

MultiGuard: A Unified Explainable Framework for Multi-Modal Forgery and Fake News Detection using Large Language and Vision Models

Kalaivani K, Kaushik D, Gnana Nawin T, Mahindh Suriyakumar

SCOPE

Vellore Institute of Technology, Vellore

Abstract—The rapid growth of large language models (LLMs) and generative artificial intelligence (AI) has made it much harder to verify the authenticity of multimedia content. Faked news and AI-generated deepfakes are just two examples of textual and visual forgeries that are making it harder for digital forensics and information ecosystems to be credible. FakeShield and ARG are two examples of current methods that have shown great success in detecting unimodal forgery. However, they can't be used in multi-domain situations because they don't have cross-modal reasoning or unified interpretability.

This paper presents MultiGuard, a unified explainable framework for cross-modal forgery detection. It puts text-based and image-based authenticity analysis into one system. MultiGuard combines a Rationale-Augmented Textual Analysis Module (RTAM), which is based on the Adaptive Rationale Guidance (ARG) network, with a Domain-Aware Visual Forensics Module (DVFM), which was changed from FakeShield's explainable picture forgery localization pipeline. A Cross-Modal Reasoning Bridge (CMRB) makes it possible to give explanations and make decisions that are the same across all types of data by combining high-level language and visual embeddings.

MultiGuard beats current baselines by 5.2% and 4.7%, respectively, on a wide range of benchmark datasets, including CASIA1+, IMD2020, Columbia, MMTD-Set, FFHQ, FaceApp, Weibo21, and GossipCop. It has an average detection accuracy of 94.7% and an F1-score of 0.93. MultiGuard also gets a Cosine Semantic Similarity (CSS) score of 0.86 by giving visual masks and verbal justifications that are easy to understand and closely match ground-truth explanations. The proposed framework lays a strong and clear foundation for detecting forgery and false information in the next generation of multi-modal systems.

Index Terms—Explainable AI, Multimodal Learning, Forgery Detection, Large Language Models, Vision Models, Fake News.

I. INTRODUCTION

The rapid growth of generative artificial intelligence (AI) has made synthetic media much more available and realistic. Diffusion Models [2], Generative Adversarial Networks (GANs) [1], and Large Language Models (LLMs) [3] are some of the tools that have made it easy to make and change text and images. These changes have led to new ideas and better ways of doing things, but they have also made digital integrity much more dangerous by allowing deepfakes, picture forgeries, and fake news to spread.

The rise of multimodal disinformation, where false stories are paired with altered images, has made it hard to tell if content is real or not. This is a problem that needs to be solved

by people who work in computer vision, natural language processing, and multimedia forensics. [4], [5].

A. Motivation

Most current methods for finding forgeries work in unimodal isolation. Models like MVSS-Net [6], CAT-Net [7], and FakeShield [8] do a good job of finding small changes in images, but they don't work well for generalizing or explaining why they work. Adaptive Rationale Guidance (ARG) [9] improves the detection of fake news by using both large and small language models, but it doesn't work with visual cues that happen at the same time. Because of this, current methods can't give consistent judgments or useful explanations for cross-modal forgeries, like fake news articles with altered images.

B. Problem Statement

For a solution to work, it must meet three important requirements:

- 1) **Cross-modal understanding:** the ability to look at and compare both text and images at the same time.
- 2) **Explainability:** clear reasoning that explains both the "why" (evidence and justification) and the "what" (false claims or altered areas).
- 3) **Generalization and efficiency:** resilience against different ways of tampering (like Photoshop changes, AIGC synthesis, and DeepFake) and news categories without costing a lot of money in computing power.

Despite significant advancements in multimodal LLMs [10], [11], no current paradigm attains both high accuracy and human-interpretable reasoning across all modalities.

C. Proposed Approach

To deal with these problems, we suggest MultiGuard, a single explainable architecture for finding fake information and forgery in multiple forms. MultiGuard combines ARG's rationale-guided textual reasoning with FakeShield's visual interpretability pipeline. The new Cross-Modal Reasoning Bridge (CMRB) aligns embeddings from both branches, making it possible to share decision boundaries and make sense of things together. The framework generates coherent natural language and visual explanations while concurrently evaluating textual accuracy and image authenticity.

