An Effective CNN-based Approach for Synthetic Face Image Detection in Pre-Social and Post-Social Media Context

Protyay Dey[0009-0002-3859-4257], Abhilasha S. Jadhav[0009-0008-9161-9509], and Kapil Rana[0000-0001-8063-8087]

Indian Institute of Technology Ropar, Rupnagar, Punjab, India

Abstract. The proliferation of image manipulation techniques, including DeepFake technology, has posed significant threats to the authenticity and credibility of images. Accurately classifying real and fake images has become crucial in fields such as forensics, security, and media authentication. However, detecting fake images during downloading and uploading from social networks is even more challenging. In this paper, we present an approach based on the EfficientNet model to learn discriminative features for classifying real and synthetic face images shared on social networks. We conducted extensive experiments using the TrueFace dataset, which comprises real and synthetic facial images shared on three major social media platforms. We employed the EfficientNet-B2 model trained on a combination of pre-social and post-social images from the TrueFace dataset. The presented approach outperforms all other methods, achieving accuracies of 99.98%, 100%, and 100% for images shared on Facebook, Telegram, and Twitter. This approach demonstrates exceptional performance when evaluated on a distinct dataset of images shared on social media platforms, separate from the images used for training.

Keywords: EfficientNetB2 · TrueFace dataset · DeepFakes · Image Forensics · Fake face identification

1 Introduction

With recent advances in mobile technology and the Internet, social media platforms have become an important channel for sharing photos and videos publicly. It is also a major source of information because a person pays more attention and is more actively engaged with visual content, substantially improving the probability of sharing such content. Digitally created deep fake images have caused great public concern. Advanced image processing tools based on generative adversarial networks (GAN) are available to everyone. These GANs can create very realistic fakes, especially for images. Recent studies have shown that even experts have difficulty distinguishing between real faces and synthetic faces generated by GANs, which are incredibly realistic [1]. This has led to the emergence of deep fakes that threaten the credibility of visual content online, especially on social media platforms. Fake profiles on social networks have already caused damage and attempts have been made to manipulate information about events such as the Russian-Ukrainian war [2]. This is an alarming trend of image manipulation that is causing the spread of misinformation and fake information [3].

Therefore, it is crucial to prioritize the development of methods that ensure trust in images shared on social media platforms. Several works in image forensics have focused on detecting image manipulation and identifying digital sources [4], [5]. These studies have shown promising results in controlled environments [6]. Recently, research communities have attempted to apply forensic methods to real-world applications, including social media platforms where digital media are shared online [7]. In the context of deep forgeries, researchers have demonstrated their ability to distinguish synthetic images from real images by training data models to detect traces of [8], [9] generators.

Creating synthetic images involves different generation methods, and data-driven detectors often face challenges when the test and training generators are distinct. Fake content tends to gain popularity on social media before forensic analysis can be performed. Post-processing used by sharing platforms can mask and weaken detection methods [10]. Post-processing functions of social networks (compression, resizing) can change the statistics specific to the image and affect the evaluation. It is crucial to develop a method to distinguish between real images posted on popular sharing sites and content created artificially using sophisticated methods. In this work, we present a CNN method based on EfficientNet [11] for real and synthetic image face recognition. Importantly, the presented method can distinguish between real and synthetic facial images shared on three major social media sites. The main contributions of this paper are as follows:

- We present a CNN-based EfficientNet-B2 model specifically designed for classifying real and synthetic facial images shared on social networks. This model includes compound scaling, which allows us to balance accuracy and computational efficiency.
- Extensive experiments were conducted on the TrueFace dataset containing real and synthetic facial images shared on popular social media platforms such as Facebook, Twitter, and Telegram.
- The robustness of our approach was evaluated in a cross-dataset environment, demonstrating that it could generalize and accurately classify real and synthetic images shared across different social media platforms.
- We performed a comprehensive evaluation of the proposed method by comparing it with alternative CNN-based approaches such as VGG19, ResNet50, DenseNet121, and MnasNet.

