Face Forgery Detection

Yina He[1](31520231154287), Yufeng Huang[1](31520231154277), Xiaoning LI[2](36920231153210), Yanxing Huang[1](23320231154446), Jingyan Tu[1](23320231154452)

[1]Class Informatics [2]Class AI

Abstract

The majority of currently available face forgery detectors primarily focus on specific forging patterns, such as noise characteristics, local texture, or frequency clues. This limitation hinders their ability to detect forgeries with unknown patterns and adapt the learned representations to novel forgery techniques which are not presented in the training data. In this paper, we develop a forgery detection framework based on reconstruction classification learning to overcome these challenges. Reconstruction learning over real images enhances the learned representations to detect forgery patterns that are even unknown, while classification learning mines the important differences between real and fake images. In order to get a better representation, instead of relying solely on the encoder in the reconstruction learning process, we introduce bipartite graphs that incorporate both encoder and decoder features in a multi-scale manner. Through enhancing the entire learning process for categorization and reconstruction, our aim is to improve the overall performance of face forgery detection. To validate our approach, we will conduct extensive experiments on substantial benchmark datasets and compare our results to current state-of-the-art methods.

Introduction

Face forgery-generating techniques have made significant strides in recent years (Bitouk et al. 2008). The breakthrough of deep learning has made it incredibly simple to create fake facial photos or videos that look incredibly real. These methods can be used by an attacker to spread false information, malign public figures, or compromise authentication, resulting in major political, social, and security repercussions (Lyu 2020). It is critical to develop reliable detection techniques in order to reduce malicious use of face forgery.

The conventional approach to training convolutional neural networks (CNN) for image classification, as commonly employed in early face forgery detection methodologies (Nguyen et al. 2019), typically involves taking a facial image as input and classifying it as either real or fake using pre-built CNN backbones. However, these straightforward CNN models often exhibit limitations in their ability to detect forgeries across a broad spectrum of facial variations. This limitation is evident in their tendency to focus on a narrow subset of faces, underscoring a lack of comprehensive understanding in forgery detection (Wang and Deng 2021).

Recent advancements in forgery detection have sought to address these limitations by leveraging specialized forgery patterns. These patterns include considerations of noise characteristics (Gu et al. 2022), local textures (Chen et al. 2021), and frequency information (Li et al. 2020a). By incorporating these nuanced features, these studies aim to enhance the identification of forgery artifacts present in manipulated facial images. However, a notable drawback is that these approaches heavily rely on learned patterns associated with specific manipulation techniques during training.

While these specialized approaches have demonstrated positive outcomes, their effectiveness is contingent upon encountering forgery patterns within the known repertoire of manipulation techniques. The real-world scenario poses a challenge, as forgeries employing previously unseen patterns or emerging manipulation techniques can easily render existing detection methods ineffective. This vulnerability arises due to the inability of these models to adapt to new patterns and perturbations introduced by evolving manipulation techniques, underscoring the need for more robust and adaptive face forgery detection strategies in practical applications.

Our main objective is to improve the learned representations for face forgery detection by addressing the aforementioned issues. We have two key considerations to achieve this goal. Firstly, instead of overfitting to specific forgery patterns in the training set, we aim to study the universal traits of real faces to develop representations that can generalize to new forgery patterns. We hypothesize that compact representations derived from the compact distribution of real samples are better equipped to distinguish unknown fake patterns from real faces. Secondly, we strive to enhance the network's ability to reason about forgery signals, ensuring that the learned representations capture the crucial differences between real and fake images. Categorization learning, which offers a comprehensive understanding of forgeries, plays a vital role in this process. Moreover, we propose the incorporation of a reconstruction network comprising both an encoder and a decoder. This novel architecture is designed to effectively capture the intricate distribution of real faces.

In addition to the conventional reconstruction loss employed in typical autoencoder frameworks, we introduce a metric-learning loss applied to the decoder. This additional loss function is strategically incorporated to induce a deliberate discrepancy within the embedding space. By doing so, we aim to create a distinct separation between the embeddings of real and fake images. The metric-learning loss serves as a guidance mechanism, steering the decoder towards a configuration that not only accurately reconstructs real faces but also exhibits a notable divergence when presented with fake images.