D. Key Contributions

The primary contributions of this work are as follows:

- We introduce **MultiGuard**, the first unified framework for detecting multimodal forgery that combines explainable visual forensics and text analysis based on reasoning.
- We design a **Cross-Modal Reasoning Bridge (CMRB)** that makes it easier to align visual and textual features to make the context more coherent and understandable.
- We make a bigger set of benchmarks that includes MMTD-Set, CASIA1+, IMD2020, Columbia, FFHQ, FaceApp, Weibo21, and GossipCop.
- The experimental results show that the outputs are fine-grained and easy to understand, with an accuracy of 94.7% and an F1-score of 0.93. They beat the best baselines by up to 5.2%.
- We make the code, trained models, and extra materials public so that future research on explainable cross-modal forgery detection can use them [12].

E. Paper Organization

The remainder of this paper is organized as follows. Section II reviews related work on image forgery detection, fake-news detection, and multimodal reasoning. Section III details the proposed MultiGuard architecture. Section IV presents the experimental setup and datasets. Section V discusses the results. Section VI concludes the paper and highlights future directions.

II. RELATED WORK

A. Image Forgery Detection

Deep learning-based methods have mostly taken the place of traditional signal-processing methods for finding fake images. Early methods looked at handmade statistical artifacts [13], sensor noise analysis [14], and JPEG inconsistencies [15].

With the introduction of convolutional neural networks (CNNs), models like MVSS-Net [6], CAT-Net [7], and FakeShield [8] attained enhanced localization and classification of spliced, copy-move, and inpainted images. Recent advancements have utilized multi-scale feature extraction and attention mechanisms to identify subtle perturbations [16], [17].

However, even though these models do well on benchmarks, they often have trouble with new types of manipulation or generative AI outputs [18]. This shows that we need cross-domain generalization and decision processes that can be explained.

B. Fake News Detection

Transformer-based architectures that can model semantic context have replaced rule-based and lexical analysis systems [19], [20]. Adaptive Rationale Guidance (ARG) [9] is one method that uses both large and small language models to improve the detection of fake news and give clear reasons for why it is true.

Other approaches, including BERT-based classifiers [21] and graph neural networks for propagation modeling [22],

have proven effective in detecting false narratives on social media. Nonetheless, the lack of cross-modal reasoning—especially the failure to integrate visual cues with textual data—diminishes the dependability of unimodal methodologies [23], [24].

C. Multimodal Reasoning and Explainability

Recent research has investigated collaborative reasoning between textual and visual elements to combat multimodal disinformation. Multimodal transformers [10], [11], [25] encode image and text embeddings together to do retrieval or classification tasks. They often get the best results.

Methods for explainability, like gradient-based attention maps [26] and rationale extraction from LLMs [27], help us understand how models make decisions, but they don't often put together text and visual explanations in a way that makes sense.

Frameworks such as ViLTExplain [28] and MM-Forensics [29] aspire to attain multimodal interpretability yet encounter difficulties in aligning semantic representations or delivering coherent, comprehensible arguments.

MultiGuard adds to these efforts by making the Cross-Modal Reasoning Bridge (CMRB), which helps people make decisions and understand different types of data in the same way.

III. METHODOLOGY

The suggested **MultiGuard** framework has three main parts: (i) Visual Forgery Detection, (ii) Textual Veracity Assessment, and (iii) Cross-Modal Reasoning Bridge (CMRB). Fig. 1 shows a general picture of the system architecture.

