Recent Advances in Copy–Move Forgery Detection: A Deep Learning Oriented Survey

Duzhihui Li [1, *], Yuhan Sun [2]

[1] School of Computer Science and Technology, Shanghai University of Electric Power, Shanghai, China

[2] International College, Chongqing University of Posts and Telecommunications, Chongqing, China

* Corresponding Author Email: lduzhihui@gmail.com

Abstract. With the widespread use of digital images in news reporting, forensic analysis, and social media, the problem of image tampering has become increasingly prominent. To address this, researchers have proposed various Copy–Move Forgery Detection (CMFD) approaches, primarily based on block segmentation, key points, and deep learning. However, with the rapid development of emerging technologies such as Transformers, Generative Adversarial Networks (GANs), and Diffusion Models, while exploratory research has emerged in image tampering detection, a systematic review of these approaches is lacking. Therefore, this article systematically reviews the development of CMFD within a three-part framework, focusing on the latest progress from 2023 to 2024. Specifically, this article reviews the classic contributions and evolution of block-based methods and analyzes the application value of key point-based methods in low-computing scenarios such as mobile devices and remote sensing. It also explores the advantages of deep learning methods in detection accuracy, robustness, and adaptability to AIGC scenarios. This paper further compares the characteristics and application scenarios of three types of methods (Transformer, GAN and Diffusion Models), and further proposes future development directions.

Keywords: Image Forensics, Copy–Move Forgery Detection, Transformer, AIGC, Deepfake.

1. Introduction

In the digital age, images are an important carrier of news, justice, academia and social interaction, but their authenticity is facing severe challenges. Copy-Move Forgery (CMF) is easy to operate and highly concealed. It often copies the internal area of the image to cover up or repeat the content without introducing external elements. It is difficult to identify with the naked eye and poses a serious threat to the credibility of information and judicial fairness [1, 2]. To solve this problem, researchers have proposed a variety of Copy-Move Forgery Detection (CMFD) methods, covering block-based, key point-based and deep learning-based technical routes [3-5]. Traditional methods dominated the mainstream between 2000 and 2015, relying on manually designed features and having strong interpretability, but with obvious shortcomings in computational efficiency and geometric robustness [6]. With the development of deep learning (DL), the emergence of end-to-end models such as Buster Net has promoted important changes in this field [5]. Existing review works have systematically summarized these methods [7], but their contents mostly focus on early research and pay insufficient attention to the Transformer, GAN and Diffusion methods that have developed rapidly in recent years [8]. At the same time, there is also a lack of in-depth discussion of the latest datasets, evaluation systems and new challenges in the context of AIGC/Deepfake [9-11]. Based on this, this paper continues the "three-part classification framework" and strives to make three supplements: first, review and summarize representative research under the three categories of methods, especially the latest results in 2023-2024; second, sort out the technology evolution path, from traditional manual features to CNN, and then to the evolution logic of Transformer/GAN and Diffusion; third, combine AIGC/Deepfake scenarios to explore cutting-edge issues such as cross-modal forensics, open world detection and explainability. Through this systematic review, this paper hopes to provide reference and inspiration for continued research in this field. Existing research shows that most CMFD methods follow a common process: preprocessing, feature extraction, feature matching and result visualization

[1,3,5]. This framework not only unifies the research basis, but also provides direction for subsequent improvements in robustness and efficiency.

2. Traditional Methods for Copy–Move Forgery Detection

2.1. Block-based Methods

Block-based detection is the earliest method proposed and most systematically studied in the field of CMF [1]. Its basic process is shown in Figure1.

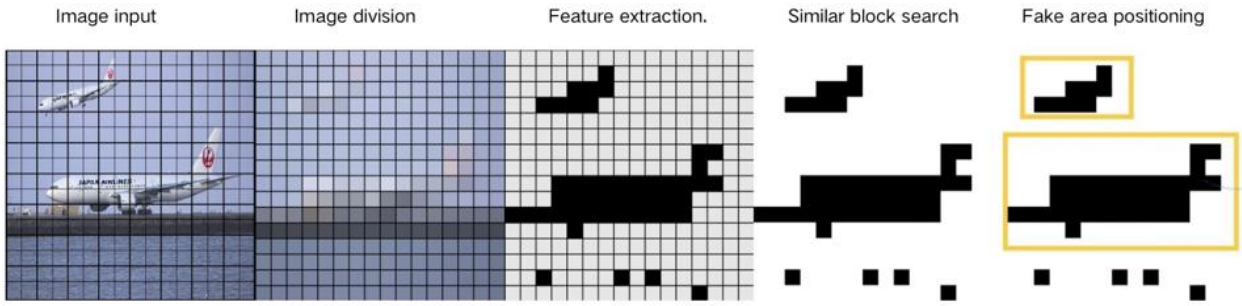

Figure 1. Basic framework of block-based methods (Picture credit: Original)

