# Unveiling Deception: A Frequency-Semantic Dual-Attention Network for Rumor Detection

1st Yuxiang Chen
*College of Cyber Science and Technology*
*Hunan University*
Changsha, China
yuxiang_chen@hnu.edu.cn

2nd Yang Tang
*College of Cyber Science and Technology*
*Hunan University*
Changsha, China
yangtangcs@hnu.edu.cn

3rd Shihui Gao
*College of Cyber Science and Technology*
*Hunan University*
Changsha, China
gaoqaq@hnu.edu.cn

4th Hui Zhou
*College of Cyber Science and Technology*
*Hunan University*
Changsha, China
huizhoucsee@hnu.edu.cn

5th Zheng Qin*
*College of Cyber Science and Technology*
*Hunan University*
Changsha, China
zqin@hnu.edu.cn

6th Lu Ou*
*School of Journalism and Communication*
*Hunan University*
Changsha, China
oulu9676@gmail.com

***Abstract*—Multimodal rumor detection aims to automatically identify real or fake news, thereby mitigating the adverse effects of misinformation. Although existing methods are effective, they pay insufficient attention to frequency features and often fail to fully integrate multimodal features. To address these challenges, we propose the FS-DARD model. First, we extract frequency features from images using the discrete cosine transform and design a spatial-frequency fusion stream based on an attention mechanism to enable comprehensive interaction between different image features. Next, we develop an image-text fusion stream that incorporates textual semantic features into the discrimination process. Finally, we introduce a gating mechanism during feature fusion to select the most representative features for detection. Experiments on three public datasets, Weibo, Twitter, and GossipCop, demonstrate the superiority of our model. Now, our code is available at https://anonymous.4open.science/r/FS-DARD-76E2.**

***Index Terms*—multimodal, rumor detection, discrete cosine transform, attention mechanism.**

## I. Introduction

The rapid evolution of social media and digital communication platforms has fundamentally reshaped the global information landscape. While these technologies have democratized information access, they have also created a fertile breeding ground for the viral spread of rumor that is intentionally and verifiably false, designed to mislead readers. The unchecked dissemination of rumor poses a severe threat to modern society. At the social level, malicious rumor undermines the credibility of mainstream media and public institutions [29]. In terms of public safety, during public health incidents or natural disasters, the spread of rumors can directly threaten people's lives [30]. On political level, malicious people manipulate public opinion through algorithms, interfering with the fairness of elections or policy implementation [31].

Rumors typically have exaggerated headlines, strong emotional language, and manipulated images. Therefore, the researchers extracted features such as word frequency, part of speech, emotional tone, and image-text semantics, and then used deep learning models to identify a post as a rumor [32]. Such methods are called **Content-based Methods**. In the early days, content-based rumor detection uses only text information as input. Early detection methods relied solely on text content [36], [37], but these are no longer effective in the current environment where multimodal misinformation is widespread. Rumor detection shifted towards a multimodal approach. Qian et al. [19] detected rumors based on a multi-layered semantic relationship modeling approach. Qi et al. [27] focused extensively on the complementary information between different modalities, the enhancement of multimodal information, and the comparison between multimodal information. Lao et al. [15] utilized fourier spectrum analysis to obtain forgery clues. However, these detection methods often focus solely on either semantic information or frequency information, neglecting the possibility of detecting misinformation from other perspectives. For example, images generated by generative adversarial network(GANs) often have distinctive frequency representations [42].

Despite significant advances in rumor detection, existing detection systems face several critical bottlenecks. [35] i) Modern fake news is rarely text-only; it often incorporates Deepfake images or manipulated videos. ii) Most existing methods focus on the semantic information of images while ignoring their physical features, such as many forgery traces that can often be detected through frequency features. Effectively fusing features across different modalities remains a significant

*Corresponding Author.