This dual-loss strategy, combining reconstruction loss and metric-learning loss, is intended to enhance the discriminative capabilities of the model. By training the network to not only faithfully reconstruct real faces but also to create a discernible gap in the embedding space, we anticipate a higher efficacy in identifying fake images, particularly those bearing unknown forgery patterns. This approach acknowledges the challenge posed by emerging manipulation techniques and variations in forgery patterns, aiming to equip the model with a more robust and adaptive mechanism for distinguishing between genuine and manipulated facial images.

To further enhance the learned representations, we will incorporate both the encoder output and decoder features, and utilize bipartite graphs that reason about forgery cues detected by the decoder features. Recent advancements in graph modeling allow for flexible and adaptive modeling of feature relationships, which is crucial for effective forgery detection (Zhao et al. 2021a). We will employ a multi-scale mechanism during the reasoning process to thoroughly explore forgery clues, considering that different face forgery techniques leave traces at varying scales. Guided by the reconstruction difference, we focus on the graph output, which serves as the final representation for classification learning, as it indicates potentially forged regions. The optimization process is performed end-to-end to jointly improve classification learning and reconstruction.

In summary, our research aims to achieve the following:

- Develop representations that can generalize to new forgery patterns by studying the common features of real face images.
- Enhance the network's ability to reason about forgery signals and capture the crucial differences between real and fake images.
- Combining reconstruction and classification learning, with the incorporation of bipartite graphs and a multi-scale mechanism to improve forgery detection performance.
- Conduct extensive experiments to validate the effectiveness of our proposed method.

Related work

Face Forgery Detection. Nowadays, many studies are proposed to boost the performance of face forgery detection(Li et al. 2020b; Gu et al. 2021; Wang and Deng 2021; Sun et al. 2022). Most early works (Lin et al. 2020; Nguyen et al. 2019; Rossler et al. 2019) model face forgery detection as a vanilla binary (real or forgery) classification problem and use neural networks to extract global features of cropped face images. However, these methods focus more on category-level differences than the subtle differences between real and fake images. Recently, there are many works proposing to focus on the manipulated features such as frequency clues, noise patterns and local textures. (Qian et al. 2020) and (Li et al. 2021) mine the frequency differences and design frequency-aware models to detect forged faces. (Zhou et al. 2017) present a two-stream deep network to detect fake faces by focusing on visual appearance and local noise in two branches, respectively. (Zhao et al. 2021a) propose a multi-attentional face forgery detector network that aggregates the low-level textural feature and high-level semantic features to discriminate real and fake samples. Although these works achieve considerable performances, they tend to overfit the training data and may experience significant performance drops on unseen forged samples.

Reconstruction Learning. Reconstruction learning has been wildly employed in unsupervised learning settings (Han et al. 2019; Liu et al. 2022; Wertheimer, Tang, and Hariharan 2021). The goal of it is to learn a representation of the input data that can accurately reconstruct the original data when fed in the model. Reconstruction learning for face forgery detection has been explored in some earlier works. (Nguyen, Yamagishi, and Echizen 2019) adopt a reconstruction network and multi-task learning for forged face detection. (Du et al. 2020) use a locality-aware autoencoder to improve the generalization capacity of the model and employ a pixel-wise mask to learn intrinsic representation from the forgery region. However, these methods perform reconstruction learning over both real and fake samples. The generalization of the learned representations is not guaranteed. In this paper, we propose to only perform reconstruction learning over real facial images to learn their common characteristics, which makes it easier for the learned representations to spot unknown forgery patterns because of the distributional discrepancy between real and fake images.

Proposed Solution

To effectively capture the fundamental differences between real and fake faces, we propose a novel framework called RCL. This framework encompasses three key components: reconstruction learning, multi-scale graph aggregation, and reconstruction guided attention, as depicted in Figure 1. The reconstruction network focuses on modeling the distributions of real face images, enabling the detection of unknown forgery patterns. Additionally, the multi-scale graph aggregation scheme aggregates the captured discrepancy information from both the encoder and decoders of the reconstruction network, allowing for comprehensive analysis of the differences between real and fake faces across multiple scales. Furthermore, the reconstruction guided attention module directs the classification network to prioritize attention towards forgery traces. The subsequent sections provide a detailed presentation of these three components.