Early studies mostly relied on manual features, including frequency domain features (such as DCT, DWT), dimensionality reduction features (such as PCA, SVD) [2], texture features (such as LBP, Gabor), and moment features (such as Hu, Zernike) [5]. These methods are effective in small-scale forgery detection, but have problems such as computational complexity and lack of robustness to rotation and scaling.

Between 2016 and 2020, researchers tried to alleviate computational overhead by accelerating matching through sparse representation, dictionary learning, and PatchMatch, while also introducing affine invariant partitioning and rotation invariant description to enhance geometric robustness [3]. These improvements outperformed traditional solutions on datasets such as MICC and CASIA, but because they still rely on manual features, they are difficult to adapt to complex deformations and diverse tampering, and are generally regarded as the results of the transitional stage [7].

In recent years, block-based methods have gradually been combined with deep learning. The typical idea is to use CNN or Siamese networks to replace traditional feature extractors, learn more robust representations from block pairs, and directly predict masks through the Encoder-Decoder structure [6]. Subsequently, the Transformer architecture was introduced to model long-range dependencies and improve the detection of large-scale copies [9]. Further research also combined GAN and Diffusion to enhance the diversity of training samples or learn robust feature distributions, showing potential in AIGC and Deepfake scenarios [8, 10, 12]. Block-based methods have gradually evolved from traditional manual features to a hybrid model of "local constraints + deep features". Although the research interest has declined, its interpretability advantage makes it still valuable in applications such as forensic evidence collection, and it is often used as a baseline for comparison and verification of new methods.

2.2. Keypoint-based Methods

Scale-Invariant Feature Transform (SIFT) is the earliest key point detection method widely used in CMFD. Its core idea is to extract stable key points in scale space and construct descriptors through directional histograms, so as to maintain high robustness under rotation, scaling and a certain degree of brightness changes [4]. SIFT was regarded as a benchmark method in early studies, but its computational complexity is high and its feature dimension is large, which limits its application in large-scale images or real-time scenarios. In order to improve efficiency, Speeded Up Robust Features (SURF) was proposed based on SIFT. It accelerates the feature extraction process by approximating the integral image and the Hessian matrix, significantly improving the detection speed while

maintaining robustness. It is more suitable for processing large-scale images and application scenarios that require fast response. The overall process of the key point-based method as shown in Figure 2.

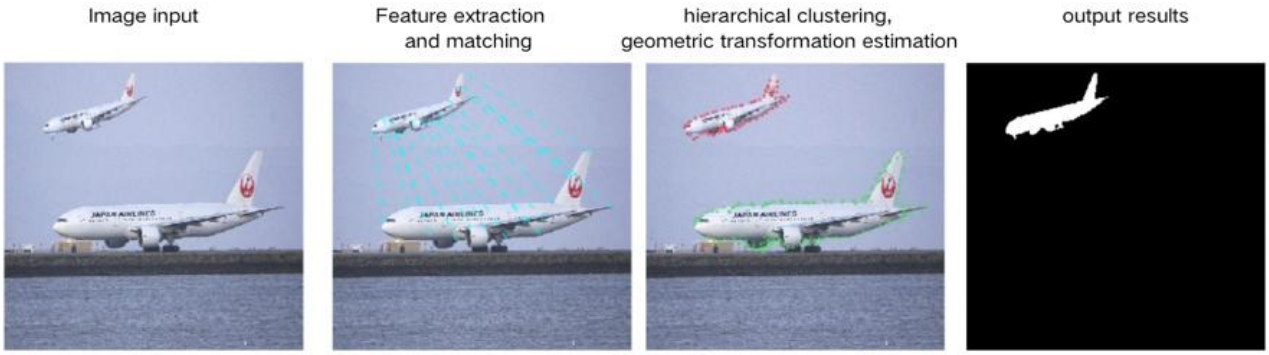

Figure 2. Basic framework of keypoint-based methods (Picture credit: Original)