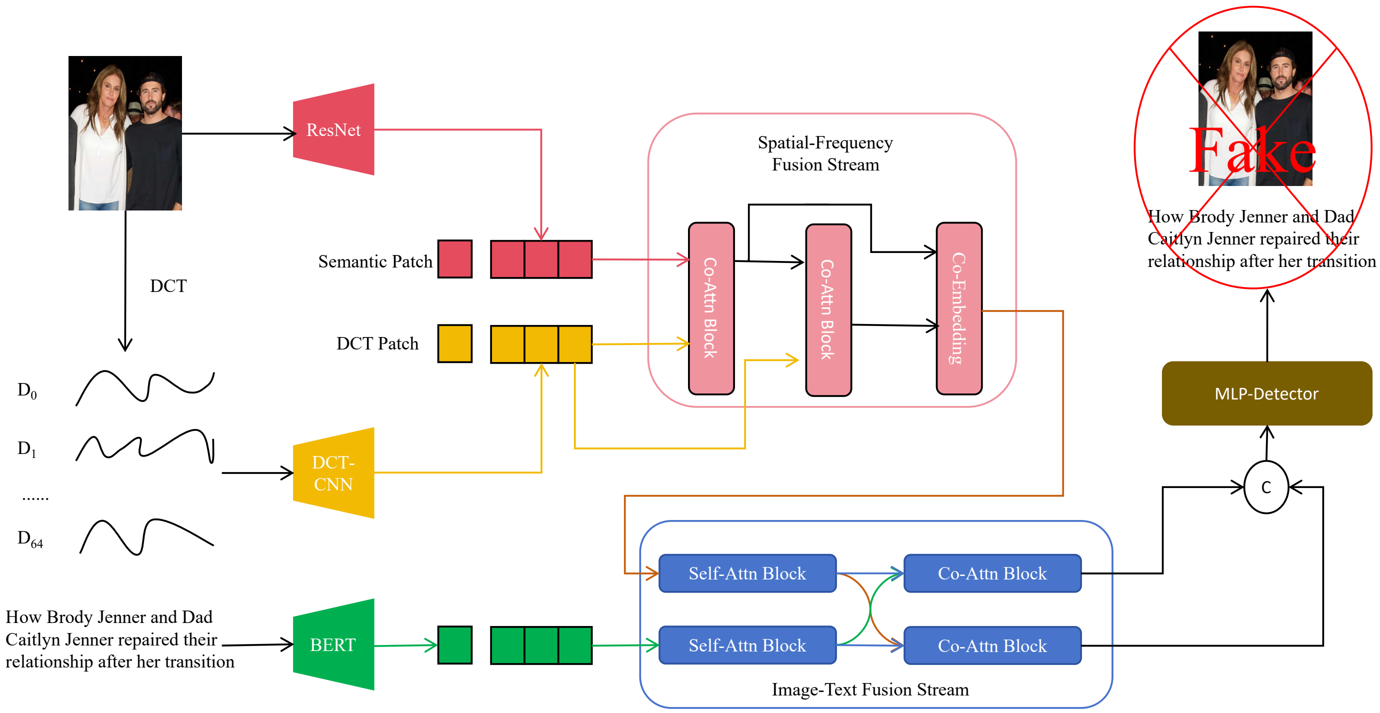


Fig. 1. FS-DARD framework detects multimodal rumors: images yield semantic (ResNet) and frequency (DCT-CNN) patches, fused via spatial-frequency blocks. Text (processed by BERT) fuses with visuals in image-text streams. An MLP-Detector aggregates features to label content ”Fake,” targeting cross-modal disinformation.

challenge.

To address the challenges mentioned above, this study proposes **Frequency-Semantic Dual-Attention Rumor Detector (FS-DARD)**. The primary contributions of our work can be summarized as follows:

- We obtained the image’s frequency features based on the discrete cosine transform and carefully designed a frequency feature extractor and a frequency-spatial fusion module.
- We designed image-text fusion module based on self-attention and cross-attention mechanisms, and optimized this module using a gating mechanism.
- Based on the modules we proposed, we present a novel end-to-end method to detect fake news on social media using only text and accompanying images.
- Extensive experiments on three datasets demonstrate that our model consistently outperforms state-of-the-art (SOTA) benchmarks, particularly on the imbalanced dataset of GossipCop, our method can also mitigate the negative impact of the data distribution and achieve better results than other models.

## II. Related Work

### A. DCT on Disinformation Detection

DCT is now widely applied in computer vision. Researchers have adapted DCT for disinformation detection to enable rapid identification of rumor patterns in the frequency domain. Kwon et al. [23] combined RGB and DCT streams to jointly learn compression artifact features in both the RGB and DCT domains, thereby achieving precise localization of image splicing regions. Zhu et al. [22] extracted and separated the high-frequency and low-frequency information of images through DCT and designed a parallel fusion stream to detect image manipulation. Dutta et al. [24] explored the combination of wavelet transform and DCT, utilizing the energy distribution across different frequency bands to quickly identify deepfake videos. However, the aforementioned studies applied DCT only to unimodal disinformation detection and did not investigate its application in scenarios involving the interaction of multiple information modalities. Wu et al. [5] applies DCT to multimodal rumor detection, it lacks selection and filtering of the DCT spectrum, leading to noise dominating the detection process.

Rumor detection and disinformation detection are two closely related but fundamentally different concepts. Although the definitions differ, the technical approaches of these two concepts overlap significantly in the field of Computer Vison, Natural Language Processing and Multimodal Learning [33], [34]. This is why we can apply the discrete cosine transform to rumor detection tasks.