Reconstruction Learning. To explore the shared characteristics of authentic faces more appropriately, we propose employing reconstruction learning to solely restore genuine facial images. Specifically, given an input image

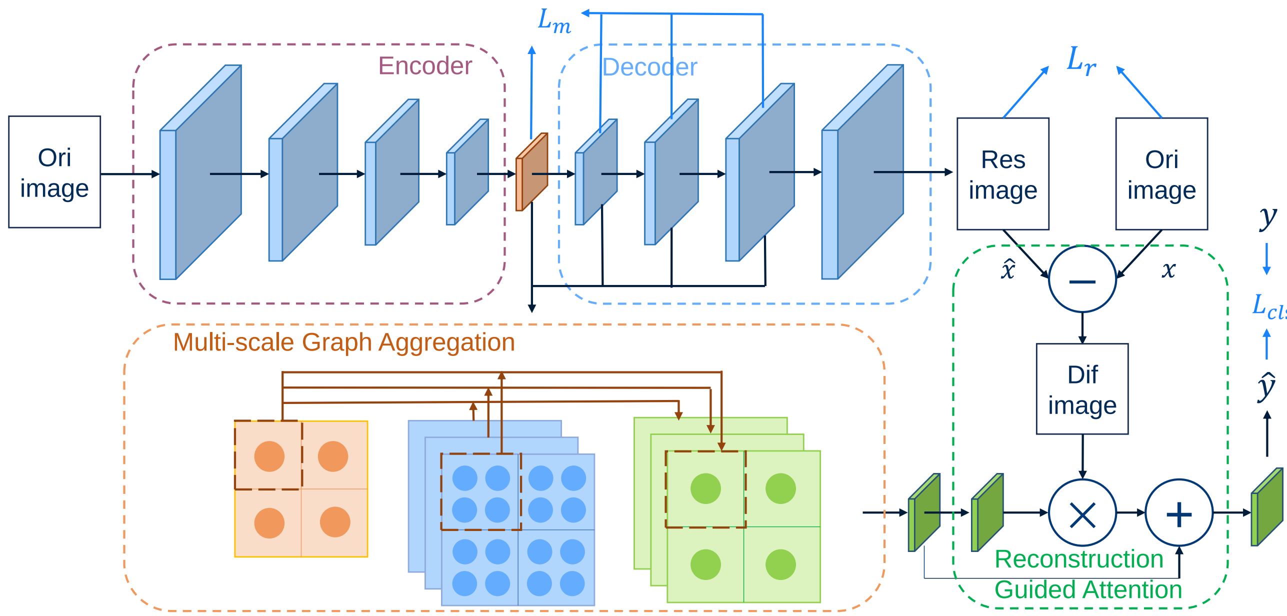

Figure 1: Schematic diagram of the proposed framework named RCL, which consists of three main schemes, i.e., reconstruction learning, multi-scale graph aggregation, and reconstruction guided attention. Our proposed method comprises an encoder-decoder network for learning image representations from both real and fake faces. The encoder output undergoes multi-scale graph aggregation to refine these representations, guided by reconstruction difference for final classification. We jointly train our system by minimizing the classification loss $\mathcal{L}_{cls}$, the reconstruction loss $\mathcal{L}_r$ (computed exclusively on real faces), and the metric-learning loss $\mathcal{L}_m$.

$X \in \mathbb{R}^{(h\times w\times 3)}$,we train a reconstruction network $\mathbb{F}$ using an encoder-decoder architecture. We incorporate some white noise into the input samples during the training process to obtain $\widetilde{x}$, aiming to acquire robust representations of real faces. Consequently, the image reconstruction procedure can be described as follows:

$$\widehat{x} = \mathcal{F}(\widetilde{x}) \tag{1}$$

During the reconstruction process, we calculate the reconstruction loss L_r between the input real images and their reconstructed versions in a mini-batch as follows:

$$L_r = \frac{1}{|R|}\sum_{i\in R}\|\widehat{X}_i - X_i\|_1 \tag{2}$$

where R denotes the set of real samples in a mini-batch and |R| is the cardinality of R.