The rest of the paper is organized as follows. Section 2 summarizes the related work. Section 3 explains the EfficientNet-B2 model for real and synthetic facial image classification. Experiments and results including the comparison with comparative models are presented in Section 4. Finally, we conclude the paper in Section 5.

2 Related Works

Researchers have focused on media forensics to address the challenges of maintaining the authenticity of manipulated visual media and shared images and videos. However, if the test generator is different from the training generator, it becomes difficult to detect fake content as it spreads quickly on social networks and evades forensic analysis. Benchmark datasets containing synthetic and real images are essential for evaluating forensic detectors in real-world scenarios.

Related works on DeepFake image classification have emerged as an important research area in recent years. The goal of deepfake image classification is to develop algorithms and models capable of distinguishing between authentic and manipulated images. The survey of [3] provides a comprehensive overview of various face manipulation techniques like Extensive Face Synthesis, Identity Swap, Attribute Manipulation, and Expression Swap, also includes deepfake generation methods, and presents an extensive review of deepfake detection methods. The study described in [12] utilized the ResNet50 model as the primary architecture to achieve their results. Initially, the researchers fine-tuned the baseline model using pre-social data. Subsequently, the fine-tuned model was further adjusted to handle datasets from Facebook, Twitter, Telegram, and a combined dataset. This multi-step fine-tuning process enabled the model to adapt and improve its performance on each specific dataset. In addition, the researchers conducted cross-testing across different datasets to thoroughly analyze and evaluate the obtained results.

The works on StyleGAN and StyleGAN2 by Karras et al [13] have significantly contributed to the field of generative adversarial networks (GANs) and image synthesis. This work introduced StyleGAN, which incorporates a style-based generator architecture that separates the high-level and low-level structure information of images. This enables control over different aspects of the generated images, such as their style, pose, and content. In this study [14], the authors perform an in-depth analysis of the image quality produced by StyleGAN. They propose improvements to the model, including modifications to the generator's architecture and loss functions, resulting in higher-fidelity generated images.

In recent years, the field of convolutional neural networks (CNNs) has witnessed significant advancements in model scaling techniques. The conventional approach involves independently scaling the depth, width, or resolution of CNNs, but this often results in sub-optimal performance or excessive computational requirements. To address these limitations [11] presents a novel compound scaling method that uniformly scales the network width, depth, and resolution while maintaining a constant resource constraint. The authors propose a scaling formula, referred to as compound scaling, which involves finding optimal scaling coefficients for each dimension. These coefficients are determined using a grid search technique, with the constraint being a specific resource metric such as the number of floating-point operations (FLOPs). To further enhance computational efficiency, EfficientNet introduces a new scaling technique, inspired by the MobileNetV2 [15]. This method focuses on reducing the number of parameters without compromising performance. It achieves this by employing a combination of depth-wise separable convolutions and inverted residual blocks.

The EfficientNet models are trained using standard data augmentation techniques, including image scaling and random cropping, and are evaluated on various benchmark datasets such as ImageNet[16]. The results demonstrate that EfficientNet achieves state-of-the-art performance while being more computationally efficient compared to previous scaling approaches. The authors achieve superior performance with fewer computational resources. The proposed methodology represents a significant contribution to the field of model scaling for CNNs and paves the way for future advancements in efficient neural network architectures.

3 Proposed Methodology

This section describes the approach for the detection of real and synthetic images, given an input face image. In the presented approach an RGB image is passed as an input to the EfficientNet-B2 model for feature extraction and further classification. EfficientNet-B2 is a deep CNN model from the EfficientNet class of CNN models. It is specifically designed to achieve a balance between model size and computational efficiency while delivering high performance in image classification problems. The architecture of EfficientNet-B2 is built upon a combination of key components, including convolutional layers, skip connections, and efficient network scaling techniques. The architecture of the EfficientNet-B2 model is shown in Fig. 2 and the building block of the model is shown in Fig. 1.