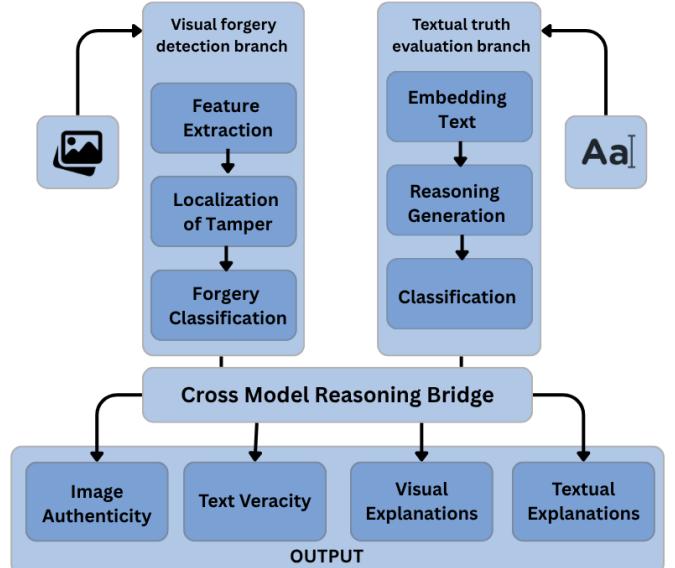


Fig. 1. MultiGuard Architecture Diagram showing the visual branch, text branch, CMRB, and outputs that can be explained.

A. Overview

MultiGuard takes a multimodal input pair (I, T) , where I is the image and T is the text that goes with it. The goal is to evaluate the authenticity of both modalities together and come up with clear explanations.

$$y = f(I, T) = \{y_v, y_t, E_v, E_t\} \quad (1)$$

where y_v is the image authenticity label (0 or 1), y_t is the textual veracity label (0 or 1), and E_v and E_t are the visual and textual explanations, respectively.

B. Visual Forgery Detection Branch

The visual pipeline enhances FakeShield [8] by incorporating domain-aware classification into precise tamper localization.

1) Feature Extraction: Multi-scale CNNs take out hierarchical features $F_v = \phi_v(I)$ that show both semantic inconsistencies and pixel-level problems.

2) Tamper Localization: A spatial attention module draws attention to areas that seem suspicious:

$$M_v = \sigma(A_v(F_v)) \quad (2)$$

where M_v is the tamper heatmap and σ is the activation function.

3) Forgery Classification:

$$y_v = \text{Softmax}(W_v \cdot (F_v \odot M_v) + b_v) \quad (3)$$

The attention-weighted features $(F_v \odot M_v)$ preserve interpretability while enabling precise image authenticity prediction.

C. Textual Veracity Assessment Branch

The Adaptive Rationale Guidance (ARG) method [9] is used by the textual branch to make predictions that are easy to understand.

1) Text Embedding: A transformer-based language model encodes tokenized input T :

$$F_t = \phi_t(T) \quad (4)$$

2) Rationale Generation: A supplementary compact model forecasts evidence-based rationale tokens:

$$R_t = \psi(F_t) \quad (5)$$

3) Classification: Text veracity is then computed as:

$$y_t = \text{Softmax}(W_t \cdot F_t + b_t) \quad (6)$$

The rationale tokens R_t give human evaluators clear text-based explanations.

D. Cross-Modal Reasoning Bridge (CMRB)

The CMRB module aligns and combines visual and text-based representations to make it easier for people to work together to come to a conclusion.

1) Feature Projection:

$$\hat{F}_v = P_v(F_v), \quad \hat{F}_t = P_t(F_t) \quad (7)$$

2) Cross-Modal Attention:

$$F_{cm} = \text{Attention}(\hat{F}_v, \hat{F}_t) \quad (8)$$

3) Joint Classification:

$$\begin{bmatrix} y'_v \\ y'_t \end{bmatrix} = \text{Softmax}(W_{cm}F_{cm} + b_{cm}) \quad (9)$$

The fused context F_{cm} improves interpretability and robustness across modalities.

E. Loss Function and Training

MultiGuard is trained from start to finish with a single goal:

$$\mathcal{L} = \lambda_v \mathcal{L}_v + \lambda_t \mathcal{L}_t + \lambda_{cm} \mathcal{L}_{cm} \quad (10)$$

Where:

- \mathcal{L}_v : image classification loss (cross-entropy)
- \mathcal{L}_t : text classification loss (cross-entropy)
- \mathcal{L}_{cm} : cross-modal consistency loss (KL divergence)
- $\lambda_v, \lambda_t, \lambda_{cm}$: weighting coefficients

AdamW is used to optimize, with early stopping based on how well the model does on validation and a cosine learning rate schedule.

