MANNET: A LARGE-SCALE MANIPULATED IMAGE DETECTION DATASET AND BASELINE EVALUATIONS

Aditya Singh[2], *Saheb Chhabra*[1,2], *Puspita Majumdar*[1,2], *Richa Singh*[1], *Mayank Vatsa*[1]

[1] IIT Jodhpur, India [2] IIIT-Delhi, India

ABSTRACT

The sharing of fake content on social media platforms has become a major concern. In many cases, the same content with small variations is shared multiple times on different social media platforms. This leads to the circulation of manipulated content on the web. With the rapid advancement in deep learning algorithms, the generation of manipulated images with small variations in original images has become an easy task. These contents raise serious concerns when used for malicious activities. Therefore, detection of manipulated contents is of paramount importance. However, no large-scale dataset having manipulated images generated using both handcrafted and deep learning algorithms is available. Therefore, in this research, we have proposed a large dataset with more than 5.5 million images, termed as **ManNet** dataset. Additionally, we have benchmarked the performance of existing algorithms for manipulated image detection. The experimental results highlight that inter-set (disjoint training testing) evaluations are the major challenge of manipulated image detection.

Index Terms— Image Representation, Manipulated Image Detection, Image Transformation

1. INTRODUCTION

Social media platforms have become the new source of information and are believed to spread false news faster than any other media. Millions of manipulated images and videos are shared multiple times on these platforms on a daily basis, which results in the sudden rise of the volume of manipulated content on these platforms [1]. These manipulated multimedia contents are generated by incorporating slight modifications in the images. Also, with the advancement of Generative Adversarial Networks (GANs) [2], the task of manipulated content generation has become easier. These manipulated content pose serious concerns when used for *pornography, bullying, trolling on social media, and spreading fake news [3]*.

Several datasets [4, 5, 6, 7, 8] and algorithms [9, 10, 11, 12] have been proposed in the literature for conducting research towards detecting manipulated images with small variations. Most of the datasets are limited in size and do not contain manipulated images generated using recent deep learning-based transformations. This demands the need for a dataset to assist the researchers in designing sophisticated algorithms for the detection of manipulated images generated using modern transformations. Therefore, in this research, we have proposed a dataset, termed as **ManNet** dataset having millions of manipulated images generated using *handcrafted and deep learning techniques*. The manipulated content refers to the images generated using small variations in the input image. We have used 24 different image transformations for creating the dataset, which includes 11 non-deep learning and 13 deep learning-based image transformations. Further, we benchmarked the performance of existing algorithms for manipulated content detection by generating the same image representation against similar content images. This will help to identify multiple manipulated content generated from an original image. Fig. 1 illustrates the challenges for obtaining the same image representation for the visually dissimilar manipulated images generated using handcrafted and deep learning-based transformations corresponding to an original image.

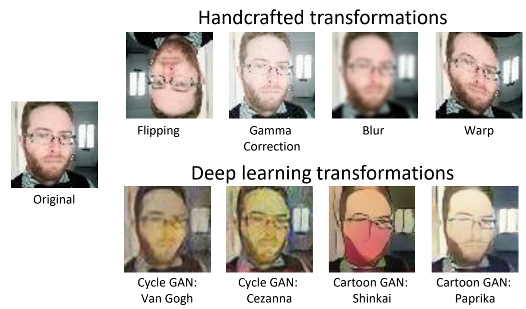

Fig. 1. Illustrating the visual difference in the manipulated images generated using handcrafted and deep learning-based transformations.

The manipulations considered in this research are widely used for duplicate and near-duplicate image generation. In this research, we have considered the manipulations that are widely used for duplicate and near-duplicate image generation. In the literature, some of the widely used datasets for duplicate image detection [13, 14, 15, 16, 17] include, Corel Photo CD Collection [5], MIRFlickr1M and MIRFlickr60K [18], UKBench [19], INRIA [6], and California ND [4]. Duplicate images are created in the Corel Photo CD Collection by applying 50 different transformations (with varying parameters) on each image. The transformations include contrast, severe contrast, format change, rotation, crop, severe crop, intensity, severe intensity, saturation, shear, rotate+crop, and rotate+scale. MIRFlickr1M contains 1M distractor images, including low-resolution images, and MIRFlickr60K contains more than 67K images. The dataset contains more human images and is used to evaluate image search. The UKBench dataset [19] consists of 10200 object images with 640 $\times$ 480 image size. Each object has four images with different viewpoints, rotation, and illumination. INRIA dataset [6] has been created by applying different transformations on scene type images. These transformations include image compression, rotation, cropping, and image resizing. Another dataset for near-duplicate image detection named California ND dataset is proposed by [4]. This dataset contains 701 photos of user's personal photo collection having non-identical near-duplicates of real-world scenes. From the existing datasets, it is observed that the duplicate images (in our case

