Can Generative Colourisation Help Face Recognition?

Pawel Drozdowski[1], Daniel Fischer[1], Christian Rathgeb[1], Julian Geissler[1], Jan Knedlik[1], Christoph Busch[1]

Abstract: Generative colourisation methods can be applied to automatically convert greyscale images to realistically looking colour images. In a face recognition system, such techniques might be employed as a pre-processing step in scenarios where either one or both face images to be compared are only available in greyscale format. In an experimental setup which reflects said scenarios, we investigate if generative colourisation can improve face sample utility and overall biometric performance of face recognition. To this end, subsets of the FERET and FRGCv2 face image databases are converted to greyscale and colourised applying two versions of the DeOldify colourisation algorithm. Face sample quality assessment is done using the FaceQnet quality estimator. Biometric performance measurements are conducted for the widely used ArcFace system with its built-in face detector and reported according to standardised metrics. Obtained results indicate that, for the tested systems, the application of generative colourisation does neither improve face image quality nor recognition performance. However, generative colourisation was found to aid face detection and subsequent feature extraction of the used face recognition system which results in a decrease of the overall false reject rate.

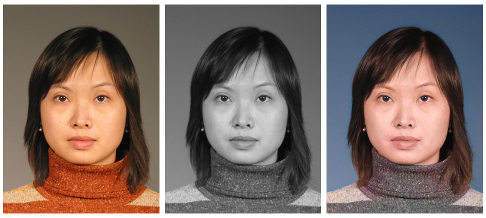
Keywords: biometrics, face recognition, face image quality, generative colourisation.

1 Introduction

Developments in deep neural networks have shown impressive improvements in diverse generative image processing tasks, *e.g.* single-image super resolution [Ha19] or inpainting [Li18]. Focusing on face images, domain-specific techniques have been established, *e.g.* face hallucination [LSF07, Ch18, GSŠ19] or face completion [Li17, Ca19]. Some of these methods have been found advantageous in various face-related vision tasks, such as face detection and recognition [Li19, MIA19]. In addition to the aforementioned generative methods, image colourisation schemes based on deep neural networks have been proposed [CYS15, ZIE16, NNE18], often for the purpose of restoring old images and film footage. Said methods are able to generate realistic colour images from greyscale images, including facial imagery. An example for applying a state-of-the-art colourisation algorithm to a face image is depicted in figure 1.

In this work, we investigate if generative colourisation can be advantageous in the context of face recognition. To this end, face image subsets of two publicly available databases [Ph98, Ph05] are converted to greyscale and colourised using two versions of a public colourisation algorithm [An19, Ke19]. Subsequently, face sample quality is assessed employing the public algorithm of FaceQnet [He19]; furthermore, standardised ISO/IEC methodology and metrics [IS06] are used to evaluate the biometric performance of the ArcFace recognition system [De19] in a scenario-based manner. The considered scenarios reflect different

[1] da/sec – Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany, {pawel.drozdowski,daniel.fischer,christian.rathgeb,christoph.busch}@h-da.de



(a) original (b) greyscale (c) colourised

Fig. 1: Example of generative face colourisation: (a) the original image, (b) the original image converted to greyscale, and (c) the colourised greyscale image using the method of Antic [An19].

relevant use cases in which colourisation methods might be applied to reference and/or probe images prior to the feature extraction. To the best of the authors' knowledge, this reproducible study represents the first investigation of the usefulness of generative colourisation in face recognition.

The remainder of this paper is organised follows: the employed image colourisation methods are described in section 2. Relevant scenarios which are considered in experiments are summarised in section 3. Experimental setup and results are presented in section 4. Conclusions are drawn in section 5.

2 Face Image Colourisation

The goal of image colouration is the addition of colour information to greyscale images (as illustrated in figure 1). Early solutions (see *e.g.* [WAM02, YS06]) required a substantial amount of input, interaction, and/or expertise from a human operator who guides the algorithm (*e.g.* by providing scribbles of colour, choosing suitable reference images, segmenting images, or providing annotations). Later developments in this area sought to automate parts of the aforementioned user interactions with some success (see *e.g.* [Li08, Ch11]). Recently, fully automated methods based on massive datasets coupled with deep learning (*e.g.* [CYS15, ZIE16, He18]) and adversarial learning (*e.g.* [Ca17, NNE18]) emerged to address the limitations of previous methods specifically for the image colourisation task, and more generally for image-to-image translation problems (*e.g.* [Is17, Zh17]).