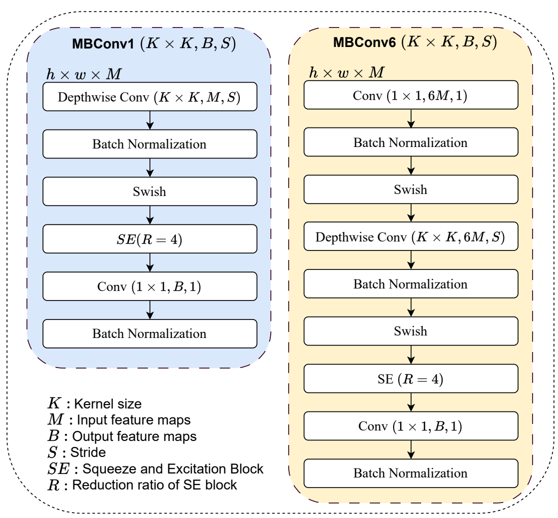

Fig. 1. The architecture of building blocks of EfficientNet.

The model consists of multiple blocks, each containing a sequence of convolutional layers, batch normalization layers, and non-linear activation functions. EfficientNet-B2 adopts compound scaling, a technique that uniformly scales the network's width, depth, and resolution while adhering to a consistent resource constraint. This scaling method enables the model to efficiently capture and represent the complex features and patterns inherent in the dataset. By employing a depth-wise separable convolution strategy, EfficientNet-B2 decomposes the traditional convolution operation into two distinct operations: a depth-wise convolution and a point-wise convolution. This approach effectively decreases the number of parameters and computational complexity, resulting in enhanced efficiency. These architectural choices highlight EfficientNet-B2's ability to trade off between model performance and computational requirements, making it well-suited for image classification problems. EfficientNet-B2 has a similar architecture to other different CNN models in the EfficientNet family (EfficientNet-(B0, B1, B2, B3, B4, B5, B6, and B7)) but with specific scaling factors determined based on the B2 configuration. The input image is then passed to the EfficientNet-B2 model. The model extracts low-level and high-level features related to real and synthetic facial images for classification.

Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels $\hat{C}_i$	#Layers $\hat{L}_i$
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	48	2
5	MBConv6, k3x3	28×28	88	3
6	MBConv6, k5x5	14×14	120	3
7	MBConv6, k5x5	14×14	208	4
8	MBConv6, k3x3	7×7	352	1
9	Conv1x1 & Pooling	7×7	1408	1
10	FC	1408	2	1

Fig. 2. Architecture and parameters of EfficientNet-B2 model [11].

4 Experiments and Results

In this section, we first present information about the datasets and settings used for our experiments. Then, we share the results obtained from our proposed approach and compare them with other comparative methods. In the subsequent subsections, we provide detailed insights into our experimental settings, describing the specific models utilized, the preprocessing techniques applied, and any data augmentation methods employed. Additionally, we outline the experiments conducted, present the evaluation metrics used, and discuss the obtained results in terms of model performance and classification accuracy.

Additionally, we conducted an ablation study to ensure optimal model selection for our image classification task in a post-social media context. In particular, we investigate the performance of three different versions of EfficientNet: B0, B1, and B2. Through a comprehensive and systematic evaluation of these models within our experimental setup, we aim to discern the most effective model among the various variations.