We employ a metric-learning loss to encourage proximity among real images and distance between real and fake images in the embedding space. For simplicity, let $F \in \mathbb{R}^{(h'\times w'\times 3)}$ represent the output features of an encoder or decoder block. By performing global average pooling on F, we obtain the feature vector $\overline{F} \in \overline{\mathbb{R}}^c$ for each input sample. The metric-learning loss is calculated as follows:

$$L_m = \frac{1}{N_{RR}}\sum_{i\in R, j\in R} d(\overline{F}_i, \overline{F}_j) - \frac{1}{N_{RF}}\sum_{i\in R, j\in R} d(\overline{F}_i, \overline{F}_j) \tag{3}$$

where R, F denote the set of real and fake samples. $N_R R$ and $N_R F$ are the total number of (real, real) pairs and (real,fake) pairs, respectively. $d(\cdot,\cdot)$ is a pair-wise distant function based on the cosine distance:

$$d(a,b) = \frac{1 - \frac{a}{\|a\|_2}\cdot\frac{b}{\|b\|_2}}{2} \tag{4}$$

The first part of $\mathcal{L}_m$ encourages the learning of compact representations from genuine faces, while the second part highlights the distinctions between real and fake samples. Unlike conventional metric-learning losses applied directly to feature extractors, our proposed loss is tailored to amplify reconstruction differences, thereby facilitating reconstruction learning. Furthermore, we do not impose compactness constraints on fake data, considering their significant feature variations across different forgery techniques. We apply the metric-learning loss to the output of the last encoder block and each decoder block.

Multi-scale Graph Aggregation. To leverage the valuable information within the decoder for distinguishing between real and fake images, we introduce the metric-learning loss. However, to enhance the classification of real and fake images based on the information embedded in the decoder, we propose a multi-scale graph aggregation (MGA) module. This module combines the features from the decoder blocks and the encoder output into a bipartite graph, allowing comprehensive reasoning about forgery cues.

In the MGA module, we focus on the feature maps of a

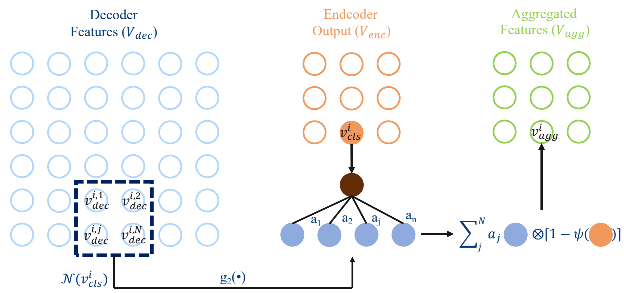

Figure 2: Illustration of the proposed multi-scale graph aggregation scheme to aggregate information in the encoder output (orange) and decoder features for a given scale (blue) to produce richer representations (green).

decoder block at a specific scale. As illustrated in Figure 2, we model both the encoder output and the decoder features. We represent F_enc and F_dec, as two sets of vertices $V_{enc} = \{v^i_{enc}\}_{i=1}^{h1\times w1}$, $V_{dec} = v^i_{deci=1}^{\;\;h2\times w2}$, where each vertex corresponds to an embedding vector of the original feature maps. For each v^i_{enc}, we define $\mathcal{N}\left(v^i_{enc}\right) = \{v^{i,j}_{dec}\}_{j=1}^N$ as the set of vertices in V_{dec} that are linked to it. N represents the number of vertices in the set. To improve the reasoning about forgery clues, the graph aggregation process aggregates the information from $\mathcal{N}\left(v^i_{enc}\right)$ to enrich the feature representations of v^i_{enc}. We keep the spatial correspondence when aggregating the information from the decoder to the encoder to model the local relationship since forgery traces usually reside in continuous local areas.

To determine their importance, we project v^i_{enc} and $v^{i,j}_{dec}$ to a shared embedding space using two neural networks, $g1(\cdot)$ and $g2(\cdot)$, resulting in $\widetilde{v}^i_{enc}, \widetilde{v}^{i,j}_{dec}$, respectively. We compute a weight coefficient a_j to indicate the relevance of v^i_{enc} and $v^{i,j}_{dec}$. Particularly, we first concatenate the vertices from the two sub-graphs, and then passing through a single-layer network ϕ to get a_j as:

$$a_j = \frac{exp(\phi(\widetilde{v}^i_{enc}\|\widetilde{v}^{i,j}_{dec}))}{\sum_{v^{i,l}_{dec}\in\mathcal{N}(v^i_{enc})} exp(\phi(\widetilde{v}^i_{enc}\|\widetilde{v}^{i,j}_{dec}))} \tag{5}$$

where $\|$ denotes the concatenation operation. We then compute $a \in [0, 1]$ valued vector based on v^i_{enc}. During information aggregation, we particularly enhance the channels of $v^{i,j}_{dec}$. when the weight of the corresponding channels of $\widetilde{v}^i_{enc}$ is small. The aggregated feature vector v^i_{agg} is computed by:

$$v^i_{agg} = \sum_{j=1}^N a_j \widetilde{v}^{i,j}_{dec} \otimes \left[1 - \left(v^i_{enc}\right)\right] \tag{6}$$