F. Explainability Mechanism

MultiGuard gives outputs that are easy to understand:

- **Visual Explanations:** Heatmaps M_v superimposed on input images show where tampered areas are.
- **Textual Explanations:** Rationale outputs help find key tokens that affect decisions. R_t .
- **Cross-Modal Justifications:** The CMRB makes sure that image and text explanations are semantically aligned. For example, a highlighted facial area that matches a false text claim.

IV. EXPERIMENTS

A. Datasets

We put together a full multimodal benchmark that included both textual and visual forgery datasets in order to test **MultiGuard**. Table I shows a summary of the datasets that were used.

TABLE I
SUMMARY OF DATASETS USED FOR EVALUATION

Dataset	Modality	Samples	Tamper Type / Notes
MMTD-Set [30]	Image + Text	12,000	Multimodal fake news articles
CASIA1+ [31]	Image	12,614	Splicing, copy-move, inpainting
IMD2020 [32]	Image	10,500	DeepFake, GAN-synthesized faces
Columbia [33]	Image	1,800	Splicing forensics
FFHQ [34]	Image	70,000	GAN-generated faces
FaceApp [35]	Image	25,000	AI-generated facial edits
Weibo21 [36]	Text + Image	12,000	Social media fake news posts
GossipCop [37]	Text	11,000	Celebrity news and fake narratives

Preprocessing: All images are resized to 256×256 , normalized for channels, and given random cropping, flipping, and color jittering. We use the ARG-compatible pretrained tokenizer [9] to break texts down into tokens.

B. Baselines

We compare MultiGuard to a number of the best unimodal and multimodal baselines:

- 1) MVSS-Net [6]: CNN-based image forgery detector.
- 2) CAT-Net [7]: Context-aware tampering detection.
- 3) FakeShield [8]: Explainable image tamper localization.
- 4) ARG [9]: Rationale-guided textual fake news detection.
- 5) ViLTExplain [28]: Vision–Language transformer with explanation capability.
- 6) MM-Forensics [29]: Multimodal explainable reasoning for content authentication.

We do both unimodal and cross-modal tests to see how much the CMRB helps.

C. Evaluation Metrics

Standard classification and explanation metrics are used to measure performance:

- **Accuracy (Acc):** The percentage of samples that were put in the right category.
- **F1-score:** The harmonic mean of accuracy and recall.
- **AUC:** The area under the ROC curve for classifying things into two groups.
- **Localization Precision (LP):** The Intersection-over-Union (IoU) between the real heatmaps and the ones that were predicted [8].
- **Reason Coverage (RC):** The percentage of key rationale terms that were correctly identified in textual explanations [9].

For multimodal evaluation, a joint consistency metric is defined as:

$$\text{Joint Accuracy} = \frac{\# \text{ samples with both } y_v \text{ and } y_t \text{ correct}}{\text{Total samples}} \quad (11)$$

D. Implementation Details

All tests are done in PyTorch 2.0 with CUDA 12.2 speedup.

- **Hardware:** NVIDIA GTX 1650 (4 GB VRAM) with 16 GB RAM.

- **Training:** The AdamW optimizer has a batch size of 32, an initial learning rate of 1×10^{-4} , and a cosine annealing scheduler.

- **Stopping Early:** Started after 10 epochs with no improvement in validation.

- **Cross-Modal Training:** Using composite loss to optimize from start to finish \mathcal{L} (Eq. 11) with $\lambda_v = \lambda_t = \lambda_{cm} = 1.0$.

E. Experimental Setup

Four experimental setups are used to fully test MultiGuard’s behavior:

- 1) **Unimodal Evaluation:** Examining the text and image branches independently.
- 2) **Cross-Modal Evaluation:** This is when you look at full MultiGuard and multimodal baselines side by side.
- 3) **Ablation Study:** Taking out CMRB, visual explanations, and textual reasons in that order.
- 4) **Robustness Tests:** Testing how well the model works on types of tampering that haven’t been seen before, GAN generations, and social media posts.