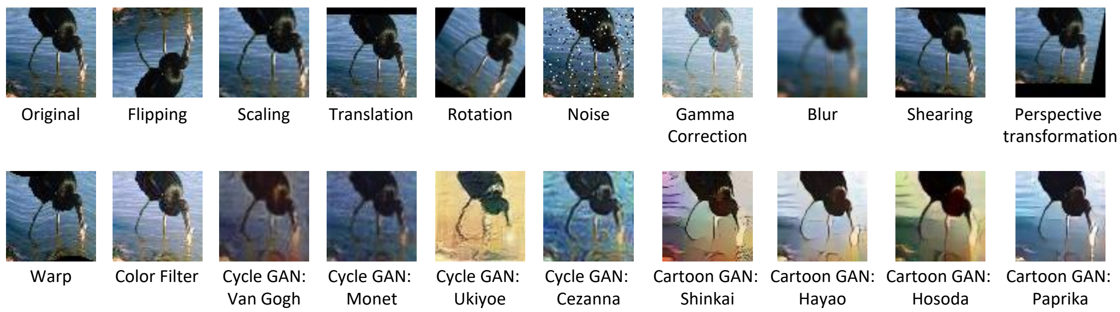

Fig. 2. Sample images of the proposed ManNet dataset with different types of transformations used to generate manipulated images.

manipulated images) in the above-mentioned datasets are limited to handcrafted and vision-based transformations. Algorithms designed on these datasets may not perform well in the real world scenarios where manipulations are performed using other types of transformations, including GANs based transformations. Hence, in the proposed ManNet dataset, we have included both handcrafted and deep learning-based transformations for generating manipulated images.

2. MANNET DATASET

The aim of this research is to detect the manipulated content available on the web by generating a unique image representation for all the variations of an image. We focus on classifying images, which are geometric, intensity, or style transformations of the same original image into a single class. For this purpose, a dataset having such transformations is required because none of the existing works have used a standard dataset for the problem. Thus, as a primary contribution, we have created the ManNet dataset [1] of more than 5.5M images with different transformations. Images from two publicly available datasets, namely, Labeled Faces in the Wild (LFW) [20] and Tiny ImageNet [21] are used to create the proposed dataset. Standard image processing techniques are used for geometric and intensity transformations. For style transformations, GAN based image-to-image translation networks, namely, CycleGAN [22], CartoonGAN [23], and StarGAN [24] are used. These image-to-image translation networks are used to transform an image's style in creative ways. Fig. 2 shows the transformations applied to an image to create manipulated images.

The proposed ManNet dataset consists of four different sets, each having different numbers of images generated using different types of transformations. Set 1 and Set 2 contain manipulated images generated using handcrafted transformations, while Set 3 and Set 4 contains images generated using deep learning-based transformations. Set 2 is a subset of Set 1 that can be used for training with limited resources. Set 3 contains images generated using CycleGAN and CartoonGAN, while Set 4 consists of images generated using StarGAN. Different protocols are used for performance evaluation in each set. In the proposed dataset, each unique image belongs to a different class, and each class has different transformations of that image. The following present the dataset design and the corresponding protocol. Table 1 summarizes the dataset statistics.

Set 1: The first set is created by applying intensity and geometric transformations on the images of the LFW and Tiny ImageNet datasets. We used 11 unique transformations for each image and varied four different parameters of each transformation. The transformations include 'Flipping', 'Scaling', 'Translation', 'Rotation', 'Salt & Pepper Noise', 'Gamma Correction', 'Blurring', 'Shearing', 'Perspective Transform', 'Warping', and 'Filters'. Table 1 shows the details of the parameters varied during the creation of the transformed images. For each original image, 44 (11×4), different transformed images are created, resulting in 45 images (including original image) per class.