There exist different repositories with deep learning-based greyscale image colourisation software; however, many of them have certain limitations. For example, [E.17] only handles relatively low resolution images, while [Zh19] requires some user input in a semi-automatic process. In this work, we utilise one of the most recently published image colourisation methods called "DeOldify" [An19]. The software is based on concepts from [Zh18] and [He17], as well as a novel (as of yet unpublished) GAN pre-training strategy. In addition to the current version of the software, we also test an older version thereof [Ke19], which uses a different GAN training strategy inspired by [Ka17]. The authors of the software provide three pre-trained models: "artistic", "stable", and "video". We use the "stable" model, since according to the authors it is expected to achieve the best results *i.a.* for portraits, which are the use case considered in our paper. The used software which is applied to original images previously converted to greyscale[2] convinces with ex-

[2] Using ImageMagick command `magick in.png -grayscale Rec709Luma out.png`, see `https://imagemagick.org/script/command-line-options.php#grayscale`

cellent visual results and is easy and flexible to use. It should also be noted that this paper constitutes a preliminary study on this subject; future works (see section 5) may include considering a more comprehensive selection of image colourisation methods.

Figure 2 shows examples of colourised greyscale images (from the used facial image databases, see section 4.1) generated by the aforementioned methods. The results look mostly clean and realistic; furthermore, the newer model appears to produce more visually pleasing results. Note, that the colourised images are not identical to the corresponding original colour images. Those differences are inevitable: the colourisation algorithm needs to assign new pixel values in three dimensions (RGB) to pixel values with variation only along one dimension (intensity or luminance). It is possible for different colours to have the same luminance value, but different hue or saturation. Therefore, there exists no inherent "correct" solution to the task of colourising a greyscale image.

(a) original (b) greyscale (c) colourised [An19] (d) colourised [Ke19]

Fig. 2: Examples of reference (top row) and probe (bottom row) images: (a) the original images, (b) the original images converted to greyscale, (c)-(d) the colourised greyscale images.

3 Scenarios

As illustrated in figure 3, we consider five different scenarios which result in different pairings of compared reference and probe images:

- *Scenario 1*: baseline scenario; original reference and probe face images are used.
- *Scenario 2*: original reference image is used; probe image is converted to greyscale.
- *Scenario 3*: reference and probe images are converted to greyscale.
- *Scenario 4*: reference and probe images are converted to greyscale and colourised.
- *Scenario 5*: original reference image is used; probe image is converted to greyscale and colourised.

The second scenario might represent a surveillance or automated border control scenario in which a greyscale probe image is compared against a colour reference image. Accordingly, the third scenario reflects a use case where only greyscale images are processed by the face recognition system. Note that this applies for many older face recognition systems which utilise handcrafted feature extractors. The last two scenarios involve the application of colourisation to greyscale images. Specifically, colourisation is applied to reference and probe images or only to the probe image, respectively.

4 Experiments

The following subsections describe the experimental setup (section 4.1), conducted quality assessment (section 4.2), and biometric performance measurements (section 4.3).

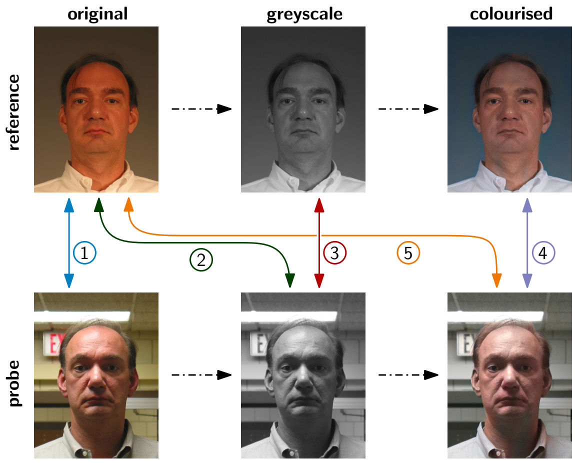


Fig. 3: Overview of applied the image processing chain (dash-dotted lines) and pairings (coloured lines) of reference and probe face images for the considered scenarios.

4.1 Experimental Setup

Subsets of two publicly available face image databases, *i.e.* FERET [Ph98] and FRGCv2 [Ph05], were used in the experiments. For reference images frontal faces with neutral expression have been manually chosen and ICAO compliance has been verified [In15]. Wherever possible, probe images from different acquisition session were preferentially chosen in order to obtain a realistic scenario. Examples of probe and reference images of both face image subsets are depicted in figure 2 and figure 3. The experiments are conducted in biometric verification mode. The number of subjects, corresponding reference and probe images, as well as the resulting number of genuine and impostor comparisons are listed in table 1.