We propose a multi-scale approach to extract comprehensive forgery information by mining traces resulting from different manipulation techniques. Specifically, we aggregate the output features of the encoder with each block output of the decoder in a multi-scale manner. Using a sigmoid function and two fully-connected layers, we concatenate the aggregated features v^i_{agg} in different scales with v^i_{enc} to produce an enhanced feature vector v^i_{enh} with the same channel dimension as v^i_{enc}. Finally, we spatially assemble the enhanced feature vectors v^i_{enh} to obtain the enhanced feature maps F_{enh} for reconstruction-guided attention.

Reconstruction Guided Attention. With the constraints of the reconstruction network, the visually distinct reconstructed forged faces compared to the input forged faces prompt us to utilize the reconstruction difference to highlight potential manipulated traces. Therefore, to indicate the probably manipulated traces, we propose the reconstruction guided attention module to guides the classification network to prioritize forgery traces. Given the reconstructed image $\widehat{X}$ nd the original image X, we first compute their difference in pixel level to get the difference mask m as:

$$m = \left|\hat{X} - X\right| \tag{7}$$

Given F_{enh} the enhanced feature maps, we compute the attention map based on the difference mask and apply it to F_{enh} spatially to get $F\prime_{enh}$. Then, we add $F\prime_{enh}$ and F_{enh} to obtain the attended output features:

$$F'_{enh} = \sigma(f_1(m)) \otimes f_2(F_{enh}) \tag{8}$$

$$F_{att} = F'_{enh} + F_{enh} \tag{9}$$

where f_1, f_2 represent the convolutional operations, σ is the sigmoid function, and $\otimes$ denotes the element-wise multiplication. To maintain simplicity, we exclude the spatial size of these tensors and employ bilinear interpolation to ensure proper spatial sizing for the mentioned operations.

Loss Function. The total loss function $\mathcal{L}$ of the proposed framework includes the reconstruction loss and the metric-learning loss for reconstruction learning, together with the cross-entropy loss $\mathcal{L}_{cls}$ for binary classification:

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda_1\mathcal{L}_r + \lambda_2\mathcal{L}_m \tag{10}$$

Experiments

Experimental Setup

Datasets. To assess the effectiveness of our proposed method and compare it with existing approaches, we conducted comprehensive evaluations on the widely utilized FaceForensics++ (FF++) dataset (Rossler et al. 2019). This dataset has gained prominence for its diversity and realism, comprising four distinct manipulation techniques: Deepfakes (DF), Face2Face (F2F), FaceSwap (FS), and NeuralTextures (NT). Each technique introduces unique challenges, reflecting the evolving landscape of facial manipulation methods.

The FF++ dataset is meticulously curated, drawing from various sources to ensure a representative and varied collection of manipulated facial data. This diversity is crucial for evaluating the robustness of forgery detection methods, as it exposes models to a broad spectrum of facial manipulations encountered in real-world scenarios.

