side, we extract these auxiliary parameters and the target image from the reversible operated image, and further restore the original image from the target image with the extracted parameters.

In the following contents, we will give out the general frameworks to do reversible steganography, reversible adversarial example, reversible visual transformation and reversible image processing respectively.



Fig. 1: Reversible data hiding.

2. EMERGING APPLICATIONS

2.1 Reversible Steganography

The popular method for privacy protection is encryption. But the messy codes of ciphertext with special form are easy to cause the attention of attackers. Therefore, covert storage hiding the existence of data has been widespread concerned. It is clear that covert storage requires two properties at the same time, i.e., undetectability and reversibility.

Steganography is a secure tool designed for covert communication, and the most successful steganographic algorithms [14], [15] are devoted to embedding messages while minimizing the total distortion, which can achieve the strong undetectability under the advanced steganalysis [16], [17]. However, steganography will destroy host image irreversibly. Different from covert communication, as for covert storage the image here is used as a special kind of storage medium that needs to be erasable like a disk. After deleting the stored data, the storage medium must be restored to its original state. To make the image erasable so that the storage space can be used repeatedly. Besides the undetectability, reversibility is also desired. From this point, RDH is suitable for covert storage, by which the cover image can be losslessly restored after the message being extracted. But traditional RDH algorithms are not secure under steganalysis.

The high undetectability of steganography is mainly achieved by modifying the complex regions of images. However, most of RDH algorithms give the priority of modifications to pixels in smooth regions due to that pixels in smooth areas can be predicted more accurately. That's the reason why traditional RDH cannot resist steganalysis. Recently, Hong et al. [11] give out the first RDH algorithm which has much higher undetectability than traditional ones. The undetectability is mainly achieved by embedding messages into PEs with small absolute values, but PEs in complex regions are preferentially modified through a sorting technique. Then Zhang et al. [12] improve Hong et al.'s method by giving the priority to PEs with the larger absolute values for accommodating messages.

To endow RDH with the undetectability, we should define inconsistent distortion metrics for each pixel according to its local variance. Generally speaking, the distortion metric in smooth region should be defined higher than that in noisy region. For the convenience of handling, inconsistent distortion metrics are quantized as multi-distortion metrics. Assume that these inconsistent distortion metrics are clustered into K ($K \leq N$, N is the volume of host elements) classes. Accordingly, the host sequence \mathbf{X} is segmented into K subsequences denoted as $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,N_i})$, where N_i is the volume of \mathbf{x}_i . The elements in \mathbf{x}_i will share the same distortion metric $d_i(x,y)$, where $1 \leq i \leq K$. Then after giving the embedding rate R , the rate-distortion problem of RDH under multi-distortion metrics is formulated as

$$\begin{aligned} & \sum_{i=1}^K N_i \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} P_{X_i}(x) P_{Y_i|X_i}(y|x) d_i(x,y) \\ \text{minimize } & \frac{\quad}{N} \\ & \sum_{i=1}^K N_i H(Y_i) = R + \sum_{i=1}^K N_i H(X_i) \\ \text{subject to } & \frac{\quad}{N} = R + \frac{\quad}{N} \end{aligned} \quad (1)$$

To minimize the average distortion between the host sequence \mathbf{X} and the generated marked sequence \mathbf{Y} , each sub-OTPM $P_{Y_i|X_i}(y|x)$ is desired for $i = 1, 2, \dots, K$, by which we can optimally modify the K sub-sequences and embed the corresponding messages into each sub-host-sequence. The unified framework for estimating the corresponding K sub-OTPMs $P_{Y_i|X_i}(y|x)$ is presented in [13]. After getting these sub-OTPMs, we perform recursive code construction (RCC) to finish the message embedding. Reversible steganography [11]-[13] has the reversibility as traditional RDH and the ability of undetectability as traditional steganography, which will be rather valuable in the applications of covert storage. Based on the above, one framework of reversible steganography minimizing the distortion is given as Fig. 2.

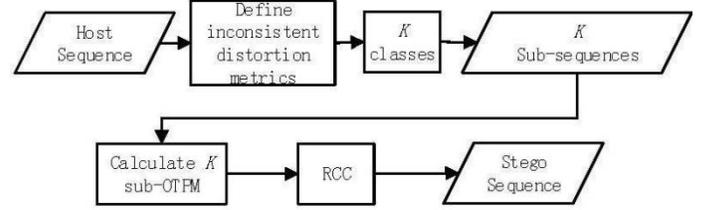


Fig. 2: Reversible steganography.