For experimental evaluation, the set is divided into disjoint training and testing partitions. For the training set, 500 images are randomly sampled from each of the 200 classes in the Tiny ImageNet dataset, totaling to 100,000 images. Additionally, 500 images are randomly sampled from the LFW dataset. Thus, a total of 100,500 original images are obtained for the training set. Next, 44 different transformations are applied to the original images to create the corresponding manipulated images. As mentioned earlier, each unique image belongs to a different class. Therefore, the training set contains a total of 4,522,500 images belonging to 100,500 classes. For the testing set, 50 images from each class of the Tiny ImageNet dataset and 50 images from the LFW dataset are randomly sampled by excluding the training set samples. Thus, the testing set contains 10,050 original images. For generating manipulated images, we use different parameters (from training) for each transformation and obtain a total of 452,250 images.

Set 2: This set is a sampled version of Set 1 and is about 10% of its size. It is prepared for training with limited resources. Here, 50 images per class of the Tiny ImageNet dataset and 50 images from the LFW dataset are sampled for the training set. On the other hand, 5 images per class of the Tiny ImageNet dataset and 50 images from the LFW dataset are sampled for the testing set. Different transformations are applied to the images of the training and testing sets for manipulated image generation. Similar to Set 1, the parameters used for transformations in the training set are different from the testing set. A total of 452,250 images belonging to 10,050 classes and 45,225 images belonging to 1,005 classes are obtained for the training and testing sets, respectively.

Set 3: The third set is created by applying GAN based style transformations to the images, using 'CycleGAN 'and 'CartoonGAN'. We used 4 transformations, namely, Van-Gogh, Monet, Ukiyoe, and Cezanne for CycleGAN. Similarly, Shinkai, Hayao, Hosoda, and Paprika transformations are used for CartoonGAN. Therefore, 8 different transformed images are obtained corresponding to an original image, resulting in 9 images (including original image) per class.

Similar to Set 2, 50 images per class of the Tiny ImageNet dataset and 50 images from the LFW dataset are sampled for the training set, resulting in a total of 10,050 images. For the testing set,

[1] The dataset and protocol files are available at: http://www.iab-rubric.org/resources/ManNet.html

Table 1. Statistics of the proposed ManNet dataset (5.5 million images).

Type		Training					Testing				
		Original Images	Dataset Images		Transformations	Total Images (Original + Transformed)	Original Images	Dataset Images		Transformations	Total Images (Original + Transformed)
			Tiny ImageNet	LFW				Tiny ImageNet	LFW		
Handcrafted	**Set 1**	100500	100000	500	11 x 4	4522500	10050	10000	50	11 x 4	452250
	Set 2	10050	10000	50	11 x 4	452250	1005	1000	5	11 x 4	45225
GANs	**Set 3**	10050	10000	50	8	90450	1005	1000	5	8	9450
	Set 4	450	0	450	5	2700	100	50	50	52	5200
Total = 5580025											

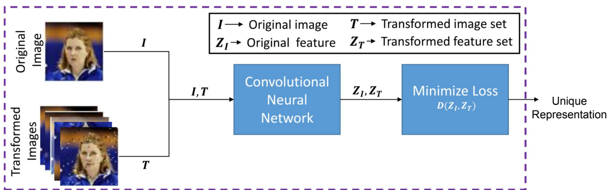

Fig. 3. Pipeline for obtaining unique representation corresponding to original and transformed images.

5 images per class are sampled from the Tiny ImageNet dataset and 5 images from the LFW dataset, totaling 1050 images. After applying style transformations, a total of 90450 and 9450 images are obtained for the training and testing sets, respectively.

Set 4: This set consists of GAN based style transformations of face images using StarGAN. We generated 5 facial transformations, black hair, brown hair, blonde hair, age, and gender, resulting in 6 images (including original image) per class. For this set, 500 face images (sampled from the LFW dataset) of the training set of Set 1 are used. From these images, 450 images are used for the training set, and the remaining 50 are used for the testing set of Set 4. Next, 5 different style transformations of StarGAN are applied to the images of the training set, resulting in a total of 2700 images belonging to 450 classes. For the testing set, we additionally sampled 50 images from random classes of the Tiny ImageNet dataset, totaling 100 original images. 52 different transformations (44 intensity and geometric transformations used in Set 1, and 8 style transformations used in Set 3) are applied to the images sampled from the Tiny ImageNet dataset. On the other hand, 5 style transformations of StarGAN are applied multiple times on the face images of the testing set to obtain 52 transformed images corresponding to an original face image. Thus, the testing set contains 5200 images belonging to 100 classes.