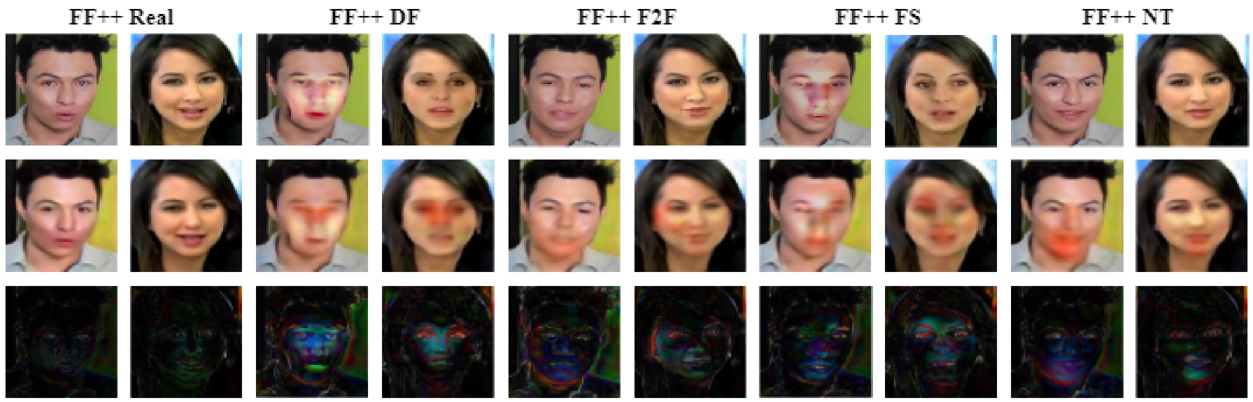

Figure 3: Reconstruction visualization of the proposed method on the FaceForensics++ dataset The first row displays the input images. The second row and the third row show reconstruction results and pixel-level differences, respectively.

Evaluation Metrics. To evaluate our method, we report the most commonly used metrics in related studies(Afchar et al. 2018; Chen et al. 2021; Rossler et al. 2019), including Accuracy (Acc) and Area Under the Receiver Operating Characteristic Curve (AUC).

Implementation Details. The proposed framework is implemented based on the Xception(Chollet 2017) architecture. We train it with a batch size of 32, using the Adam optimizer with an initial learning rate of 2e-4 and a weight decay of 1e-5. A step learning rate scheduler is employed to adjust the learning rate. $\lambda 1$ and $\lambda 2$ in Equation 10 are empirically set to 0.1. For data augmentation, we only use random horizontal flipping.

Experimental Results

Intra-testing. We compare our proposed method with other methods. As shown in Table 1, for FF++ dataset, our proposed method outperforms other approaches.

Table 1: Intra-testing comparisons.

Methods	Acc(%)	AUC(%)
Xception	86.86	89.30
Add-Net	87.50	91.01
MultiAtt	88.69	90.40
RCL(ours)	**91.03**	**95.02**

Cross-testing. To assess how well our method generalizes to unfamiliar manipulations, we perform cross-dataset experiments. This involves training our models on FF++ dataset and subsequently testing them on WildeDeepfaks and Celeb-DF(Li et al. 2020b) datasets. By conducting these experiments, we can evaluate the performance of our method on detecting forgeries that are outside the training dataset. The results are shown in Table 2. From the table, we can know that our method generally outperforms all the listed methods in terms of performance on unseen test data, often with a significant margin.

Table 2: Cross-testing in terms of AUC (%) by training on FF++ dataset

Methods	WildeDeepfake(%)	Celeb-DF(%)
	AUC	AUC
Xception	67.72	61.80
Add-Net	62.35	64.78
MultiAtt	59.74	68.01
RCL(ours)	**64.31**	**69.06**

Reconstruction visualization. In order to provide a clear understanding of reconstruction learning, we have visualized the results of the reconstruction network and compared them with the original input in Figure 3. The visualization allows us to observe that genuine faces can be reconstructed effectively with minimal blurring, while the manipulated regions in fake faces cannot be accurately restored. The difference masks further highlight the disparities between real and manipulated faces, thereby revealing potential indications of forged regions. It is important to note that our method is solely trained using image-level supervision.

Conclusion

In this paper, we present a novel approach to detect facial forgery, emphasizing the exploration of shared compact representations of genuine facial features to discern the disparities between authentic and manipulated faces. Our inventive multi-scale graph aggregation module seamlessly integrates encoder output and decoder features into bipartite graphs across multiple scales, facilitating a comprehensive analysis of forgery indicators. Concurrently, we introduce the reconstruction-guided attention module to steer the model's focus towards potential forgery traces. Rigorous experiments and intricate visualizations confirm the resilience and applicability of our method across well-established benchmark datasets.

References

Afchar, D.; Nozick, V.; Yamagishi, J.; and Echizen, I. 2018. Mesonet: a compact facial video forgery detection network. In *2018 IEEE international workshop on information forensics and security (WIFS)*, 1–7. IEEE.

Bitouk, D.; Kumar, N.; Dhillon, S.; Belhumeur, P.; and Nayar, S. K. 2008. Face swapping: automatically replacing faces in photographs. In *ACM SIGGRAPH 2008 papers*, 1–8.

Chen, S.; Yao, T.; Chen, Y.; Ding, S.; Li, J.; and Ji, R. 2021. Local relation learning for face forgery detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 1081–1088.

Chollet, F. 2017. Xception: Deep Learning with Depthwise Separable Convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1800–1807.