4.1 Dataset

We evaluate our proposed approach on the TrueFace dataset [12]. To the best of our knowledge, the TrueFace dataset is currently the only publicly available dataset that includes both real and synthetic facial images shared on social media platforms. It provides images to perform experiments to determine the difference between real and synthetic images. It consists of 210,000 face images divided into two sets: pre-social and post-social. The pre-social set consists of 70,000 real and 80,000 synthetic images. All the images in the pre-social set are of size $1024 \times 1024 \times 3$. The synthetic images were created using two popular generator models called StyleGAN and StyleGAN2. Each GAN model generated 40,000 synthetic images. The real images present in the TrueFace dataset are included from the FFHQ dataset [13]. The FFHQ Dataset was originally designed as a benchmark for Generative Adversarial Networks (GANs) [17], as a reliable source of authentic facial images. The sample images of the pre-social set are shown in 3. The post-social dataset contains 60,000 images shared on three popular social media platforms: Facebook, Twitter, and Telegram where each social media platform has 20000 images: 10000 real images and 10000 synthetic images. The reason there are three different versions of a single image is that each social media platform applies different post-processing to the uploaded images while storing them on their servers. Analyzing all three variants provides us with useful insights into a model's behavior when provided with images having different post-processing applied to them.

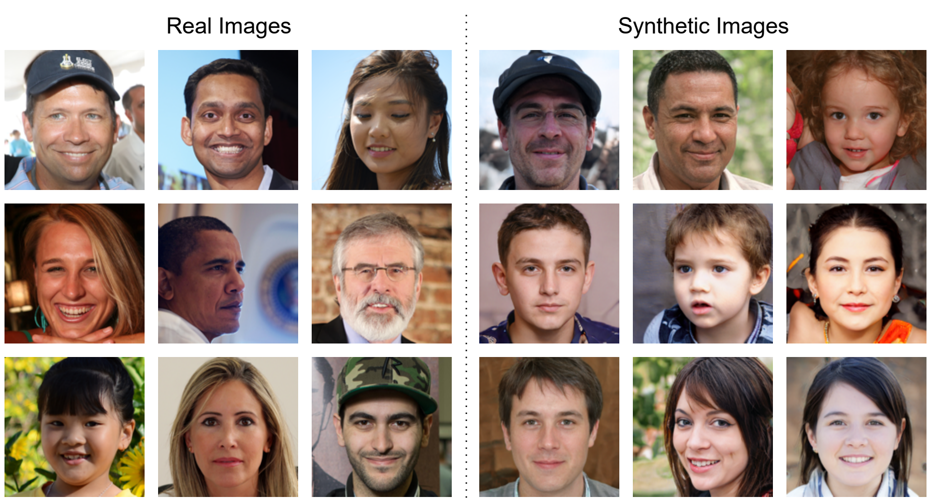

Fig. 3. Illustration of sample images from the pre-social set of TrueFace dataset.

4.2 Experimental Setting

The experiments were conducted on a system consisting of an A100 Tensor Core GPU of 40GB memory and a 3.39 GHz AMD-EPYC-7742 CPU with 64GB RAM. The TrueFace dataset has been divided into two sets: the training set and the testing set in a proportion of 70% and 30%, respectively. We employed a seed-based random splitting technique to distribute the images evenly across both sets to ensure reproducibility and integrity. This approach ensured that our models were trained and evaluated consistently on a predetermined set of images throughout our experiments. The purpose of implementing this strategy was to mitigate potential biases or inconsistencies that may arise from using different subsets of images for training and testing. All images were resized to a dimension of $224 \times 224 \times 3$. The models aimed to perform binary classification, distinguishing between real and synthetic input images. The experiments were conducted using the PyTorch 2.0.1 framework and CUDA version 11.7 for GPU acceleration. A mini-batch size of 64 was used across all experiments. The models were trained for a maximum of 100 epochs, employing an early stopper with a patience factor of 10. The early stopper monitored the loss to determine when to stop training. We picked the model with maximum testing accuracy. We utilized the AdamW optimizer with an initial learning rate of 3×10^{-4}, benefiting from its regularization capabilities by separating weight decay from exponential moving averages compared to traditional Adam. Additionally, a learning rate scheduler was applied, with a minimum learning rate set to 10^{-6} and a patience of 10.