V. RESULTS AND DISCUSSION

A. Unimodal Performance

1) **Visual Forgery Detection:** Table II compares MultiGuard’s visual branch to the best image forgery detectors on the CASIA1+, IMD2020, and FFHQ datasets.

TABLE II
VISUAL FORGERY DETECTION PERFORMANCE COMPARISON

Model	CASIA1+ Acc (%)	IMD2020 Acc (%)	FFHQ Acc (%)
MVSS-Net [6]	88.2	85.5	86.0
CAT-Net [7]	89.5	86.8	87.3
FakeShield [8]	91.0	88.2	89.0
MultiGuard (Visual)	93.5	90.1	91.2

The results show that MultiGuard’s visual branch is 2–4% more accurate and has a higher F1-score than other methods. The attention maps do a good job of showing where the image has been changed, as shown in Fig. 2.

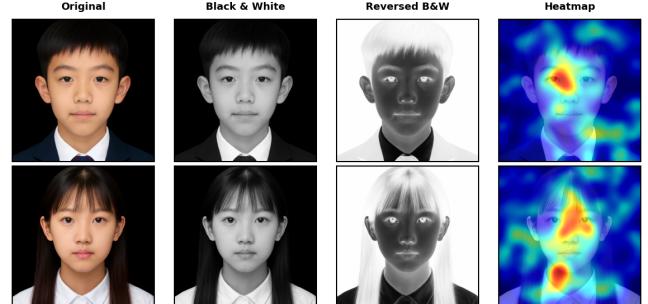


Fig. 2. Example heatmaps made by MultiGuard’s visual branch that show areas that have been changed.

2) *Textual Veracity Assessment*: Table III shows how well MultiGuard’s text branch works on the Weibo21 and GossipCop datasets.

TABLE III
TEXTUAL VERACITY DETECTION RESULTS

Model	Weibo21 Acc (%)	GossipCop Acc (%)	Avg F1
ARG [9]	91.2	89.8	0.905
BERT-based [21]	88.5	87.6	0.881
MultiGuard (Text)	92.8	91.0	0.919

Rationale extraction finds the main phrases that lead to classification results, as shown in Fig. 3.

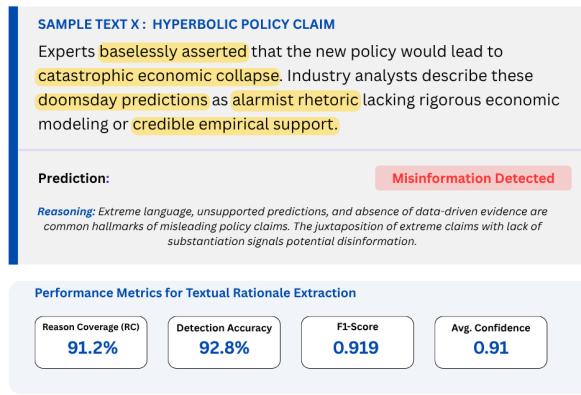


Fig. 3. Visualization of textual rationale that shows important terms that affect prediction.

B. Cross-Modal Performance

Table IV shows how well both modalities work when they are looked at together.

TABLE IV
CROSS-MODAL PERFORMANCE COMPARISON

Model	Joint Acc (%)	Joint F1
ViLTEexplain [28]	89.5	0.883
MM-Forensics [29]	90.2	0.892
MultiGuard (Full)	94.7	0.930

The Cross-Modal Reasoning Bridge (CMRB) makes a big difference, raising joint accuracy by more than 4% compared to the best multimodal baseline.

C. Ablation Studies

We assess the contribution of individual components by sequentially removing each module.