3. BASELINE EXPERIMENTAL RESULTS

We have evaluated the performance of two deep learning algorithms; namely, triplet loss [25] and contrastive loss [26] and one handcrafted algorithm, i.e., spatial histogram [27] on the proposed ManNet dataset. For deep learning based approaches, feature representation of the input images is learned using triplet loss and contrastive loss. On the other hand, spatial histogram is used to extract the feature representation by concatenating the histograms of the patches of an input image. In this research, these feature representations are used to detect manipulated images. Fig. 3 shows the pipeline for obtaining unique representation corresponding to original and transformed images using learning-based algorithm.

For experimental evaluation, two different experiments are performed: *1) Intra Set* and *2) Inter Set*. The first experiment represents the scenario where the network is trained and evaluated on the same set. The second experiment represents the scenario where the network is trained on a specific set and evaluated on other sets. For both the experiments, results on Sets 2 and 3 are shown in the main paper. The rest of the results are shown in the supplementary file [2]. The implementation details and protocol for training the networks for all the experiments are discussed below.

Implementation Details: We have segregated the implementation details of deep learning and handcrafted algorithms.

Triplet & Contrastive Loss: A deep CNN network is used for generating the embeddings for the images. The network consists of 4 convolutional layers, each followed by batch normalization and ReLu activation. Finally, the fully connected layers output an embedding of dimension 32. We also use dropout to prevent the network from over-fitting[28]. During training with contrastive loss, the network is trained for 10 epochs with Adam optimizer. The learning rate is set to 0.001 and batch size to 16. A margin $m = 8$ is used for the experiments. During triplet training, the network is trained for 10 epochs with a learning rate of 0.001 and batch size 16 . Adam optimizer is used and margin α is set to 8 during training. Code is implemented in Tensorflow. All the experiments are performed on Nvidia GeForce GTX 1080 Ti.

Spatial Histogram: Each image is divided into 4 cells. For each cell, the hash that we generate is of length 120, and consequently, the length of the hash code of the whole image is of length 480. Because all values of the embedding are binary (0 or 1), the embedding can be stored in 480 bits or 60 bytes.

Protocol: Experiments are performed on disjoint training and testing splits corresponding to each set. To evaluate the performance of the trained network on the testing set, a single image gallery set and a probe set with multiple images are used. Here, our aim is to retrieve the original image used to generate the manipulated images. Original and its manipulated versions must have the same image representation. Therefore, the gallery set consists of original images corresponding to each class. The probe set consists of all transformed images corresponding to each class. For each image in the probe set, the top k images in the gallery set are identified. These top k images correspond to the smallest Euclidean distance between the feature vector of the gallery images and the probe image. To compute the accuracy on each set, the fraction of images in the probe set having its corresponding image in the top k closest images of the gallery set is taken. We have shown the results at $k = 1$ in the main paper and reported the results at $k = 1$, $k = 5$, and $k = 10$ in the supplementary file. We have reported the overall accuracy by taking the average accuracy corresponding to each transformation.

3.1. Intra Set Experiments

Tables 2 and 3 show the results of the intra set experiment for Set 2 and Set 3, respectively. It is important to note that Set 2 consists

[2]Supplementary file can be found at: `https://github.com/Saheb-Chhabra/ICASSP`

Table 2. Results for the intra set experiment. The network is trained and evaluated on Set 2. The accuracy is computed for the top 1 closest matches.

Transformations	Triplet Loss	Contrastive Loss	Spatial Histogram
Flipping	96.04%	61.12%	33.86%
Scaling	99.60%	94.43%	83.66%
Translation	99.85%	98.11%	96.84%
Rotation	98.83%	96.54%	71.62%
Salt & Pepper Noise	99.88%	93.41%	92.29%
Gamma Correction	93.43%	9.33%	0.67%
Blurring	99.25%	96.44%	94.63%
Shearing	99.85%	98.31%	87.69%
Perspective Transform	98.73%	90.50%	86.77%
Warping	100.00%	100.00%	98.76%
Filters	92.81%	22.01%	33.48%
Overall	**98.03%**	**78.20%**	**70.93%**

Table 3. Results for the intra set experiment. The network is trained and evaluated on Set 3. The accuracy is computed for the top 1 closest matches.

