

Detection of Image Tampering Using Multiscale Fusion and Anomalousness Assessment

Yichen WANG^a, Lijuan LIU^{b,1} and Tingting HUANG^b

^a School of Computer Communication and Engineering, Dalian Jiaotong University, Dalian, Liaoning, China

^b School of Software, Dalian Jiaotong University, Dalian, Liaoning, China

Abstract. In response to the growing importance of detecting maliciously altered images to mitigate their harmful effects, we propose a deep learning-based image tampering detection method that incorporates multiscale fusion and anomaly assessment. This approach addresses the limitations of existing methods that often struggle to detect diverse tampering types and exhibit suboptimal precision and localization performance. The proposed method employs a Channel-Spatial Attention module to enhance feature representations extracted from multiscale input images, thereby capitalizing on both spatial and channel-wise dependencies within the data. Furthermore, it uses a Z-score scoring mechanism and an LSTM-based mechanism to effectively capture and evaluate anomalous regions within the image. These components collectively contribute to a more robust identification of the manipulated content. For training supervision, we introduce a binary cross entropy loss, which jointly optimizes pixel-level classification and regression tasks, ensuring accurate tampering detection and localization. Experimental evaluations demonstrate that our method significantly outperforms prevailing tampering detection techniques, exhibiting an increase in AUC values ranging from 21% to 62% and achieving up to a 99.8% improvement in the best F1 score. Specifically, on benchmark datasets CASIA1.0, Coverage, and NIST16, our method attains F1 scores of 0.673, 0.714, and 0.981, respectively, underscoring its superior performance across diverse scenarios and tampering types.

Keywords. Image tampering, digital image blind forensics, multiscale fusion, anomaly detection, attention module

1. Introduction

Digital images serve as primary information carriers within communication platforms and are extensively acquired and disseminated by individuals^[1]. The advent of image editing software and deep generative models such as Generative Adversarial Networks (GANs) has rendered the manipulation of images increasingly facile. However, malicious tampering and dissemination by ill-intentioned actors pose significant challenges to societal security and stability, engendering harmful effects. Consequently, the field of image forensics has garnered escalating attention.

Digital image forensics is divided into two principal classes: active and passive forensics (also referred to as blind forensics)^[2]. Active forensics involves the a priori

¹ Corresponding Author: Lijuan Liu, liulj@djtu.edu.cn

embedding of information, such as digital signatures and watermarks, into digital images, enabling their integrity to be assessed during detection to determine whether manipulation has occurred. In contrast, passive or blind forensics operates without reliance on prior knowledge, relying solely on the intrinsic properties of the image itself for analysis.

Compared with the former, the latter approach offers broader applicability and lower costs, making digital image blind forensics the prevailing focus of current research endeavors. The methodology employed in this study is based on blind forensic techniques for image tamper detection^[3-5].

Image tampering techniques are currently categorized into three main categories: copy-move, splicing, and removal^[6], as depicted in Figure 1. Currently, there is an escalating and sustained interest among researchers within the academic sphere, with tampering detection methods evolving into both traditional and deep learning-based methodologies. Traditional approaches primarily leverage local image attributes, such as color, edges, and texture for detection. Notable algorithms encompassing these strategies include Scale-Invariant Feature Transform (SIFT)^[7], its enhanced version SURF (Speeded Up Robust Features)^[8], Discrete Cosine Transform (DCT)^[9], and Local Binary Patterns (LBP)^[10], among others.

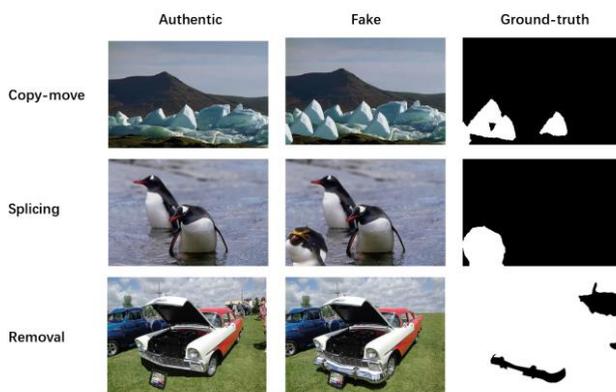


Figure 1. Examples of prevalent manipulation methods.