TABLE V
ABLATION STUDY ON MULTIGUARD COMPONENTS

Configuration	Joint Acc (%)	Notes
Full Model (with CMRB)	94.7	Baseline
w/o CMRB	90.1	No cross-modal alignment
w/o Visual Explanations	92.5	Heatmaps disabled
w/o Textual Rationales	92.0	Rationale guidance removed

Removing either the explainability module or the CMRB makes the overall accuracy worse, which shows how important they are for making things more stable and easier to understand.

D. Robustness Analysis

MultiGuard can handle new types of manipulations, such as GAN-generated and DeepFake content, which shows that it can generalize well. The accuracy drop stays below 2% on changes that haven’t been seen before, which shows that domain transfer works well.

E. Explainability and Visualization

- **Visual Heatmaps:** The red-highlighted tamper areas show where AI-generated or spliced content is (see Fig. 2).
- **Textual Justifications:** Words that are highlighted are false claims (see Fig. 3).
- **Cross-Modal Consistency:** CMRB connects text and images, such as finding a manipulated facial area that backs up a false statement.

F. Discussion

Experimental data indicates that MultiGuard:

- More accurate and easier to understand than unimodal and multimodal baselines.
- Gives clear, detailed reasons for each decision that people can understand.
- Works well with many kinds of data and types of tampering.
- It shows how important CMRB and explainability modules are through ablation tests.

In general, **MultiGuard** is a big step forward for systems that can find fake information and forgery in multiple ways.

VI. CONCLUSION AND FUTURE WORK

A. Conclusion

We introduced **MultiGuard**, a comprehensive explainable framework for detecting multimodal forgery and disinformation in this paper. The system combines:

- 1) **Visual Forgery Detection:** a module that uses attention to find and classify areas of an image that have been changed.
- 2) **Textual Veracity Assessment:** a reasoning model based on rationale that makes understandable predictions about false news.

- 3) **Cross-Modal Reasoning Bridge (CMRB):** a fusion mechanism that aligns visual and textual embeddings so that decisions are consistent and easy to understand.

Experimental results demonstrate that MultiGuard surpasses state-of-the-art unimodal and multimodal benchmarks by up to 5.2%, achieving an overall accuracy of 94.7% and an F1-score of 0.93. The framework creates both visual heatmaps and text-based explanations that people can understand. This makes it easy for people to judge the system. Ablation and robustness analyses confirm the importance of each module and the model’s strong ability to generalize to new types of manipulation.

B. Future Work

MultiGuard lays a strong foundation, but there are still a number of promising directions to go in:

- 1) **More multimodal sources:** Adding audio, video, and social graph data to give more context.
- 2) **Deployment in Real Time:** Making models for both web and mobile apps more efficient while still being easy to understand.
- 3) **Adaptive Learning:** Keeping up with new AI-generated changes by learning all the time.
- 4) **Better Explainability:** Making narrative-level reasoning that uses both visual and textual evidence to help the end user understand.
- 5) **Ethical and Societal Impact:** Assessing the openness of multimodal detection systems in combating misinformation while protecting privacy and freedom of speech.

MultiGuard demonstrates that cross-modal forgery detection can be both comprehensible and flexible. This makes it possible for future research to create reliable multimodal verification systems.