2.2 Reversible Adversarial Example

As for one deep neural network (DNN), its parameters are optimized by reducing the loss gradually on the training set. After optimizing these parameters, DNN can be successful applied in image classification, speech recognition, and so on. However, as for the trained network, one input with the carefully selected small perturbation perhaps will get a completely different result. Such input to fool the network is called an adversarial example. One can perturb an image to misclassify network while keeping the image quality to be imperceptible for human eyes. Even worse, an attacker can use the trained classifier to generate adversarial example, and then use it to fool another model.

The users create adversarial example to fool DNNs, but the created adversarial example must be controlled the users themselves. That is to say, the creators must be able to restore the adversarial example and can not let adversarial example to fool their own networks. Usually, the image to generate adversarial example is the sensitive and important one such as military imagery, which can not be damaged with loss. After modification the adversarial example can be deemed as a noisy image, although the distortion sometimes may be not sensitive to human eyes. However, the further processing on adversarial example must will be interfered. For example, the adversarial example must will affect the feature extraction and thus result in the decrease of processing accuracy. As for some important applications such as military system, police system, and so on, the slight decrease of accuracy perhaps will result in serious consequences. Therefore, reversible adversarial example is desired.

One simple way for creating adversarial example proposed by Goodfellow *et al* [18] is adding the perturbation on host image, and the added perturbation is the direction in image space which yields the highest increase of the linearized cost. The hyper-parameter ϵ is applied to limit the distance between the adversarial image X_{adv} and the host image X . Specifically,

$$X_{adv} = X + \epsilon \times \text{sign}(\nabla_x J(X, y_{true})) \quad , \quad (2)$$

where y_{true} is the target fooling class.

In [19], we firstly present the concept of reversible adversarial example, whose framework is shown as Fig. 3. We reversibly embed the perturbation into adversarial example to get reversible adversarial example. To restore the original image, we first restore adversarial example after extracting the perturbation from reversible adversarial example, and further subtract perturbation from adversarial example to return the original image.

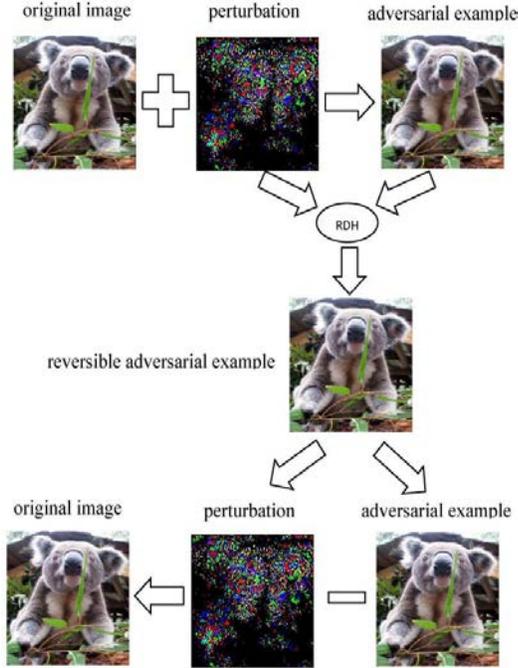


Fig. 3: Reversible adversarial example.

2.3 Reversible Visual Transformation

Traditional secure data hiding algorithms [14], [15] are effective for embedding a small part of messages into a large cover, such as an image. But for image transmission and storage, the image itself is the secret file to be protected. Therefore, we need large capacity data hiding method to hide image. Reversible visual transformation is usually used for image protection, which reversibly transforms a secret image to a freely-selected target image with the same size and gets a camouflage image similar to the target image.

Lai *et al.* [20] propose the first work about visual transformation, which generates transformed image by replacing each target block with one corresponding similar secret block, and then record the location mapping accounting for the main auxiliary information with RDH. By this method, the target image must be similar to the secret image. What is more, the camouflage image is in the poor visual quality.

Generally speaking, as long as the mean and the standard deviation of each secret block are adjusted to be similar with those of the corresponding target block, the secret image will be masqueraded as the target image visually. Lee *et al.*'s method [21] can transform a secret image to a freely selected target image via color transformation [23], which greatly improves the visual quality of camouflage image. However, by Lee *et al.*'s method the

secret image cannot be losslessly reconstructed due to that the adopted color transformation [23] is not reversible.