Although traditional tampering detection methods demonstrate commendable robustness against image noise and JPEG compression, they generally exhibit limitations in that they are often effective only in detecting tampering within scenarios involving single, uncomplicated objects. Among the current surge and maturation of deep learning frameworks, numerous methods have emerged that actively learn and extract tampering-specific features using deep learning models. These methods exhibit superior performance in complex regions containing multiple objects and have consequently seen widespread adoption.

In their scholarly contribution, Fu et al.^[11] proposed a novel methodology for identifying copy-move forgery in digital images that combines feature fusion with density-based clustering. This technique leverages the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to effectively filter out spurious correspondences, thus diminishing the incidence of false positive detections within the forgery identification procedure. Kwon et al.^[12] presented an innovative end-to-end

architecture known as the Compressed Artifacts Network, which combines both RGB and DCT signal streams. This network concurrently learns the forensic attributes associated with compression artifacts across these dual domains, thereby addressing the challenge of accurately pinpointing altered areas within manipulated images. Zhang et al. [13] presented a multitask network, denoted SE-Network, tailored for splice localization. This network incorporates a squeeze-and-excitation attention module to facilitate the fusion of features. Wu et al. [14] proposed an Iid-net designed to detect image removal tampering. The network relies on NAS-driven feature extraction modules that are optimized to uncover subtle yet distinctive traces indicating image removal tampering. Bappy et al. [15] constructed a hybrid neural network that intertwines LSTM (Long Short-Term Memory) units within an encoder-decoder structure. Hu et al. [16] proposed a spatial pyramid attention network (SPAN) that, by integrating a spatial pyramid attention mechanism, enables a focus on various scales of detail within the image, thereby enhancing the accuracy in pinpointing manipulated areas. Guo et al. [17] introduced a hierarchical fine-grained Tampering Detection Localization Network (HiFiNet), which addresses image manipulation stemming from conventional image editing techniques and deep generative models such as GANs. These methods address specific tampering techniques and exhibit inaccurate localization when confronted with input images of varying scales. In response to the problems presented in the literature, this paper presents MFDA-Net (Multiscale Fusion and Anomaly Detection Networks), an approach grounded in multiscale fusion and anomaly detection. MFDA-Net employs High-Resolution Representation Network (HRNet), a lightweight network architecture, as its base model to enhance training efficiency, upon which further improvements are made.

In the encoder component, the Convolutional Block Attention Module (CBAM) is integrated into the multiscale features of the image, thereby augmenting the network's capacity to discern tampered regions. Within the decoder portion, a Z-score mechanism and an LSTM mechanism are employed to capture and assess anomalous region features, enabling the network to more effectively detect anomalous regions in images and identify diverse tampering methods.

The contributions of this study can be summarized as follows:

- (1) The introduction of a multiscale approach to image tampering detection that is capable of addressing tampering in images of differing scale dimensions;
- (2) Use of the Z-score mechanism to recast the image tampering detection problem as anomaly detection, leveraging an LSTM network for evaluation, which permits the detection of images subjected to various tampering techniques;
- (3) The substantiation of the proposed method's advanced nature and superiority through evaluations on multiple datasets, assessing the tampering detection performance of the model.

2. MFDA-Net

This study introduces an anomaly detection network based on multiscale fusion to address the limitations of existing methods that handle only finite-scale variations. The HRNet serves as the backbone encoder, facilitating information exchange via dense connections across different scales and effectively integrating features from diverse scales, thereby enabling a more comprehensive extraction of tampering characteristics in images. The architecture of the proposed network is shown in Figure 2.