Tab. 1: Overview of face image subsets from the FERET and FRGCv2 face databases.

Database	Subjects	Images		Comparisons	
		Reference	Probe	Genuine	Impostor
FERET	529	529	791	791	147,712
FRGCv2	533	984	1,726	3,298	144,032

For quality assessment, the FaceQnet algorithm [He19] is used.[3] This public face sample quality estimator is based on deep learning and returns a quality score (*i.e.* high values indicate good quality). In order to measure biometric performance, the widely used state-of-the-art ArcFace system [De19] is employed which has shown competitive recognition performance among open-source face recognition systems.[4] For a pair of reference and probe face images, this system returns a distance score (*i.e.* low values indicate high similarity). Note that when presented with a biometric sample, the face recognition system might internally perform some kind of colour space transformation(s). The utility of the individual colour channels for the purposes of face recognition has been investigated for

[3] FaceQnet has been shown to achieve convincing results, is open-source, and a pre-trained model is available, see `https://github.com/uam-biometrics/FaceQnet`

[4] ArcFace is open-source with a pre-trained model available at `https://github.com/deepinsight/insightface`

older systems by [BH08]. However, it is out of scope for this paper, as it investigates the effects of generative colourisation on facial recognition.

Tab. 2: Overview of face sample quality results.

Database	Colour	Mean	Median	St. Dev.	Minimum	Maximum
FERET	Original	0.616	0.616	0.050	0.460	0.777
	Greyscale	0.614	0.615	0.051	0.433	0.767
	Colourised [An19]	0.608	0.606	0.049	0.448	0.756
	Colourised [Ke19]	0.608	0.607	0.047	0.454	0.755
FRGCv2	Original	0.617	0.615	0.051	0.449	0.802
	Greyscale	0.622	0.620	0.053	0.426	0.794
	Colourised [An19]	0.616	0.614	0.053	0.435	0.777
	Colourised [Ke19]	0.615	0.613	0.052	0.453	0.811

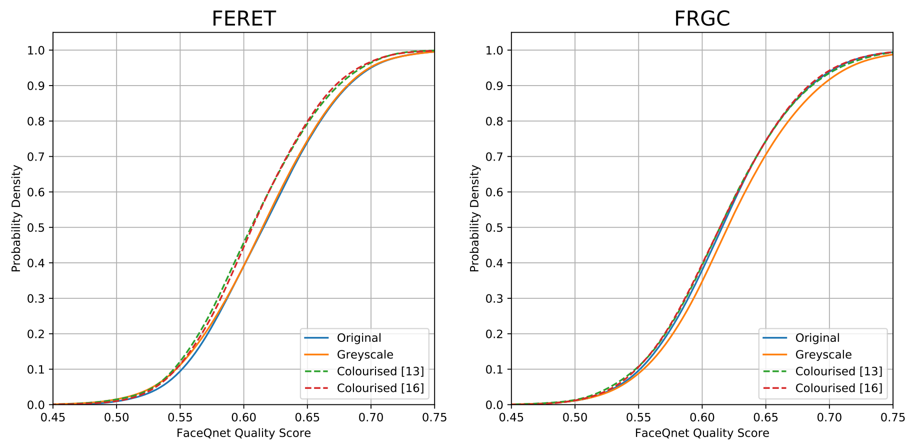
Biometric performance is evaluated in terms of False Non-Match Rate (FNMR) and False Match Rate (FMR). In addition, the Failure-to-Acquire rate (FTA) is measured as the proportion of verification attempts for which the system fails to capture or locate an image or signal of sufficient quality [IS06]. The False Reject Rate (FRR) is then estimated as the proportion of genuine verification transactions that will be incorrectly denied. This includes transactions denied due to failures-to-acquire as well as those denied due to false non-match decisions, $\text{FRR} = \text{FTA} + \text{FNMR} \times (1 - \text{FTA})$ [IS06]. More precisely, the FNMR and FRR are estimated at a false match probability of 0.1%, referred to as $\text{FNMR}_{0.1}$ and $\text{FRR}_{0.1}$, respectively. This operation point is recommended in the guidelines of European Agency for the Management of Operational Cooperation at the External Borders (FRONTEX) [FR15]. Genuine comparisons are performed for all of the previously described scenarios, while impostor comparisons are only performed for the first baseline scenario. That is, the decision thresholds estimated from the baseline scenario are used in all scenarios. Additionally, a decidability measure (d') [Da00] calculated as $d' = |\mu_g - \mu_i| / \sqrt{\frac{1}{2}(\sigma_g^2 + \sigma_i^2)}$ is reported, where μ_g and μ_i represent the means of the genuine and impostor score distributions and σ_g and σ_i their standard deviations, respectively.