To make the transformation reversible, a novel reversible visual transformation (RVT) scheme is presented by using shift transformation [22]. Before shift transformation, a nonuniform clustering algorithm is utilized to match secret blocks and target blocks, which largely reduces the amount of accessorial information for recording indexes of secret blocks. To further reduce the amount of accessorial information, the correlations among three color channels and inside each color channel are explored [24]. Therefore, not only the visual quality of camouflage image improved a lot by dividing secret and target image into smaller blocks for transformation, but also the reversibility is achieved.

As for RVT, there are two steps to generate camouflage image, the first step is dividing secret image and target image into small blocks and transforming secret image to one target image to get a transformed image, and in the second step we embed some auxiliary parameters into the transformed image by RDH. At the receiver's side, by the decode processes of RDH we restore the auxiliary parameters and the transformed image from the camouflage image firstly, and then the secret image is restored from the transformed image with the help of the restored auxiliary parameters. The diagram RVT is as Fig. 4.

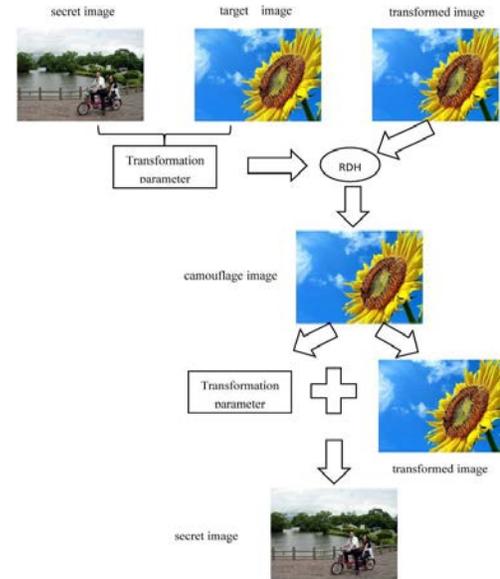


Fig. 4: Reversible visual transformation.

2.4 Reversible Image Processing

Image processing is rather popular, and people process their images to desired results through various of tools [25], [26]. Nowadays, most of image processing methods are irreversible, that is to say after processing the image we will damage the original copy. However, sometimes the client may not satisfy with the processed result, then the irreversibility will result in great inconvenience. Of course, the users can save the original copy as a backup before processing it, which will cost much more storage space. Indeed, Google's Picasa's automatic image enhancement system is one of such examples, who stores the original image in a separate folder as the backup. Instead of storing both the original and the processed images, Apple and Google Photos keep the

original image and a small record file of the applied enhancements. The enhanced image will be displayed each time by re-enhanced the original image with the help of the record file. However, the enhanced image can be only correctly viewed on their own software. What is more, processing original image for displaying each time will waste many computing resources.

As shown in Fig. 5, technological process of reversible image processing [27] is described as follows. By using algorithms or softwares, the original image is processed to the desired result regarded as target image. Since the target image is obtained from the original image, the correlations between original image and target image are high, thus can be explored to compress the original image effectively. We get the reversible secret image similar to the target image by embedding the compressed secret image into the target image with an RDH scheme.

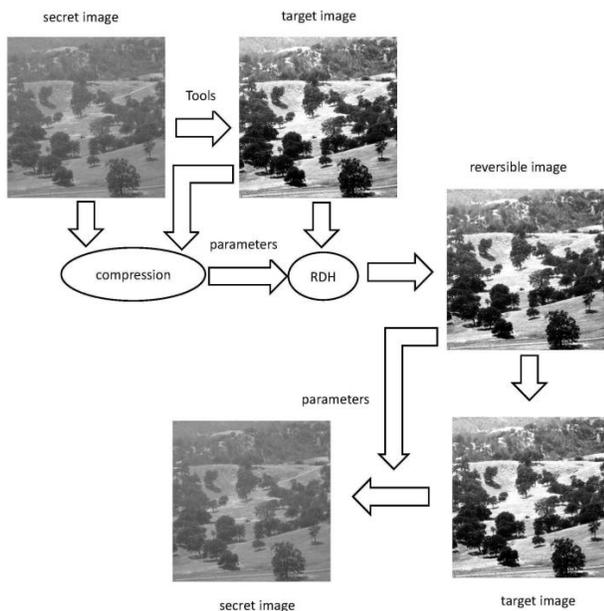


Fig. 5: Reversible image processing.

Some image processing algorithms only process local regions of one image, in such case, the amount of information for recording the processed region is small, and it is much easy for reversibility. Taking inpainting as example shown in Fig. 6, we cut out the person in original image, and then inpaint the remained content to make it natural. To make the operation reversible, we only need to reversibly embed the cropped person into the inpainted image to get the reversible inpainted image.