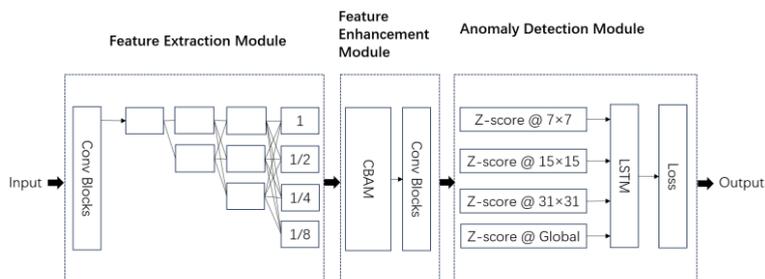


Figure 2. Architecture of the MFDA-Net model.

MFDA-Net comprises three modules: feature extraction, feature enhancement, and anomaly detection. In the feature extraction module, the HRNet is utilized, processing the input raw image to generate four outputs of progressively decreasing resolution, namely downsampling to one-fourth, one-half, one-eighth, and maintaining the original image size. The dense interconnections between these scales not only tackle scale variability but also facilitate information exchange, ensuring that each feature contains ample information.

The feature enhancement module employs a CBAM to further boost the extracted multiscale features by assigning spatial and channel weights, culminating in feature fusion through a concatenation operation.

The anomaly detection module standardizes local anomaly features in the image using the Z-score technique and uses an LSTM network to derive the final detection outcome by aggregating information from feature maps at multiple resolutions.

The anomaly detection module employs the Z-score method to normalize local anomaly features in the image and subsequently harnesses an LSTM network to assess and integrate the information derived from feature maps at multiple levels of resolution, thereby yielding the ultimate anomaly detection result.

2.1. CBAM

In the realm of deep learning, CBAM^[18] functions as an attention module implemented for tasks including, but not limited to, image classification and object localization. It endeavors to augment the model's inherent ability to discriminate and emphasize salient features throughout the visual task processing pipeline. Embedded within a convolutional neural network, CBAM comprises a two-step process featuring a Channel Attention Module (CAM) followed by a Spatial Attention Module (SAM). Both CAM and SAM compute separate attention coefficients along the channel and spatial dimensions, respectively, which are then pointwise multiplied with the incoming feature maps to refine and enhance their quality. This strategic refinement process significantly elevates the precision and effectiveness of the model. Figure 3 shows the structure of the CBAM module.

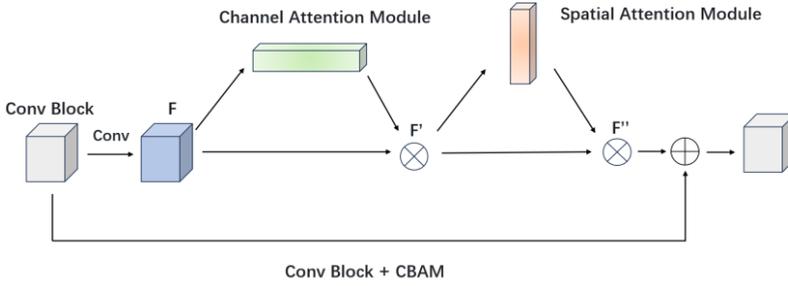


Figure 3. Structure of the CBAM module

The CAM utilizes global average pooling and fully connected layers to assess the significance of each individual channel within the input data. Based on these computations, it dynamically adjusts the weightings of the channels, thus applying a learned reweighting scheme to the input accordingly. Conversely, the SAM uses convolutional and pooling operations to assess the significance of each spatial location and applies multiplicative adjustments at those positions in the input.

In this study, for convolutional blocks within the feature extraction module, convolutional operations yield extracted features F , which are then subjected to CAM to obtain channel-wise attended features M_c . Subsequently, a tensor multiplication operation combines M_c with the original features F , yielding optimized features F' . This process is followed by a similar operation involving F' and features that have undergone spatial attention optimization, ultimately producing the final optimized feature map F'' . The computational procedure is as follows:

$$\begin{aligned} \mathbf{F}' &= \mathbf{M}_c(\mathbf{F}) \otimes \mathbf{F}, \\ \mathbf{F}'' &= \mathbf{M}_s(\mathbf{F}') \otimes \mathbf{F}' \end{aligned} \quad (1)$$

where \otimes denotes the tensor multiplication operation.