After SIFT and SURF, researchers have proposed a variety of lightweight keypoint detection algorithms, such as ORB (Oriented FAST and Rotated BRIEF), BRISK (Binary Robust Invariant Scalable Keypoints), and KAZE features. These methods usually use binary descriptors, which have the advantages of low storage overhead and fast matching speed, and are particularly suitable for real-time detection and mobile device scenarios [6]. For example, ORB significantly reduces the computational cost while ensuring a certain degree of robustness; BRISK shows good consistency in multi-scale environments; KAZE extracts feature points through nonlinear scale space, better preserving image details. In addition, some studies have attempted to combine keypoint features with texture, color, or gradient information to form hybrid methods. For example, fusing SIFT with LBP features can enhance the detection effect of low-texture areas, while the scheme combined with color histogram is more robust under illumination changes. Such methods are often better than single features in terms of accuracy, but also increase the computational complexity. In practical applications, they have shown high practical value and are particularly suitable for scenarios that require a balance between robustness and efficiency.

3. Deep Learning Methods for Copy-Move Forgery Detection Summary

Over the past decade, the rise of deep learning has transformed the field of CMFD. Compared to traditional methods of manually extracting features, deep learning automatically learns to extract features from data. Especially after 2016, CNN, Siamese networks, Transformer, and more recently GAN and Diffusion models have been used in this field [8-10].

3.1. Methods Based on Pre-trained Networks

VGGNet, ResNet, and other networks pre-trained on ImageNet were first introduced into CMFD to extract image features. This makes it easy to use visual features for detection even when a large amount of data has been tampered with. This method generally uses CNN to obtain high-level semantic features of the image, and then uses similarity judgment (such as Euclidean distance and SVM classifier) to determine whether there is a duplicate area [6]. From a large number of papers, it can be seen that this method still has good robustness in the face of noise and compression environments, but it is sensitive to geometric transformations (rotation, scaling), and its generalization ability also depends on the pre-trained model itself. In general, this type of method uses existing visual knowledge to complete the task. The effect is significant in the early stage, but its ability to deal with complex tampering is limited. The basic process of the method based on pre-trained networks is shown in Figure 3.

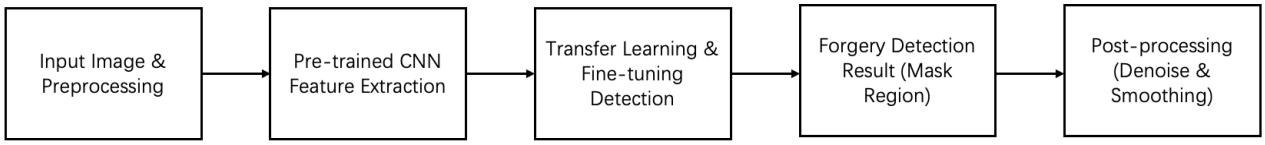

Figure 3. Basic framework of pretrained network-based methods (Picture credit: Original)

3.2. Methods Based on Fusion Networks

In view of the limitations of pre-trained networks, scholars realized that the core of CMFD is not "difference" but "similarity", that is, the correspondence between two regions within the image. Therefore, fusion networks became popular, such as Siamese networks and two-stream CNNs. They input two image blocks into a network with shared parameters and measure the similarity through the learned feature vectors. This can more accurately characterize the "relationship between the two blocks" and is particularly sensitive to small-scale copying and forgery. However, this requires a lot of calculations. Large images require thousands of candidate block pairs, which is very time-consuming and computationally expensive [5]. At the same time, the emergence of fusion networks is an improvement in methods, allowing detection to move from "single-point feature comparison" to "relationship modeling." The basic process of the method based on fusion networks is shown in Figure 4.