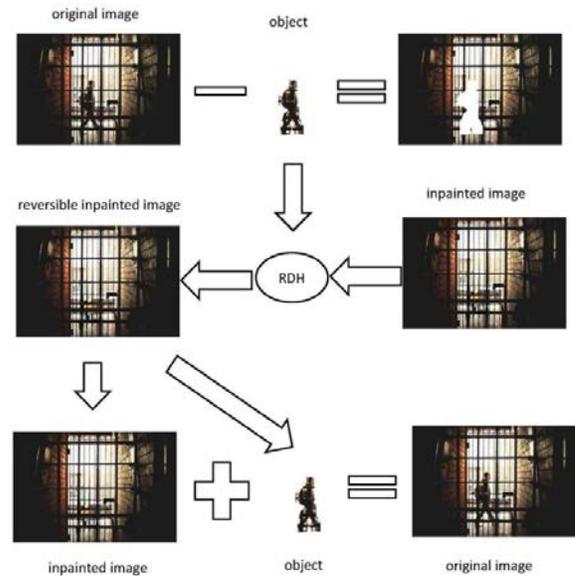


Fig. 6: Reversible image inpainting.

3. CONCLUSION AND DISCUSSION

In the past years, the motivation of RDH is mainly about integrity authentication. Now, we present some innovate applications based on RDH, containing reversible steganography, reversible image processing, reversible adversarial example, reversible visual transformation. Of course, besides those, we believe that style transfer [28]–[31], colorization [32], [33], and so on, can be also reversible. These applications have valuable prospects and extend the application of RDH a lot.

There are still many difficulties to be solved in these reversible operations. For example, compared to adversarial example the accuracy of reversible adversarial example will be slightly descended. As for some complex image processing methods, the auxiliary information for restoring original image is too much to be embedded by RDH, or will greatly degrade the quality of the processed image. In the future, we will try to overcome those difficulties to make these reversible operations be more better.

4. ACKNOWLEDGMENTS

Our thanks to ACM SIGCHI for allowing us to modify templates they had developed. This work was supported in part by the Natural Science Foundation of China under Grant U1636201 and 61572452.

5. REFERENCES

- [1] A. Khan, A. Siddiq, S. Munib, and S. A. Malik, "A recent survey of reversible watermarking techniques," *Information sciences*, vol. 279, pp. 251–272, 2014.
- [2] B. Ou, X. Li, Y. Zhao, R. Ni and Y. Q. Shi, "Pairwise Prediction-Error Expansion for Efficient Reversible Data Hiding," *IEEE Trans. Image Processing*, vol. 22, no. 12, pp. 5010-5021, Dec. 2013.
- [3] B. Ou, X. Li, J. Wang and F. Peng, "High-fidelity reversible data hiding based on geodesic path and pairwise prediction-