Through these steps, the CBAM module adaptively selects and attends to the most salient channels and positions within the feature map, conditioned on the specific input and task at hand. Compared with other attention mechanism modules, CBAM exhibits heightened explainability and flexibility, making it suitable for various visual tasks. The CBAM module has demonstrated its ability to significantly enhance model performance in many tasks.

2.2. Z-score

Z-score normalization, also referred to as standard score normalization, is a statistical technique that transforms raw data into a standardized form characterized by a zero mean and unit standard deviation, thereby enabling comparability among distinct datasets or among different features within the same dataset.

The calculation formula for the Z-score is as follows:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (2)$$

$$z = \frac{x - \mu}{\sigma} \quad (3)$$

where σ represents the standard deviation, μ denotes the mean, and x signifies the value of an individual feature.

In this study, the Z-score is used to compute anomaly scores for input data or intermediate features. Given that the Z-score quantifies the degree of deviation of each feature relative to a reference region—namely, the region where the feature is situated—it is imperative that this reference region be sufficiently large to avoid misclassifying features as anomalies, yet not so extensive as to encompass anomalous regions and thereby mistakenly label normal features.

To address this balance, a multi-resolution approach is adopted, with the reference region defined as four distinct regions of varying resolutions: 7×7 , 15×15 , 31×31 , and the original feature size. By calculating the Z-score, i.e., the deviation of each feature from its respective regional (or global) mean and standard deviation, saliently deviant features are effectively highlighted while minimizing the propensity for false positive detections. This, in consequence, facilitates the model's accurate focus on potential anomaly regions or critical features of interest.

2.3. LSTM

LSTM networks, constituting a variant of neural networks, employ gating mechanisms to retain long-range dependency information, frequently proving instrumental in addressing distant correlations within sequential data. Such capabilities are particularly advantageous for discerning intricate and multifaceted semantic content within images and determining whether they have been tampered with. Bunk et al. [19] were the first to introduce LSTM to the realm of image forensic analysis for tamper detection, feeding resampling-based image features into an LSTM to derive a feature vector used for assessing the presence of image manipulation.

In this study, Z-score deviations obtained after Z-score normalization across various resolutions are sequentially fed into LSTM units. This process yields predicted results regarding the tampered regions (anomalies) within the image.

2.4. Loss Function

The binary cross entropy loss function, denoted as L_{BCE} , is employed to supervise the detection and localization losses associated with the manipulated regions within the network. The standard mask is partitioned into four distinct scales corresponding to its original size, half-size, quarter-size, and eighth-size, labeled G1, G2, G3, and G4. Each of these scale-specific masks is subjected to loss computation against the corresponding generated prediction masks M1–M4. Subsequently, the final prediction mask M_p is paired with the ground truth mask GT for a separate loss calculation.

The overall loss function L_{total} can thus be represented as follows:

$$L_{\text{total}} = L_{\text{BCE}}(M_p, GT) + \frac{1}{4} \sum_{m=1}^4 (L_{\text{BCE}}(M_m, G_m)) \quad (4)$$

3. Experiments

3.1. Datasets

To evaluate the performance of the proposed model, we conducted experiments using the CASIA^[20], Coverage^[21], and NIST16 datasets^[22]. The division of the datasets is presented in Table 1.

Table 1. Dataset partitioning

Datasets	CASIA1.0	Coverage	NIST16
Training	1377	161	900
Testing	343	39	224
Total	1720	200	1124

The CASIA1.0 dataset comprises two categories of data: genuine and tampered images. Genuine images were sourced from the Corel image database, which offers a rich diversity of natural scenes. The tampered images, amounting to 921 in total, are generated through splicing operations applied to genuine images, thereby emulating prevalent splicing forgery practices encountered in actual image manipulation scenarios. The dataset's variety and complexity render it a crucial benchmark for research within the relevant domain.