Transformations	Triplet Loss	Contrastive Loss	Spatial Histogram
Cycle GAN			
Van-gogh	99.60%	94.03	5.07%
Monet	100.00%	97.61	3.78%
Ukiyoe	98.21%	80.10	0.70%
Cezanna	99.80%	97.21	4.08%
Cartoon GAN			
Shinkai	98.71%	92.04	7.76%
Hayao	98.41%	91.24	6.27%
Hosoda	99.00%	90.05	3.38%
Paprika	99.60%	94.23	8.56%
Overall	**99.17%**	**92.95**	**4.94%**

of images with only geometric and intensity transformations, while Set 3 consists of images with deep learning-based transformations. It is observed that for both sets, learning-based algorithms perform better. For geometric and intensity transformations, the accuracy is lowest for 'Flipping', 'Gamma Correction', and 'Filters' transformations. For instance, the accuracy is 61.12%, 9.33%, and 22.01%, corresponding to these transformations using contrastive loss. Thus, contrastive loss is not robust against generation of unique image representation for these transformations. *This highlights that obtaining unique image representation for some transformations is difficult compared to other transformations.*

3.2. Inter Set Experiments

This experiment represents the real-world scenario where the distribution of testing images are different from training images. Tables 4 and 5 show the results of the inter set experiment for Set 2 and Set 3, respectively. It is observed that deep learning algorithms perform worse when the network is trained on Set 2 and tested on Set 3. For triplet loss and contrastive loss, an overall accuracy of 36.06%, and 17.37% is obtained, while spatial histogram algorithm achieves an accuracy of 4.94%. On the other hand, deep learning algorithms perform better on Set 2 when the network is trained on Set 3. This shows that the network training on Set 3 is more generalizable. It is important to note that Set 3 consists of GANs generated (deep learning) transformed images, while Set 2 contains images generated using simple image transformations (handcrafted). This shows

Table 4. Results for the inter set experiment. The network is trained on Set 2 and evaluated on Set 3. The accuracy is computed for the top 1 closest matches.

Transformations	Triplet Loss	Contrastive Loss	Spatial Histogram
Cycle GAN			
Van-gogh	42.69%	9.95%	5.07%
Monet	32.74%	6.07%	3.78%
Ukiyoe	12.94%	0.10%	0.70%
Cezanna	40.60%	8.26 %	4.08%
Cartoon GAN			
Shinkai	41.59%	8.56 %	7.76%
Hayao	42.49%	9.85%	6.27%
Hosoda	37.61%	6.07%	3.38%
Paprika	37.91%	7.56%	8.56%
Overall	**36.06%**	**17.37%**	**4.94%**

Table 5. Results for the inter set experiment. The network is trained on Set 3 and evaluated on Set 2. The accuracy is computed for the top 1 closest matches.

Transformations	Triplet Loss	Contrastive Loss	Spatial Histogram
Flipping	30.02%	32.71%	33.86%
Scaling	85.05%	74.83%	83.66%
Translation	50.52%	55.42%	96.84%
Rotation	29.73%	36.34%	71.62%
Salt & Pepper Noise	99.93%	99.53%	92.29%
Gamma Correction	96.47%	45.02%	0.67%
Blurring	100.00%	99.90%	94.63%
Shearing	100.00%	99.85%	87.69%
Perspective Transform	56.77%	58.26%	86.77%
Warping	96.72%	95.47%	98.76%
Filters	99.95%	94.85%	33.48%
Overall	**76.83%**	**72.02%**	**70.93%**

that a network trained with images generated using deep learning algorithms generalizes better compared to the one trained with handcrafted transformed images. *In the experiments, we observe that inter set evaluations are more challenging for baseline approaches, and future research should focus on these variations.*

4. CONCLUSION

This research presents a large scale dataset with more than 5.5 million images for manipulated image detection. The proposed ManNet dataset is created by generating manipulated images using handcrafted and deep learning-based approaches. None of the existing datasets contain images generated using traditional and modern transformation techniques. The proposed ManNet dataset may assist researchers in designing algorithms for original image retrieval along with manipulated image detection. It should also encourage researchers to develop a universal algorithm for detecting different types (deep learning and handcrafted) of manipulated images.

Acknowledgements

This research is supported through a grant from BPR&D, Ministry of Home Affairs, India. P. Majumdar is partly supported by DST Inspire Ph.D. Fellowship. M. Vatsa is partially supported through Swarnajayanti Fellowship.