4.2 Quality Assessment

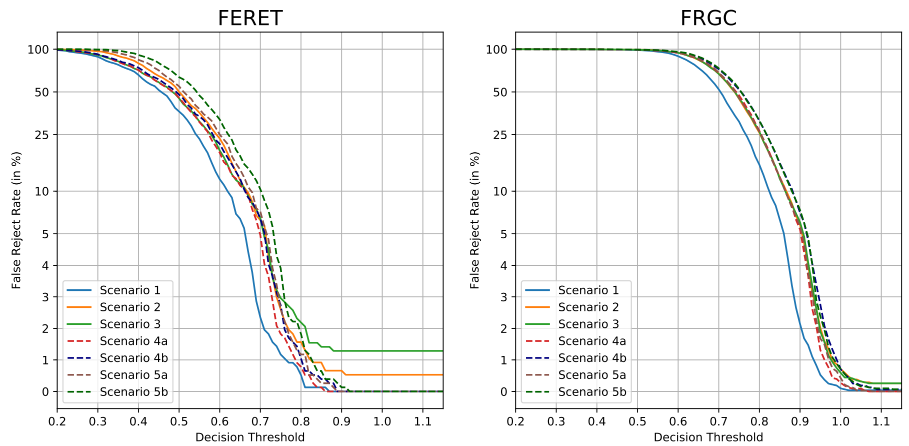
Figure 4(a) depicts the cumulative distribution function of sample quality scores. Corresponding statistical properties are summarised in table 2. The sample quality of the colourised images is generally lower compared to the greyscale images; however, the differences are very small. Hence, for the used method [He19], applied colourisation techniques do not yield improvements. This does not necessarily mean that colourisation might not be helpful in general – other algorithms in the pipeline might benefit from it, or it could be the case that other colourisation methods might improve the sample quality.

4.3 Performance Evaluation

Obtained biometric performance rates are summarised in table 3. On both databases practical biometric performance is achieved. Generally, better performance rates are obtained on the FERET database which contains more constrained face images. Further, it can be observed that increased FTAs are obtained for scenarios in which greyscale images are processed. However, by using image colourisation, the FTAs are reduced to that of the baseline system. That is, colourisation is found helpful to reduce the FTA on the greyscale



(a) Cumulative distribution of face sample quality.



(b) FRRs in relation to decision threshold.

Fig. 4: Experimental results.

images. This is also reflected in the FRRs which are plotted in figure 4(b) (the FRRs values begin at 100% for the low decision thresholds, as the used face recognition system works with dissimilarity-based comparison scores). Example face images for which the feature extraction failed on greyscale images but succeeded on the corresponding colourised images are shown in figure 5. Focusing on the algorithmic performance rates, *i.e*. FNMR, the application of colourisation yields generally worse comparison scores (higher dissimilarity) compared to scenarios in which greyscale images are processed directly.

5 Conclusion

Deep learning-based generative image colourisation techniques show impressive visual results for converting greyscale images to colour images. In this work, we investigated the usefulness of such techniques for facial recognition. For this purpose, open-source face image quality assessment and recognition tools are evaluated on two public databases

Tab. 3: Overview of biometric performance rates.

Database	Scenario	d'	FTA (%)	$FNMR_{0.1}$ (%)	$FRR_{0.1}$ (%)
FERET	1	9.290	0.000	0.000	0.000
	2	9.208	0.530	0.000	0.530
	3	8.793	1.288	0.000	1.288
	4 [An19]	8.739	0.000	0.000	0.000
	4 [Ke19]	8.553	0.000	0.000	0.000
	5 [An19]	9.344	0.000	0.000	0.000
	5 [Ke19]	9.306	0.000	0.000	0.000
FRGCv2	1	8.436	0.037	0.310	0.347
	2	7.872	0.258	0.446	0.703
	3	7.816	0.258	0.450	0.707
	4 [An19]	7.821	0.000	0.474	0.474
	4 [Ke19]	7.498	0.037	0.520	0.557
	5 [An19]	7.914	0.000	0.454	0.454
	5 [Ke19]	7.664	0.037	0.526	0.562

(a) greyscale (b) colourised [An19] (c) colourised [Ke19]

Fig. 5: Example face images for which the feature extraction failed on the greyscale images but succeeded on the colourised images.

considering scenarios where face images are converted to greyscale and colourised using state-of-the-art colourisation software.