The Coverage dataset comprises 100 spliced images, each constructed by incorporating at least one or more additional elements into an existing image containing one or multiple elements within the same scene. Notably, all spliced images in this dataset undergo specialized post-processing treatments.

The NIST16 dataset includes three types of tampering techniques: copy-move, splicing, and image inpainting, encompassing 564 manipulated images. Tampering operations within the dataset were post-processed to obscure visible artifacts, and tampered region masks were used for model assessment.

3.2. Experimental Settings and Evaluation Metrics

The proposed model uses the PyTorch deep learning framework and adopts an end-to-end training approach. All input images are consistently transformed into uniform dimensions of 256×256 pixels. Furthermore, the start of the learning rate parameter is set to 0.00001. The Adam optimizer was employed, and the graphics card used was an NVIDIA GeForce RTX 4060. As the backbone feature extraction network, the pre-trained hrnetv2_w32 model provided by HRNet official on ImageNet is used, retaining its original parameter configuration.

Tamper detection is classified as a classification problem because tampered regions typically constitute a portion of the entire forged image, necessitating the identification of whether each pixel in the image belongs to a tampered area. Consequently, the distribution of positive and negative samples is highly imbalanced. Thus, this study employs the F1-score and AUC (Area Under Curve) metrics to assess the performance of the proposed model. The F1 score serves as a pixel-level evaluation metric in image detection and is defined as follows:

$$F_1 = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (5)$$

In Equation (5), TP represents the number of correctly classified pixels within the tampered region, FP denotes the number of incorrectly classified pixels from the authentic region, and FN signifies the number of incorrectly classified pixels within the tampered region. The F1 score is a composite measure that considers both precision and recall, and it is the harmonic average of the two, with values ranging from 0 to 1, where 1 indicates the best performance.

AUC corresponds to the area under the ROC (Receiver Operating Characteristic) curve, serving as a rating for binary classification models, generally ranging from 0.5 to 1. Higher values of both the F1 score and AUC indicate superior localization performance of the model.

3.3. Contrastive Experiments

To demonstrate the performance of the proposed model in this work, it is compared with several prevailing mainstream tamper detection models, namely PSCC [23], ManTraNet [24], HiFiNet [17], and HRNet[25], using the NIST16 dataset. Corresponding AUC and F1 scores are computed for each model, as presented in Table 2.

Table 2. Comparative Results of Different Models on the NIST16 Dataset.

Methods	AUC	F1
PSCC	0.654	0.676
ManTraNet	0.493	0.980
HiFi-Net	0.628	0.491
HRNet	0.662	0.963
Ours	0.798	0.981

To assess the generalizability of the proposed model, experiments are conducted on three distinct datasets: CASIA1.0, Coverage, and NIST16. The corresponding AUC and F1 scores for the model's performance on each dataset are calculated and tabulated in Table 3.

Table 3. Performance of the Proposed Model Across Various Datasets

Methods	AUC	F1
CASIA1.0	0.602	0.673
Coverage	0.592	0.714
NIST16	0.798	0.981

An examination of Tables 2 and 3 reveals that the model presented in this study exhibits substantial improvement over various leading mainstream models, attaining the highest overall performance with respective AUC and F1 scores of 0.798 and 0.981. These results substantiate the notion that the methodologies adopted in this study have significant advantages over those employed in comparable existing studies.

The model demonstrates a slight variation in performance across different datasets. Specifically, the lowest F1 value is observed on the CASIA1.0 dataset, which is attributed to the fact that this dataset predominantly comprises small-sized images, thereby impeding precise tamper localization. Conversely, the model achieves its lowest AUC value on the Coverage dataset, likely due to the relatively fewer number of images contained therein, which may lead to insufficient learning of the specific characteristics inherent in the Coverage dataset, hence the lower AUC score.

3.4. Ablation Analysis

To evaluate the necessity of individual modules within the proposed model, an ablation study is conducted on the NIST16 dataset. Using AUC and F1 as evaluation metrics, the experimental outcomes are summarized in Table 4.