4.3 Results and analysis

We thoroughly evaluate our approach through three experiments using different combinations of the TrueFace dataset to assess its effectiveness. In the first experiment, models were trained on the pre-social set of the TrueFace dataset. The second experiment involved training on a combined variant of the TrueFace dataset, which included both pre-social and post-social set images. Lastly, cross-dataset testing was conducted, where models were trained exclusively on each social media platform image and evaluated on the same and other social media platform images. We conduct a set of experiments to evaluate the performance of the proposed approach along with a comparative analysis with different CNN-based approaches including VGG19 [18], ResNet50 [19], DenseNet12 [20], MnasNet [21].

Table 1. Comparative analysis of different CNN based methods on pre-social set of TrueFace dataset.

Models	# Parameters	Accuracy	F1-score
VGG19	144M	99.95	99.95
ResNet50	26M	99.71	99.71
DenseNet121	7.97M	99.82	99.83
MnasNet	4M	99.88	99.88
EfficientNet-B2	9.2M	**99.97**	**99.97**

Results on pre-social dataset In this experiment, the proposed approach is evaluated on the pre-social set of the TrueFace dataset. The main focus of the performed experiment is to achieve optimal results by leveraging lightweight models with fewer parameters. The proposed method utilizing EfficientNet-B2 [11] provides the maximum accuracy of 99.97% and F1-score of 99.97% on the pre-social set as shown in Table 1. This can be attributed to its advanced architecture, parameter efficiency transfer learning capabilities, and effective optimization techniques. The architectural design of EfficientNet B2 incorporating depth, width, and resolution scaling allows for accurate capturing and representation of crucial image features. With a smaller number of parameters, EfficientNet-B2 achieves better generalization and effectively mitigates overfitting issues resulting in improved classification accuracy. Enhancing its adaptability for tasks like real vs synthetic image classification. Furthermore, the utilization of the AdamW optimizer significantly enhances regularization and generalization by decoupling weight decay from exponential moving averages.

Table 2. Accuracy of various models when trained over the combined dataset and cross-tested across different splits of TrueFace dataset.

Models	Pre-Social	Telegram	Twitter	Facebook	Combined
VGG19	99.09	99.90	99.92	99.87	99.73
ResNet50	97.90	99.82	99.82	99.78	99.39
DenseNet121	98.94	99.90	99.92	99.93	99.73
MnasNet	99.40	99.93	99.93	99.92	99.79
EfficientNet-B2	**99.85**	**100**	**100**	**99.98**	**99.96**

Results on combined dataset In this experiment, we combined the pre-social and post-social sets, each containing an equal number of images. The post-social dataset consisted of 20,000 images from each social media platform (Facebook, Twitter, and Telegram), resulting in a total of 60,000 images. Additionally, we included 20,000 images from the pre-social dataset. Our aim is to propose and evaluate a robust approach that can perform real vs synthetic image classification even if an image is post-processed using social media platforms. We conducted tests on both the pre-social and post-social test datasets to perform a comparative analysis. This approach allowed us to assess how well the models performed on images from different sources and gain a better understanding of their ability to generalize across diverse datasets. Table 2 shows the accuracy of comparative approaches when tested on different splits of the TrueFace dataset. From the table 2, it can be that the proposed approach consistently achieves high accuracy across all splits of the TrueFace dataset. Specifically, it demonstrates remarkable performance in classifying real and synthetic images, with accuracy ranging from 99.85% to 100% on the different splits of the TrueFace dataset. This indicates its ability to accurately differentiate between real and synthetic images shared across diverse social media platforms.