5. REFERENCES

[1] Puspita Majumdar, Akshay Agarwal, Mayank Vatsa, and Richa Singh, "Facial retouching and alteration detection," in *Handbook of Digital Face Manipulation and Detection*, pp. 367–387. Springer, 2022.

[2] Alec Radford, Luke Metz, and Soumith Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[3] Anubhav Jain, Puspita Majumdar, Richa Singh, and Mayank Vatsa, "Detecting gans and retouching based digital alterations via dad-hcnn," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 672–673.

[4] Amornched Jinda-Apiraksa, Vassilios Vonikakis, and Stefan Winkler, "California-nd: An annotated dataset for near-duplicate detection in personal photo collections," in *International Workshop on Quality of Multimedia Experience*. IEEE, 2013, pp. 142–147.

[5] "Corel Photo CD Collection," `http://apps.corel.com/dell/paintshop/uk/photo_album_6/download.html`.

[6] "INRIA dataset," `http://lear.inrialpes.fr/~jegou/data.php`.

[7] Bihan Wen, Ye Zhu, Ramanathan Subramanian, Tian-Tsong Ng, Xuanjing Shen, and Stefan Winkler, "Coverage—a novel database for copy-move forgery detection," in *IEEE International Conference on Image Processing*. IEEE, 2016, pp. 161–165.

[8] Jing Dong, Wei Wang, and Tieniu Tan, "Casia image tampering detection evaluation database," in *IEEE China Summit and International Conference on Signal and Information Processing*. IEEE, 2013, pp. 422–426.

[9] Puspita Majumdar, Akshay Agarwal, Richa Singh, and Mayank Vatsa, "Evading face recognition via partial tampering of faces," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

[10] Zehao Chen and Hua Yang, "Attentive semantic exploring for manipulated face detection," in *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2021, pp. 1985–1989.

[11] Chao Yang, Zhiyu Wang, Huawei Shen, Huizhou Li, and Bin Jiang, "Multi-modality image manipulation detection," in *IEEE International Conference on Multimedia and Expo*. IEEE, 2021, pp. 1–6.

[12] Shuaibo Li, Shibiao Xu, Wei Ma, and Qiu Zong, "Image manipulation localization using attentional cross-domain cnn features," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[13] KK Thyagharajan and G Kalaiarasi, "A review on near-duplicate detection of images using computer vision techniques," *Archives of Computational Methods in Engineering*, pp. 1–20, 2020.

[14] Yi Zhang, Shizhou Zhang, Ying Li, and Yanning Zhang, "Single-and cross-modality near duplicate image pairs detection via spatial transformer comparing cnn," *Sensors*, vol. 21, no. 1, pp. 255, 2021.

[15] G Kalaiarasi, M Maheswari, Prathima Devadas, and M Selvi, "Detecting near duplicate images using hellinger correlation coefficient," *International Journal of Advanced Research in Engineering and Technology*, vol. 11, no. 6, 2020.

[16] Weiming Hu, Yabo Fan, Junliang Xing, Liang Sun, Zhaoquan Cai, and Stephen Maybank, "Deep constrained siamese hash coding network and load-balanced locality-sensitive hashing for near duplicate image detection," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4452–4464, 2018.

[17] Emanuel Poppa, *A benchmark protocol for evaluating near duplicate detection performance*, Ph.D. thesis, Politecnico di Torino, 2020.

[18] "MIRFlickr dataset," `http://press.liacs.nl/mirfickr/`.

[19] "UKBench dataset," `https://archive.org/details/ukbench`.

[20] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," 2007.

[21] "Tiny ImageNet," `https://tiny-imagenet.herokuapp.com`.

[22] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *IEEE Conference on Computer Vision*, 2017, pp. 2223–2232.

[23] Yang Chen, Yu-Kun Lai, and Yong-Jin Liu, "Cartoongan: Generative adversarial networks for photo cartoonization," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9465–9474.

[24] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8789–8797.

[25] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[26] Raia Hadsell, Sumit Chopra, and Yann LeCun, "Dimensionality reduction by learning an invariant mapping," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2006, vol. 2, pp. 1735–1742.

[27] Shijun Xiang, Hyoung-Joong Kim, and Jiwu Huang, "Histogram-based Image Hashing Scheme Robust Against Geometric Deformations," in *9th Workshop on Multimedia & Security*, 2007.

[28] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," in *The Journal of Machine Learning Research 15.1*, 2014.