Table 4. Ablation Experiment Results

Methods	AUC	F1
HRNet	0.662	0.963
HRNet + CBAM	0.689	0.980
HRNet + SE ^[26]	0.548	0.979
HRNet + SK ^[27]	0.634	0.977
Ours	0.798	0.981

This study conducts five sets of experiments comprising the following configurations: the original HRNet backbone network, the addition of a CBAM attention mechanism, the integration of an SE attention mechanism, the inclusion of an SK attention mechanism, and the complete proposed model.

The experimental results reveal that the proposed model exhibits a marked enhancement compared with the bare HRNet backbone network. In the NIST16 dataset, the AUC and F1 scores escalated from 0.662 to 0.798 and from 0.963 to 0.981, respectively. In the comparative analysis of attention mechanisms, it is evident that the model incorporating the CBAM attention mechanism outperforms the others, exhibiting an AUC improvement of 26% over the model with the SE attention mechanism and a 9% increase in AUC relative to the model with the SK attention mechanism.

With respect to the incorporation of anomaly detection, the proposed model, when contrasted against the one solely augmented with the CBAM attention mechanism, registers a 16% boost in AUC. This substantiates the effectiveness of the anomaly detection module.

4. Conclusion

This paper presents an image tampering detection network based on multiscale fusion and anomaly assessment. Employing HRNet as the backbone architecture for the proposed model, the input images undergo feature extraction and enhancement through HRNet's multi-resolution processing, which selectively emphasizes critical channel-wise and spatial features. Subsequently, the extracted features are subjected to Z-score normalization and further aggregation via an LSTM in the anomaly detection module, culminating in the determination of the presence of tampering. Experimental findings demonstrate that, when compared against other prevailing approaches on the NIST16 dataset, as well as across evaluations on three public benchmark datasets—CASIA1.0, Coverage, and NIST16—this novel tampering detection method consistently exhibits substantial improvements, affirming its efficacy.

Acknowledgment

This work was supported by the Liaoning Provincial Natural Science Foundation General Project(2022-MS-341); and the Humanities and Social Sciences Research of

Dalian Jiaotong University - Supporting the Special Research Project on the Integration and Development of Humanities and Social Sciences (General Project).