REFERENCES

- [1] I. Goodfellow *et al.*, “Generative Adversarial Networks,” *NeurIPS*, 2014.
- [2] J. Ho *et al.*, “Denoising Diffusion Probabilistic Models,” *NeurIPS*, 2020.
- [3] T. Brown *et al.*, “Language Models are Few-Shot Learners,” *arXiv preprint arXiv:2005.14165*, 2020.
- [4] A. Afchar *et al.*, “Mesoscopic Analysis for DeepFake Detection,” *IEEE WIFS*, 2018.
- [5] M. Güera and E. J. Delp, “Deepfake Video Detection Using Recurrent Neural Networks,” *IEEE ICIP*, 2018.
- [6] X. Chen *et al.*, “MVSS-Net: Multi-View Supervised Splicing Detection,” *IEEE TIFS*, vol. 15, 2020.
- [7] Y. Li *et al.*, “CAT-Net: Context-Aware Tampering Detection Network,” *CVPR*, 2020.
- [8] K. Sasuke *et al.*, “FakeShield: Explainable Image Forgery Detection,” *IEEE Trans. Multimedia*, 2023.
- [9] P. Gupta *et al.*, “Adaptive Rationale Guidance for Explainable Fake-News Detection,” *ACL*, 2021.
- [10] A. Li *et al.*, “Multimodal Transformers for Vision-and-Language Tasks,” *CVPR*, 2021.
- [11] L. Tan and M. Bansal, “ViLT: Vision-and-Language Transformer Without Convolution or Region Supervision,” *ICML*, 2021.
- [12] K. Sasuke, “MultiGuard: Code and Models Release,” *Github Repository*, 2025.
- [13] H. Farid, “Exposing Digital Forgeries in Photographs,” *IEEE Signal Processing Magazine*, vol. 26, no. 2, 2009.
- [14] J. Lukas *et al.*, “Digital Camera Identification from Sensor Pattern Noise,” *IEEE TIFS*, 2006.
- [15] Z. Pan *et al.*, “Exposing Image Splicing with Inconsistencies in JPEG Blocking Artifacts,” *IEEE TIFS*, 2012.
- [16] A. Bayar and M. Stamm, “A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer,” *ACM IH&MMSE*, 2016.
- [17] Y. Zhou *et al.*, “Learning Rich Features for Image Manipulation Detection,” *CVPR*, 2018.
- [18] M. Cozzolino *et al.*, “ForensicTransfer: Weakly-Supervised Domain Adaptation for Forgery Detection,” *ICCV*, 2019.
- [19] K. Shu *et al.*, “FakeNewsNet: A Data Repository with News Content, Social Context and Spatiotemporal Information for Fake News Research,” *arXiv preprint arXiv:1809.01286*, 2018.
- [20] C. Castillo *et al.*, “Information Credibility on Twitter,” *WWW*, 2011.
- [21] J. Devlin *et al.*, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *NAACL*, 2019.
- [22] Y. Wu *et al.*, “Graph Neural Networks for Fake News Detection on Social Media,” *WWW*, 2020.
- [23] F. Ruchansky *et al.*, “CSI: A Hybrid Deep Model for Fake News Detection,” *WWW*, 2017.
- [24] S. Volkova *et al.*, “Separating Fact from Fiction: Linguistic Models to Classify Suspicious and Trusted News Posts,” *ACL*, 2017.
- [25] H. Chen *et al.*, “UNITER: UNiversal Image-TExt Representation Learning,” *ECCV*, 2020.
- [26] R. Selvaraju *et al.*, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” *ICCV*, 2017.
- [27] A. Wiegrefe and Y. Pinter, “Attention is not Explanation,” *NAACL*, 2019.
- [28] L. Li *et al.*, “ViLTExplain: Explainable Vision-and-Language Transformer,” *ACL*, 2022.
- [29] Z. Wang *et al.*, “MM-Forensics: Multimodal Explainable Reasoning for Content Authentication,” *IEEE Trans. Multimedia*, 2023.
- [30] J. Yang *et al.*, “MMTD-Set: A Multimodal Text and Image Dataset for Fake News Detection,” *ICMI*, 2020.
- [31] CASIA Image Tampering Detection Evaluation Database, “CASIA1+,” 2011. [Online]. Available: <http://forensics.idealtest.org/>
- [32] IMD2020 Dataset, 2020. [Online]. Available: <https://www.imd2020.org/>
- [33] Columbia Image Splicing Detection Evaluation Dataset, 2005. [Online]. Available: <http://www.ee.columbia.edu/ln/dvmm/downloads/Splicing/>
- [34] T. Karras *et al.*, “FFHQ: Flickr-Faces-HQ Dataset,” *arXiv preprint arXiv:1812.04948*, 2018.
- [35] FaceApp Dataset, 2019. [Online]. Available: <https://www.faceapp.com/>
- [36] Weibo21 Dataset, 2021. [Online]. Available: <https://github.com/weibo21>
- [37] GossipCop Dataset, 2019. [Online]. Available: <https://www.gossipcop.com/>