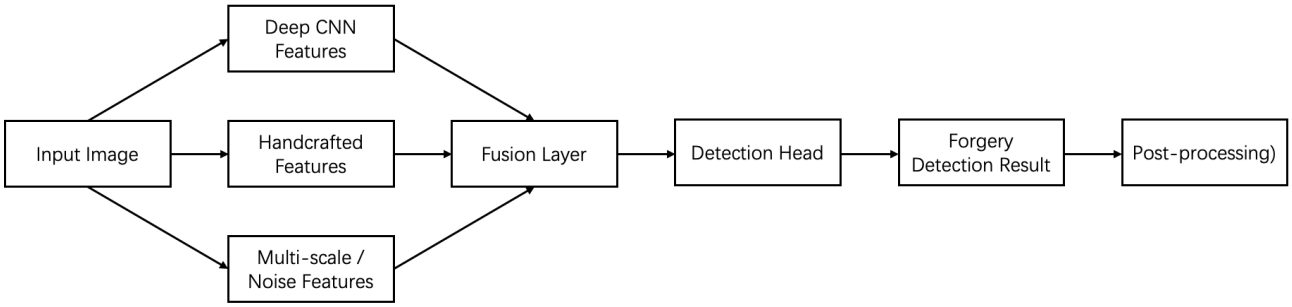

Figure 4. Basic framework of fusion network-based methods (Picture credit: Original)

3.3. Methods Based on Transformer

Since 2021, the successful application of the Transformer model in computer vision has also promoted its development in image tampering detection. Unlike CNNs that rely on local convolution, Transformers can model long-range dependencies through the self-attention mechanism, as shown in Figure 5, making them more suitable for identifying large-scale CMFD. Studies have shown that architectures such as Vision Transformer (ViT) and Swin Transformer have surpassed ResNet in performance on large-scale datasets such as CoMoFoD++, and have shown stronger robustness in geometric transformation scenarios [9,11].

At the same time, researchers have also proposed a hybrid architecture: combining the local modeling advantages of CNN with the global dependency capabilities of Transformer to balance detail capture and overall consistency. However, Transformer also faces new challenges in CMFD: its training is highly dependent on data scale, and it is difficult to realize its full potential in the absence of large-scale labeled data; in addition, its computational and graphics memory overhead is significantly higher than that of traditional CNN, limiting its application in real-time or low-computing environments.

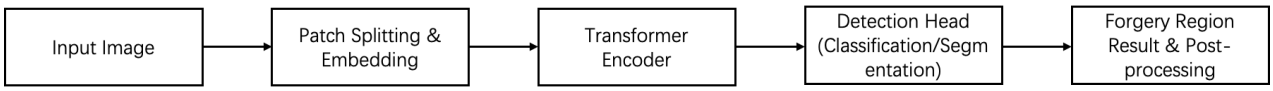

Figure 5. Basic framework of transformer-based methods (Picture credit: Original)

3.4. Methods Based on GAN/Diffusion-related Models

In recent years, with the explosion of AICG and Deepfake technology, the types and complexity of CMF technology have increased dramatically. CMFD has to face the huge challenge of generative forgery. Based on this, some studies have begun to use GAN and Diffusion models for CMFD as shown in Figure 6. GAN adversarial generation technology is mainly used to synthesize a large number of samples for training to improve the robustness of the model; Diffusion models use their powerful generation and reconstruction capabilities to identify those areas that are obviously deviated from the natural distribution. Experiments show that these methods have better generalization capabilities in cross-dataset testing [10] [12]. However, GAN and Diffusion methods also have limitations. On the one hand, their lack of interpretability makes their application in high-credibility scenarios such as judicial evidence collection still questionable; on the other hand, the computational cost of training and inference is significantly higher than that of CNNs. For example, Diffusion-based methods perform well in new forgery detection, but often require several times the training time and video memory consumption of CNNs. Overall, the introduction of these generative models has expanded the research boundaries of CMFD, but further exploration is still needed in terms of efficiency and application implementation.