- error expansion," *Neurocomputing*, vol. 226, pp. 23-34, Feb, 2017.
- [4] J. Wang, J. Ni, X. Zhang and Y. Shi, "Rate and distortion optimization for reversible data hiding using multiple histogram shifting," *IEEE Trans. cybernetics*, vol.47, no. 2, pp. 315-326, Feb, 2017.
- [5] H. Yao, C. Qin, Z. Tang, and Y. Tian, "Guided filtering based color image reversible data hiding," *Journal of Visual Communication and Image Representation*, vol. 43, pp. 152-163, 2017.
- [6] D. Hou, W. Zhang, K. Chen, S. Lin and N. Yu, "Reversible Data Hiding in Color Image with Grayscale Invariance," in *IEEE Trans. Circuits and Systems for Video Technology*. doi: 10.1109/TCSVT.2018.2803303.
- [7] J. Tian, "Reversible Data Embedding Using a Difference Expansion," *IEEE Trans. on Circuits System and Video Technology*, vol. 13, no. 8, pp. 890-896, Aug. 2003.
- [8] Z. Ni, Y. Shi, N. Ansari, and S. Wei, "Reversible Data Hiding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 3, pp. 354-362, 2006.
- [9] W. Zhang, X. Hu, X. Li, and N. Yu, "Recursive Histogram Modification: Establishing Equivalency between Reversible Data Hiding and Lossless Data Compression," *IEEE Trans. Image Processing*, Vol. 22, no. 7, pp. 2775-2785, July 2013.
- [10] D. Hou, W. Zhang, Y. Yang, and Y. Nenghai, "Reversible Data Hiding under Inconsistent Distortion Metrics," *IEEE Trans. Image Processing*, vol. 27, no. 10, pp.5087-5099, 2018.
- [11] W. Hong, T. Chen, J. Chen. "Reversible data hiding using Delaunay triangulation and selective embedment," *Information Sciences*, vol. 308, pp. 140-154, 2015.
- [12] Z. Zhang and W. Zhang, "Reversible steganography: Data hiding for covert storage," in *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pp. 753-756, 2015.
- [13] D. Hou, W. Zhang, Y. Yang and N. Yu, "Reversible Data Hiding under Inconsistent Distortion Metrics" in *IEEE Trans. Image Processing*, vol. 27, no. 10, pp. 5087-5099, 2018.
- [14] B. Li, M. Wang, X. Li, S. Tan, and J. Huang, "A strategy of clustering modification directions in spatial image steganography," *IEEE Trans. Information Forensics and Security*, vol. 10, no. 9, pp. 1905-1917, 2015.
- [15] B. Li, M. Wang, J. Huang, and X. Li, "A new cost function for spatial image steganography," in *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 4206-4210, 2014.
- [16] T. Pevny, P. Bas, and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," *IEEE Trans. Information Forensics and Security*, vol. 5, no. 2, pp. 215-224, 2010.
- [17] Fridrich J, Kodovsky J. "Rich models for steganalysis of digital images," *IEEE Trans. Information Forensics and Security*, vol. 7, no. 3, pp. 868-882, 2012.
- [18] I. J. Goodfellow, J. Shlens, C. Szegedy, "Explaining and harnessing adversarial examples," *International Conference on Learning Representations*, 2015.
- [19] J. Liu, D. Hou, W. Zhang, and N. Yu, "Reversible Adversarial Examples," *International Conference on Communications, Signal Processing, and Systems*, 2018.
- [20] I.-J. Lai and W.-H. Tsai, "Secret-fragment-visible mosaic image-a new computer art and its application to information hiding," *IEEE Trans. Information Forensics and Security*, vol. 6, no. 3, pp. 936-945, 2011.
- [21] Y.-L. Lee and W.-H. Tsai, "A new secure image transmission technique via secret-fragment-visible mosaic images by nearly reversible color transformations," *IEEE Trans. Circuits and Systems for Video Technology*., vol. 24, no. 4, pp. 695-703, 2014.
- [22] D. Hou, W. Zhang, and N. Yu, "Image camouflage by reversible image transformation," *Journal of Visual Communication and Image Representation*, vol. 40, Part A, pp. 225-236, 2016.
- [23] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 34-41, 2001.
- [24] D. Hou, C. Qin, W. Zhang, and N. Yu, "Reversible Visual Transformation via Exploring the Correlations within Color Images," *Journal of Visual Communication and Image Representation*, vol. 53, pp. 134-145, 2018.
- [25] R. C. Gonzalez, *Digital image processing*. Pearson Education India, 2009.
- [26] Q. Fan, D. Chen, L. Yuan, G. Hua, N. Yu, and B. Chen, "Decouple learning for parameterized image operators", in *ECCV*, pp. 455-471, 2018.
- [27] D. Hou, W. Zhang, Z. Zhan, R. Jiang, Y. Yang, and N. Yu, "Reversible image processing via reversible data hiding," in *Proc. IEEE Int. Conf. Digital Signal Processing (DSP)*, pp. 427-431, Oct. 2016.
- [28] L. Gatys, A. S. Ecker, and M. Bethge, "Texture synthesis using convolutional neural networks," In *Advances in Neural Information Processing Systems*, pages 262-270, 2015.
- [29] D. Chen, J. Liao, L. Yuan, N. Yu, G. Hua, "Coherent Online Video Style Transfer", in *ICCV*, pp. 1105-1114, 2017.
- [30] D. Chen, L. Yuan, J. Liao, N. Yu, and G. Hua, "Stylebank: An explicit representation for neural image style transfer", in *CVPR*, pp. 2770-2779, 2017.
- [31] D. Chen, L. Yuan, J. Liao, N. Yu, and G. Hua, "Stereoscopic neural style transfer", in *CVPR*, 2018.
- [32] G. Larsson, M. Maire, and G. Shakhnarovich, "Learning representations for automatic colorization", In *Proc. ECCV*. 577-593, 2016.
- [33] M. He, D. Chen, J. Liao, P. V. Sander, and L. Yuan, "Deep Exemplarbased Colorization", *ACM Transactions on Graphics*, 2018