In the conducted experiments, the effects of colourisation on sample quality were insignificant and did not result in improvements. To fully evaluate the impact of colourisation on biometric performance, more experiments with larger and more unconstrained datasets, as well as more facial recognition systems are needed to produce statistically significant results. The scenario-based evaluation of the comparison scores indicated generally inferior comparison scores for the colourised images compared to the direct use of greyscale images. These results are logically comprehensible since colourisation only aims at producing plausible colour images based on learned statistics which may vary for each image, *i.e.* colourised face images of various images a single subject may look different.

Finally, it was observed that colourisation can reduce the FTA, *i.e.* face detection and feature extraction exhibit more robustness if colourisation is applied. However, this may also highly depend on the used face recognition system. This preliminary study opens several avenues of potential research. Future works in this area may include testing other facial recognition and quality estimation methods, as well as different image colourisation schemes. Furthermore, the study could be extended to other applications of facial biometrics, such as biometric identification and classification of demographic attributes.

Acknowledgements

This research work has been funded by the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References

[An19] Antic, J.: , DeOldify – A Deep Learning based project for colorizing and restoring old images (and video!). `https://github.com/jantic/DeOldify`, 2019.

[BH08] Bours, P.; Helkala, K.: Face Recognition Using Separate Layers of the RGB Image. In: Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP). IEEE, pp. 1035–1042, 2008.

[Ca17] Cao, Y.; Zhou, Z.; Zhang, W.; Yu, Y.: Unsupervised diverse colorization via generative adversarial networks. In: Machine Learning and Knowledge Discovery in Databases. Springer, pp. 151–166, 2017.

[Ca19] Cai, J.; Han, H.; Shan, S.; Chen, X.: FCSR-GAN: Joint Face Completion and Super-resolution via Multi-task Learning. Trans. on Biometrics, Behavior, and Identity Science (TBIOM), 2019.

[Ch11] Chia, A. Y.-S.; Zhuo, S.; Gupta, R. K.; Tai, Y.-W.; Cho, S.-Y.; Tan, P.; Lin, S.: Semantic colorization with internet images. In: Trans. on Graphics (TOG). volume 30. ACM, p. 156, 2011.

[Ch18] Chen, Y.; Tai, Y.; Liu, X.; Shen, C.; Yang, J.: FSRNet: End-to-End Learning Face Super-Resolution with Facial Priors. In: Conf. on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 2492–2501, 2018.

[CYS15] Cheng, Z.; Yang, Q.; Sheng, B.: Deep Colorization. In: Int. Conf. on Computer Vision (ICCV). IEEE, pp. 415–423, 2015.

[Da00] Daugman, J.: Biometric Decision Landscapes. Technical Report UCAM-CL-TR-482, University of Cambridge - Computer Laboratory, 2000.

[De19] Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S.: Arcface: Additive angular margin loss for deep face recognition. In: Conf. on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 4690–4699, 2019.

[E.17] E.Wallner: , Coloring black and white images with deep learning. `https://blog.floydhub.com/colorizing-b-w-photos-with-neural-networks/`, 2017.

[FR15] FRONTEX: . Best Practice Technical Guidelines for Automated Border Control ABC Systems, 2015.

[GSŠ19] Grm, K.; Scheirer, W. J.; Štruc, V.: Face hallucination using cascaded super-resolution and identity priors. Trans. on Image Processing (TIP), pp. 2150–2165, 2019.

[Ha19] Ha, V. K.; Ren, J.-C.; Xu, X.-Y.; Zhao, S.; Xie, G.; Masero, V.; Hussain, A.: Deep Learning Based Single Image Super-resolution: A Survey. Int. J. of Automation and Computing (IJAC), 16(4):413–426, 2019.

[He17] Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S.: GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In: Int. Conf. on Neural Information Processing Systems (NIPS). ACM, pp. 6626–6637, 2017.

[He18] He, M.; Chen, D.; Liao, J.; Sander, P. V.; Yuan, L.: Deep exemplar-based colorization. Trans. on Graphics (TOG), 37(4):47, 2018.