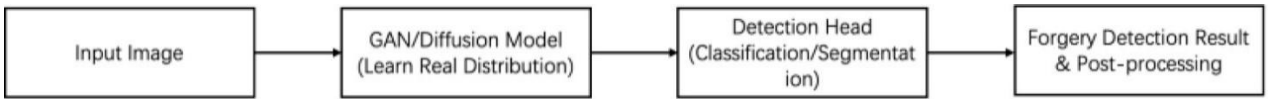

Figure 6. Basic framework of GAN/Diffusion-related methods (Picture credit: Original)

4. Comparsion and Discussion

4.1. Comparative Analysis and Trend Exploration

As shown in Figure 7, the Block-based method has evolved from traditional manual features, improved optimization to deep learning fusion. The traditional method is intuitive and has high positioning accuracy, but it has high computational complexity and is not robust to rotation and scaling [1,2,6]; the improved method has improved efficiency and robustness through sparse representation, Patch Match acceleration and affine invariant partitioning, but it still relies on manual features and has limited adaptability [3, 4, 7]; the hybrid mode of fusion deep learning maintains local constraints while learning stronger representations with the help of CNN or Transformer, showing better performance in complex scenarios [6, 9, 10, 12].

Figure 7. Development timeline of block-based methods (Picture credit: Original)

Block-based methods have been gradually marginalized in mainstream research, remaining primarily as tools for experimental baselines or forensic evidence collection that require interpretability. However, their concepts are not completely outdated: in low-computing environments, block-based frameworks still have certain application value, and their integration with deep features has also provided important inspiration for subsequent methods.

4.2. Comparison and Practical Applications

Overall, keypoint methods offer advantages over block matching methods: they are more robust to rotation, scaling, and illumination variations, require less feature storage, and have fast matching speeds, making them suitable for large images and real-time applications. However, they can be difficult to extract stable keypoints in low-texture or smooth areas, which can easily lead to missed detections.

Keypoint methods continue to demonstrate unique value in practical applications. For example, their lightweight and computationally intensive nature makes them an important complement to deep learning methods in the rapid preprocessing of large-scale image screening on social media, satellite remote sensing monitoring, and forensic evidence collection. Lightweight methods such as ORB and BRISK remain particularly promising in mobile and embedded devices. In recent years (2023–2024), some research has proposed hybrid methods that combine deep feature compression with fast keypoint matching. These methods further reduce computational overhead while maintaining accuracy, making real-time CMFD possible. Table 1 shows a comparison between block-based and keypoint-based methods.

Table 1. Comparative analysis of block-based and keypoint-based CMFD approaches

Method Category	robustness	Detection speed	Applicable Scenarios	Development trends in recent years
Block-based	Sensitive to rotation/scaling; somewhat robust to compression	Large number of features and slow speed	Small-scale forgery, early research	Gradually replaced, mainly as a baseline
Keypoint-based	More robust to rotation/scale/brightness changes	Fewer features and faster matching speed	Large images, real-time detection, and low computing power scenarios	Lightweight direction still has research value

4.3. Method Comparison and Critical Discussion

Without considering computing power, deep learning methods demonstrate advantages in copy-move forgery detection: they are more robust in more complex scenarios and data; Transformer and Diffusion models offer significantly higher detection accuracy and can be applied to a wide range of copy-move forgeries. Furthermore, with the explosion of AICG and deepfake technologies, deep learning offers broader application opportunities, addressing both traditional and new AIGC forgeries. A comparison of traditional and deep learning methods is shown in Table 2.

However, deep learning also faces significant challenges: Without extensive annotated data, it is difficult to maintain stability; its performance varies across different datasets, and its generalization capabilities need improvement; and its mathematical interpretability is significantly lower than that of traditional methods, making it difficult to intuitively trace evidence of forgery.

Table 2. Comparison between traditional and deep learning-based CMFD methods

Dimension	Traditional Methods (Block/Keypoint)	Deep Learning Methods (CNN/Transformer/GAN)
Robustness	Effective against compression and noise, but sensitive to rotation and scaling	High accuracy across diverse tampering scenarios, robust to geometric transformations
Speed	Feature matching is time-consuming and relatively slow	Training is costly, but inference is fast (with GPU support)
Data Dependency	No need for training datasets	Requires large-scale annotated datasets
Generalization	Performs reasonably well across datasets	Prone to overfitting, significant performance drop across distributions
Explainability	High, results can be traced back to specific blocks/keypoints	Low, often considered a black box; requires support from XAI research
Application Scenarios	Small images, low-computation devices	Large-scale forensics, AIGC/Deepfake image detection

4.4. Summary and Emerging Trends

It can be seen that deep learning methods are gradually becoming the mainstream direction of CMFD: Transformer and Diffusion models are the focus of the future, especially in the AIGC scenario where these two models have outstanding performance [9, 11, 13]. However, in the face of increasing computing power requirements, it is obvious that unlimited increase in computing power is impossible. Lightweight research will also become popular for usability in mobile terminals and low computing power environments [8, 12]. Finally, interpretability and judicial credibility are also issues that deep learning must face in the next step. Deep learning relies on a large amount of data and lacks mathematical interpretability, making it difficult to truly apply it to legal or forensic practice [7, 10].