Dolhansky, B.; Howes, R.; Pflaum, B.; Baram, N.; and Ferrer, C. C. 2019. The Deepfake Detection Challenge (DFDC) Preview Dataset.

Du, M.; Pentyala, S.; Li, Y.; and Hu, X. 2020. Towards generalizable deepfake detection with locality-aware autoencoder. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 325–334.

Gu, Q.; Chen, S.; Yao, T.; Chen, Y.; Ding, S.; and Yi, R. 2022. Exploiting fine-grained face forgery clues via progressive enhancement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 735–743.

Gu, Z.; Chen, Y.; Yao, T.; Ding, S.; Li, J.; Huang, F.; and Ma, L. 2021. Spatiotemporal inconsistency learning for deepfake video detection. In *Proceedings of the 29th ACM international conference on multimedia*, 3473–3481.

Han, Z.; Wang, X.; Liu, Y.-S.; and Zwicker, M. 2019. Multi-angle point cloud-VAE: Unsupervised feature learning for 3D point clouds from multiple angles by joint self-reconstruction and half-to-half prediction. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 10441–10450. IEEE.

Li, J.; Xie, H.; Li, J.; Wang, Z.; and Zhang, Y. 2021. Frequency-aware discriminative feature learning supervised by single-center loss for face forgery detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 6458–6467.

Li, L.; Bao, J.; Zhang, T.; Yang, H.; Chen, D.; Wen, F.; and Guo, B. 2020a. Face x-ray for more general face forgery detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5001–5010.

Li, Y.; Yang, X.; Sun, P.; Qi, H.; and Lyu, S. 2020b. Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3204–3213.

Lin, M.; Ji, R.; Wang, Y.; Zhang, Y.; Zhang, B.; Tian, Y.; and Shao, L. 2020. Hrank: Filter pruning using high-rank feature map. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1529–1538.

Liu, X.; Liu, X.; Liu, Y.-S.; and Han, Z. 2022. Spu-net: Self-supervised point cloud upsampling by coarse-to-fine reconstruction with self-projection optimization. *IEEE Transactions on Image Processing*, 31: 4213–4226.

Lyu, S. 2020. Deepfake detection: Current challenges and next steps. In *2020 IEEE international conference on multimedia & expo workshops (ICMEW)*, 1–6. IEEE.

Nguyen, H. H.; Fang, F.; Yamagishi, J.; and Echizen, I. 2019. Multi-task learning for detecting and segmenting manipulated facial images and videos. In *2019 IEEE 10th international conference on biometrics theory, applications and systems (BTAS)*, 1–8. IEEE.

Nguyen, H. H.; Yamagishi, J.; and Echizen, I. 2019. Capsule-forensics: Using capsule networks to detect forged images and videos. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2307–2311. IEEE.

Qian, Y.; Yin, G.; Sheng, L.; Chen, Z.; and Shao, J. 2020. Thinking in frequency: Face forgery detection by mining frequency-aware clues. In *European conference on computer vision*, 86–103. Springer.

Rossler, A.; Cozzolino, D.; Verdoliva, L.; Riess, C.; Thies, J.; and Nießner, M. 2019. Faceforensics++: Learning to detect manipulated facial images. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1–11.

Ruff, L.; Vandermeulen, R. A.; Görnitz, N.; Binder, A.; Müller, E.; Müller, K.-R.; and Kloft, M. 2019. Deep semi-supervised anomaly detection. *arXiv preprint arXiv:1906.02694*.

Sun, K.; Yao, T.; Chen, S.; Ding, S.; Li, J.; and Ji, R. 2022. Dual contrastive learning for general face forgery detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 2316–2324.

Wang, C.; and Deng, W. 2021. Representative forgery mining for fake face detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14923–14932.

Wertheimer, D.; Tang, L.; and Hariharan, B. 2021. Few-shot classification with feature map reconstruction networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8012–8021.

Zhao, H.; Zhou, W.; Chen, D.; Wei, T.; Zhang, W.; and Yu, N. 2021a. Multi-attentional deepfake detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2185–2194.

Zhao, Y.; Yan, K.; Huang, F.; and Li, J. 2021b. Graph-based high-order relation discovery for fine-grained recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15079–15088.

Zhou, P.; Han, X.; Morariu, V. I.; and Davis, L. S. 2017. Two-stream neural networks for tampered face detection. In *2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW)*, 1831–1839. IEEE.