[He19] Hernandez-Ortega, J.; Galbally, J.; Fiérrez, J.; Haraksim, R.; Beslay, L.: FaceQnet: Quality Assessment for Face Recognition based on Deep Learning. In: Int. Conf. on Biometrics (ICB). IEEE, 2019.

[In15] International Civil Aviation Organization: . Machine Readable Passports – Part 9 – Deployment of Biometric Identification and Electronic Storage of Data in eMRTDs, 2015.

[IS06] ISO/IEC JTC1 SC37 Biometrics: . ISO/IEC 19795-1:2006. Information Technology – Biometric Performance Testing and Reporting – Part 1: Principles and Framework, 2006.

[Is17] Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A. A.: Image-to-image translation with conditional adversarial networks. In: Conf. on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 1125–1134, 2017.

[Ka17] Karras, T.; Aila, T.; Laine, S.; Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196, 2017.

[Ke19] Kelley, D.: , DeOldify – A Deep Learning based project for colorizing and restoring old images. `https://github.com/dana-kelley/DeOldify`, 2019.

[Li08] Liu, X.; Wan, L.; Qu, Y.; Wong, T.-T.; Lin, S.; Leung, C.-S.; Heng, P.-A.: Intrinsic colorization. In: Trans. on Graphics (TOG). volume 27. ACM, p. 152, 2008.

[Li17] Li, Y.; Liu, S.; Yang, J.; Yang, M.-H.: Generative Face Completion. In: Conf. on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 3911–3919, 2017.

[Li18] Liu, G.; Reda, F. A.; Shih, K. J.; Wang, T.-C.; Tao, A.; Catanzaro, B.: Image Inpainting for Irregular Holes Using Partial Convolutions. In: European Conf. on Computer Vision (ECCV). Springer, pp. 85–100, 2018.

[Li19] Li, P.; Prieto, L.; Mery, D.; Flynn, P. J.: On Low-Resolution Face Recognition in the Wild: Comparisons and New Techniques. Trans. on Information Forensics and Security (TIFS), 14(8):2000–2012, 2019.

[LSF07] Liu, C.; Shum, H.-Y.; Freeman, W. T.: Face Hallucination: Theory and Practice. Int. J. of Computer Vision (IJCV), 75(1):115–134, 2007.

[MIA19] Mathai, J.; Iacopo, M.; Abd-Almageed, W.: Does Generative Face Completion Help Face Recognition? In: Int. Conf. on Biometrics (ICB). IEEE, 2019.

[NNE18] Nazeri, K.; Ng, E.; Ebrahimi, M.: Image colorization using generative adversarial networks. In: Int. Conf. on Articulated Motion and Deformable Objects. Springer, pp. 85–94, 2018.

[Ph98] Phillips, J.; Wechsler, H.; Huang, J.; Rauss, P.: The FERET Database and Evaluation Procedure for Face Recognition Algorithms. Image and Vision Computing Journal (IMAVIS), 16(5):295–306, 1998.

[Ph05] Phillips, P. J.; Flynn, P. J.; Scruggs, T.; Bowyer, K. W.; Chang, J.; Hoffman, K.; Marques, J.; Min, J.; Worek, W.: Overview of the face recognition grand challenge. In: Conf. on Computer Vision and Pattern Recognition (CVPR). volume 1. IEEE, pp. 947–954, 2005.

[WAM02] Welsh, T.; Ashikhmin, M.; Mueller, K.: Transferring color to greyscale images. In: Trans. on graphics (TOG). volume 21. ACM, pp. 277–280, 2002.

[YS06] Yatziv, L.; Sapiro, G.: Fast image and video colorization using chrominance blending. Trans. on Image Processing (TIP), 15(5):1120–1129, 2006.

[Zh17] Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A. A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Int. Conf. on computer vision (ICCV). IEEE, pp. 2223–2232, 2017.

[Zh18] Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A.: Self-attention generative adversarial networks. arXiv preprint arXiv:1805.08318, 2018.

[Zh19] Zhang, R.; Zhu, J.-Y.; Isola, P.; Geng, X.; Lin, A. S.; Yu, T.; Efros, A. A.: , Interactive Deep Colorization. `https://github.com/junyanz/interactive-deep-colorization`, 2019.

[ZIE16] Zhang, R.; Isola, P.; Efros, A. A.: Colorful Image Colorization. In: European Conf. on Computer Vision (ECCV). Springer, pp. 649–666, 2016.