5. Conclusion

This paper reviews the main research routes of CMFD, including Block-based, Keypoint-based and Deep Learning methods. Block-based methods were widely adopted in the early days due to their intuitiveness and high positioning accuracy. However, with the improvement of image resolution and the complexity of tampering methods, their limitations under geometric transformations such as rotation and scaling have gradually been exposed. Keypoint-based methods are more robust and computationally efficient in rotation and scaling scenarios, so they still have application value in scenarios with limited computing power such as mobile terminals and remote sensing images. However, they are often difficult to play a role in low-texture areas. In recent years, deep learning methods have rapidly emerged, especially with the introduction of Transformer, GAN and Diffusion, which have significantly improved detection accuracy and cross-domain adaptability. However, these methods rely on large-scale data and computing power, and their generalization ability and interpretability problems have not been fundamentally solved. Overall, deep learning has become the research core of CMFD, but the future development direction still deserves in-depth consideration. For example, cross-modal forensics (such as combining images and metadata) is becoming a trend; lightweight research on mobile terminals and edge computing will also become increasingly important; at the same time, large models such as CLIP and SAM provide new opportunities for cross-task migration in this field. In addition, issues of explainability and legal credibility still require attention, and the lack of a unified benchmark is still a recognized shortcoming in academia. Solving these problems will determine whether this field can move from academic research to larger-scale practical applications.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Fridrich J, Soukal D, Lukas J. Detection of copy–move forgery in digital images. In: Proceedings of the Digital Forensic Research Workshop (DFRWS 2003), 2003: 652 – 663.

[2] Popescu AC, Farid H. Exposing digital forgeries by detecting duplicated image regions. Dartmouth College, Computer Science Technical Report TR2004 - 515, 2004.

[3] Bayram S, Sencar HT, Memon N. An efficient and robust method for detecting copy–move forgery. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2009), 2009: 1053 – 1056. IEEE.

[4] Christlein V, Riess C, Angelopoulou E. On rotation invariance in copy–move forgery detection. In: Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS 2010), 2010: 1 – 6. IEEE. https: //doi.org/10.1109/WIFS.2010.5711468

[5] Mahdian B, Saic S. Detection of copy–move forgery using a method based on blur moment invariants. Forensic Science International, 2007, 171 (2 – 3): 180 – 189.

[6] Zhou P, Han X, Morariu VI, Davis LS. Learning rich features for image manipulation detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018), 2018: 1053 – 1061. IEEE.

[7] Zedan IA, Soliman MM, Elsayed KM, Onsi HM. Copy–move forgery detection techniques: a comprehensive survey of challenges and future directions. International Journal of Advanced Computer Science and Applications, 2021, 12 (7): 256 – 265.

[8] Li D, Zhu J, Fu X, Guo X, Liu Y, Yang G, Liu J, Zha ZJ. Noise-assisted prompt learning for image forgery detection and localization. In: Proceedings of the European Conference on Computer Vision (ECCV 2024), 2024: 18 – 36. Springer.

[9] Wang J, Wu Z, Chen J, Han X, Shrivastava A, Lim SN, Jiang YG. Object Former for image manipulation detection and localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022), 2022: 2354 – 2363. IEEE.

[10] Guillaro F, Cozzolino D, Sud A, Dufour N, Verdoliva L. TruFor: leveraging all-round clues for trustworthy image forgery detection and localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023), 2023: 20606 – 20615. IEEE.

[11] Qu C, Zhong Y, Liu C, Xu G, Peng D, Guo F, Jin L. Towards modern image manipulation localization: a large-scale dataset and novel methods. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2024), 2024: 10781 – 10790. IEEE.

[12] Guo X, Liu X, Ren Z, Grosz S, Masi I, Liu X. Hierarchical fine-grained image forgery detection and localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023), 2023: 3155 – 3165. IEEE.