References

- [1] Tyagi S, Yadav D. A detailed analysis of image and video forgery detection techniques. *The Visual Computer*. 2023 Mar;39(3):813-33.
- [2] Saber AH, Khan MA, Mejbil BG. A survey on image forgery detection using different forensic approaches. *Advances in Science, Technology and Engineering Systems Journal*. 2020 Jan;5(3):361-70.
- [3] Mushtaq S, Mir AH. Digital image forgeries and passive image authentication techniques: a survey. *International Journal of Advanced Science and Technology*. 2014 Dec;73:15-32.
- [4] Farooq S, Yousaf MH, Hussain F. A generic passive image forgery detection scheme using local binary pattern with rich models. *Computers & Electrical Engineering*. 2017 Aug 1;62:459-72.
- [5] Vega EA, Fernandez EG, Orozco AL, Villalba LJ. Passive image forgery detection based on the demosaicing algorithm and JPEG compression. *IEEE Access*. 2020 Jan 6;8:11815-23.
- [6] Mehrjardi FZ, Latif AM, Zarchi MS, Sheikhpour R. A survey on deep learning-based image forgery detection. *Pattern Recognition*. 2023 Jul 2:109778.
- [7] Amerini I, Ballan L, Caldelli R, Del Bimbo A, Serra G. A sift-based forensic method for copy-move attack detection and transformation recovery. *IEEE transactions on information forensics and security*. 2011 Mar 17;6(3):1099-110.
- [8] Dhivya S, Sangeetha J, Sudhakar BJ. Copy-move forgery detection using SURF feature extraction and SVM supervised learning technique. *Soft Computing*. 2020 Oct;24(19):14429-40.
- [9] Armas Vega EA, González Fernández E, Sandoval Orozco AL, García Villalba LJ. Copy-move forgery detection technique based on discrete cosine transform blocks features. *Neural Computing and Applications*. 2021 May;33:4713-27.
- [10] Bappy JH, Roy-Chowdhury AK, Bunk J, Nataraj L, Manjunath BS. Exploiting spatial structure for localizing manipulated image regions. In *Proceedings of the IEEE international conference on computer vision 2017* (pp. 4970-4979).
- [11] Fu G, Zhang Y, Wang Y. Image Copy-Move Forgery Detection Based on Fused Features and Density Clustering. *Applied Sciences*. 2023 Jun 26;13(13):7528.
- [12] Kwon MJ, Yu JJ, Nam SH, Lee HK. CAT-Net: Compression artifact tracing network for detection and localization of image splicing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision 2021* (pp. 375-384).
- [13] Zhang Y, Zhu G, Wu L, Kwong S, Zhang H, Zhou Y. Multi-task SE-network for image splicing localization. *IEEE Transactions on Circuits and Systems for Video Technology*. 2021 Oct 27;32(7):4828-40.
- [14] Wu H, Zhou J. IID-Net: Image inpainting detection network via neural architecture search and attention. *IEEE Transactions on Circuits and Systems for Video Technology*. 2021 Apr 22;32(3):1172-85.
- [15] Bappy JH, Simons C, Nataraj L, Manjunath BS, Roy-Chowdhury AK. Hybrid lstm and encoder-decoder architecture for detection of image forgeries. *IEEE transactions on image processing*. 2019 Jan 25;28(7):3286-300.
- [16] Hu X, Zhang Z, Jiang Z, Chaudhuri S, Yang Z, Nevatia R. Span: Spatial pyramid attention network for image manipulation localization. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16 2020* (pp. 312-328). Springer International Publishing.
- [17] 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 2023* (pp. 3155-3165).
- [18] Woo S, Park J, Lee JY, Kweon IS. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV) 2018* (pp. 3-19).
- [19] Bunk J, Bappy JH, Mohammed TM, Nataraj L, Flenner A, Manjunath BS, Chandrasekaran S, Roy-Chowdhury AK, Peterson L. Detection and localization of image forgeries using resampling features and deep learning. In *2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW) 2017 Jul 21* (pp. 1881-1889). IEEE.
- [20] Dong J, Wang W, Tan T. Casia image tampering detection evaluation database. In *2013 IEEE China summit and international conference on signal and information processing 2013 Jul 6* (pp. 422-426). IEEE.

- [21] Wen B, Zhu Y, Subramanian R, Ng TT, Shen X, Winkler S. COVERAGE—A novel database for copy-move forgery detection. In 2016 IEEE international conference on image processing (ICIP) 2016 Sep 25 (pp. 161-165). IEEE.
- [22] Guan H, Kozak M, Robertson E, Lee Y, Yates AN, Delgado A, Zhou D, Kheyrkhah T, Smith J, Fiscus J. MFC datasets: Large-scale benchmark datasets for media forensic challenge evaluation. In 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW) 2019 Jan 7 (pp. 63-72). IEEE.
- [23] Liu X, Liu Y, Chen J, Liu X. Pscn-net: Progressive spatio-channel correlation network for image manipulation detection and localization. *IEEE Transactions on Circuits and Systems for Video Technology*. 2022 Jul 7;32(11):7505-17.
- [24] Wu Y, AbdAlmageed W, Natarajan P. Mantra-net: Manipulation tracing network for detection and localization of image forgeries with anomalous features. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2019 (pp. 9543-9552).
- [25] Wang J, Sun K, Cheng T, Jiang B, Deng C, Zhao Y, Liu D, Mu Y, Tan M, Wang X, Liu W. Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*. 2020 Apr 1;43(10):3349-64.
- [26] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 7132-7141).
- [27] Li X, Wang W, Hu X, Yang J. Selective kernel networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2019 (pp. 510-519).