Results on post-social dataset In this experiment, we evaluate the robustness of the proposed approach. All the comparative methods are trained on each social media platform individually and evaluated on different social media platforms (cross-dataset). The presence of unknown post-processing operations with varying parameters across different social media platforms can introduce uncertainty and potentially impact the performance of a naively trained CNN model. Furthermore, This experiment holds significance as it demonstrates the effectiveness of our proposed approach in capturing intrinsic features associated with both real and synthetic images. Table 3 shows that the proposed approach utilizing EfficientNet-B2 has performed significantly better than all other comparative methods across all the cross datasets settings. Considering different social media platforms, the proposed approach performs better in the case of Telegram. The proposed approach demonstrates exceptional accuracy of 99.05%, 99.15%, 99.55%, and 99.67% on pre-social, Telegram, Twitter, and Facebook images, respectively, when trained on Telegram images. EfficientNet ability to adapt to different resolutions and compression techniques through compound scaling enables it to effectively learn and extract meaningful features from images with varying quality. This adaptability, combined with its optimized use of computational resources, contributes to its superior performance in classifying images from diverse sources, making it a reliable choice for handling images with different resolutions and JPEG compression levels.

Table 3. Accuracy of various models trained on different datasets and tested on different social media platforms.

Training dataset	Models	Pre-Social	Telegram	Twitter	Facebook	Overall
Telegram	VGG19	96.84	96.78	93.80	97.10	96.15
	ResNet50	96.68	96.43	98.65	98.75	97.65
	DenseNet121	98.40	98.18	99.17	99.25	98.75
	MnasNet	95.80	95.58	97.73	98.67	96.85
	EfficientNet-B2	**99.05**	**99.15**	**99.55**	**99.67**	**99.38**
Twitter	VGG19	93.11	97.03	97.22	97.55	96.28
	ResNet50	93.59	97.93	97.23	98.27	96.74
	DenseNet121	96.18	98.83	98.20	99.18	98.11
	MnasNet	95.49	97.93	94.82	96.38	96.09
	EfficientNet-B2	**97.97**	**99.57**	**99.17**	**99.52**	**99.10**
Facebook	VGG19	93.75	98.22	97.63	93.67	95.97
	ResNet50	95.46	98.72	98.72	95.22	97.03
	DenseNet121	98.00	99.38	99.32	97.93	98.73
	MnasNet	97.20	99.20	98.87	97.25	98.14
	EfficientNet-B2	**98.65**	**99.68**	**99.67**	**98.87**	**99.25**

4.4 Ablation Study

In our study, we performed an ablation study on the TrueFace dataset to evaluate the performance of different variants (B0, B1, and B2) of the EfficientNet models in a post-social scenario. This study aimed to identify the most effective model for classifying synthetic and real images in the context of post-social media images, especially on the subset of telegram images because it keeps the images as close as possible to their original form and uses the smallest amount of JPEG compression [12]. We trained and evaluated three variants of the EfficientNet models, namely B0, B1, and B2, using the post-social image data. Each model was trained using the above-mentioned experimental setup.

Table 4. Accuracy of EfficientNet models when trained over the Telegram dataset and cross-tested across different sets of TrueFace dataset.

Models	Pre-Social	Telegram	Twitter	Facebook	Combined	Average
EfficientNet-B0	99.09	98.98	98.85	99.63	99.14	99.14
EfficientNet-B1	99.41	99.38	97.50	99.55	98.93	98.95
EfficientNet-B2	99.05	99.15	99.55	99.67	99.38	**99.36**

Based on the results presented in Table 4, it is clear that the performance of the three EfficientNet variants (B0, B1, and B2) was at par with each other but among these variants, EfficientNet B2 proved to be a superior performer, showing significant improvements when compared to B0 and B1.

The superiority of EfficientNet B2 demonstrates its suitability to effectively distinguish between synthetic and real images in post-social media. The model's exceptional performance demonstrates its robustness in capturing complex patterns and features in images shared on various social media platforms such as Twitter, Facebook, and Telegram.

5 Conclusion

This paper presents an approach based on the EfficientNet model for the classification of real and synthetic images shared on various social media networks. This approach is evaluated using the comprehensive TrueFace dataset, which encompasses a wide range of real and synthetic face images shared on popular platforms such as Facebook, Telegram, and Twitter. Through extensive experiments and evaluations, our proposed approach outperforms existing methods in the field, achieving remarkable accuracy rates across different social media platforms. Specifically, we achieved outstanding accuracy of 99.98%, 100%, and 100% on images shared through Facebook, Telegram, and Twitter respectively.

The results demonstrate the effectiveness and robustness of our approach in accurately classifying synthetic and real images shared on social media. Furthermore, our approach showcases its superior performance in a cross-social media platform setting, where it successfully classifies images shared on social media platforms that differ from the ones used for training. This capability highlights the adaptability and generalizability of our presented approach, making it suitable for real-world applications where images originate from diverse sources. we conducted a series of comprehensive experiments on the TrueFace dataset. We believe that our work opens up new possibilities for addressing the challenges posed by image manipulation and deepfake proliferation on social media platforms. In the future, potential areas of focus may include conducting real-time image validation studies, development of an application for image classification, expanding the dataset with more diverse social media platform images and real-world data, and by considering the challenges posed by sharing of images across multiple social media networks and adversarial attacks.

References

1. Lago, F., Pasquini, C., Böhme, R., Dumont, H., Goffaux, V., Boato, G.: More real than real: A study on human visual perception of synthetic faces [applications corner]. IEEE Signal Processing Magazine **39**(1), 109–116 (2021)
2. Kshetri, N., DeFranco, J.F., Voas, J.: Is it live, or is it deepfake? Computer **56**(07), 14–16 (2023)
3. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., Ortega-Garcia, J.: Deepfakes and beyond: A survey of face manipulation and fake detection. Information Fusion **64**, 131–148 (2020)
4. Rana, K., Singh, G., Goyal, P.: MSRD-CNN: Multi-scale residual deep CNN for general-purpose image manipulation detection. IEEE Access **10**, 41267–41275 (2022)
5. Singh, G., Goyal, P.: SDCN2: A shallow densely connected CNN for multi-purpose image manipulation detection. ACM Transactions on Multimedia Computing, Communications and Applications **18**(3s), 1–22 (2022)
6. Verdoliva, L.: Media forensics and deepfakes: an overview. IEEE Journal of Selected Topics in Signal Processing **14**(5), 910–932 (2020)
7. Rana, K., Singh, G., Goyal, P.: SNRCN2: Steganalysis noise residuals based CNN for source social network identification of digital images. Pattern Recognition Letters **171**, 124–130 (2023)
8. Marra, F., Saltori, C., Boato, G., Verdoliva, L.: Incremental learning for the detection and classification of gan-generated images. In: 2019 IEEE international workshop on information forensics and security (WIFS). pp. 1–6. IEEE (2019)
9. Nguyen, T.T., Nguyen, Q.V.H., Nguyen, D.T., Nguyen, D.T., Huynh-The, T., Nahavandi, S., Nguyen, T.T., Pham, Q.V., Nguyen, C.M.: Deep learning for deepfakes creation and detection: A survey. Computer Vision and Image Understanding **223**, 103525 (2022)
10. Pasquini, C., Brunetta, C., Vinci, A.F., Conotter, V., Boato, G.: Towards the verification of image integrity in online news. In: 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). pp. 1–6. IEEE (2015)
11. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114. PMLR (2019)
12. Boato, G., Pasquini, C., Stefani, A.L., Verde, S., Miorandi, D.: TrueFace: A dataset for the detection of synthetic face images from social networks. In: 2022 IEEE International Joint Conference on Biometrics (IJCB). pp. 1–7. IEEE (2022)
13. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4401–4410 (2019)
14. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of stylegan. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8110–8119 (2020)

15. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)
16. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
17. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Communications of the ACM **63**(11), 139–144 (2020)
18. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
19. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
20. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017)
21. Tan, M., Chen, B., Pang, R., Vasudevan, V., Sandler, M., Howard, A., Le, Q.V.: MnasNet: Platform-aware neural architecture search for mobile. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2820–2828 (2019)