### B. Multimodal Rumor Detection Methods

In addition to the content-based method introduced in the introduction section, there are three other methods that can be used for rumor detection. The **Propagation-based Methods** focus on how news spreads on social media. Liu et al. [25] propose the first real time rumor debunking algorithm and dataset Twitter15. Ma et al. [26] introduced the concept of propagation trees and developed Twitter16 dataset based on Twitter15. Zhang et al. [28] detect rumors by extracting social event elements and change them to Bipartite Ad Hoc Event Trees. In addition to the methods mentioned above, there are also **Social Context-based Methods** and **Knowledge-based Methods**. These methods are achieved by constructing a user relationship network [38], [39] and introducing external knowledge [40], [41], respectively. Propagation-based methods suffer from detection latency, requiring sufficient spread before analysis, and depend heavily on restricted structural data [43]. Social context-based methods excel by analyzing propagation patterns and user reactions but suffer from significant detection latency and vulnerability to social manipulation [44]. In contrast, knowledge-based methods provide high explainability by verifying claims against authoritative databases; yet, they are hindered by the incompleteness and time-lag of knowledge graphs, especially regarding breaking news [45].

## III. Methodology

### A. Overview of Problem and Model

There exists a pair of image and text $(I, T)$ representing news to be published. Our goal is to design a discriminative model that determines whether the news corresponds to a real event or a rumor. The model outputs 1 to indicate that the image-text pair is false, and 0 to indicate that it is true. FS-DARD is the proposed model, with its detailed architecture shown in Figure 1. FS-DARD takes spatial and spectral features of images, as well as semantic features of text, as inputs. It includes a spatial-spectral fusion module, an image-text fusion module, and a discriminator composed of a multilayer perceptron.

### B. Multimodal Feature Extraction

*a) Spatial Feature Extraction:* Spatial features are used to describe and analyze image content. To effectively extract these spatial features, we employed ResNet-50 [2] for feature extraction. The extracted features were then passed through an embedding layer to produce a set of tensors, denoted as $S = [S_1, S_2, ..., S_n]$. This set of tensors represents the spatial and semantic features of the image.

*b) Spectrum Feature Extraction:* The study [1] found that fake news often involves re-compressed or tampered images that show periodicity in the frequency domain. These forgery cues can be easily captured by convolutional neural networks (CNNs). Therefore, analyzing the image spectrum provides an alternative perspective for revealing image forgery cues. We carefully designed a **DCT-CNN Network** to capture forgery cues in the image spectrum. The network architecture, inspired by [3], [5], is illustrated in Figure 2. The DCT-CNN Network is a specialized feature extraction module that integrates 2D DCT with a multi-branch lightweight CNN; it functions by first converting input images into the frequency domain to isolate low-frequency coefficients, then processing these via a residual-based stem and three parallel convolution branches for multi-scale feature learning, and finally utilizing a Squeeze-and-Excitation (SE) attention mechanism [14] alongside global average pooling to generate a robust, flattened frequency-domain feature representation. Finally, the spectral features are also embedded as a set of tensors, denoted as $F = [F1, F2, ..., Fn]$.

*c) Text Feature Extraction:* Text semantic features are extracted by BERT [4]. Each word in the sentence is encoded into a feature vector, resulting in a series of patch embeddings. This is denoted as $T = [T_1, T_2, ..., T_n]$.

### C. Multimodal Feature Fusion

*a) Gated Attention:* Gated Attention [6] achieves selective attention to important information and filters out noise by adding a data-dependent gating signal after the value. In our model, we apply this mechanism to cross-attention in order to select crucial information from different modalities. For example, when reading news accompanied by images, we typically first concentrate on one modality, such as the title text. Then, we look for images related to that text to serve as supporting evidence. After examining the images, we assess the truthfulness of the news. Introducing an attention gate in this evaluation process means that, when processing information, our brains initially filter out non-essential details from both images and text, relying only on the key information to make a judgment. We apply gating exclusively to co-attention because self-attention inherently reinforces information.

Gated Attention's calculation method is similar to that of multi-head attention [10]. Firstly, We use linearly project $W$ on the input Query (Matrix $Q$), Key (Matrix $K$), and Value (Matrix $V$) to unify their dimensions to the model embedding dimension.

$$Q = Q * W_Q, K = K * W_K, V = V * W_V. \tag{1}$$

Then, split the projected Q, K, V into $h$ heads, where each head has a dimension of $d_i = d_{model}/h, i \in (1, h)$. Thirdly, compute attention weights for each head and weight V by these weights.

$$Y_i = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i, i \in (1, h). \tag{2}$$

Finally, we concatenate the outputs of all heads to get the result $Y$. After add a gate, our calculation method should be:

$$Y' = g(Y, X, W_\theta, \sigma) = Y \odot \sigma(XW_\theta), \tag{3}$$

where $X$ is the input used to calculate the gating score, $W_\theta$ is learnable parameters in gating, $\sigma$ is activation function and $Y'$ is the result of attention calculation.

*b) Attention Block and Fusion Stream:* Following the Transformer architecture, we combined the attention mechanism with Add&Norm and Feed-Forward Network (FFN) layers to get Self-Attn and Co-Attn blocks. Based on these two attention blocks, we designed two fusion streams. In Spatial-Frequency Fusion Stream, spatial feature embeddings and DCT feature embeddings are input into two Co-Attn Blocks to refine the correlation between spatial semantics and frequency domain patterns. Then, the Co-Embedding module integrates these into a unified spatial-frequency fusion feature, capturing

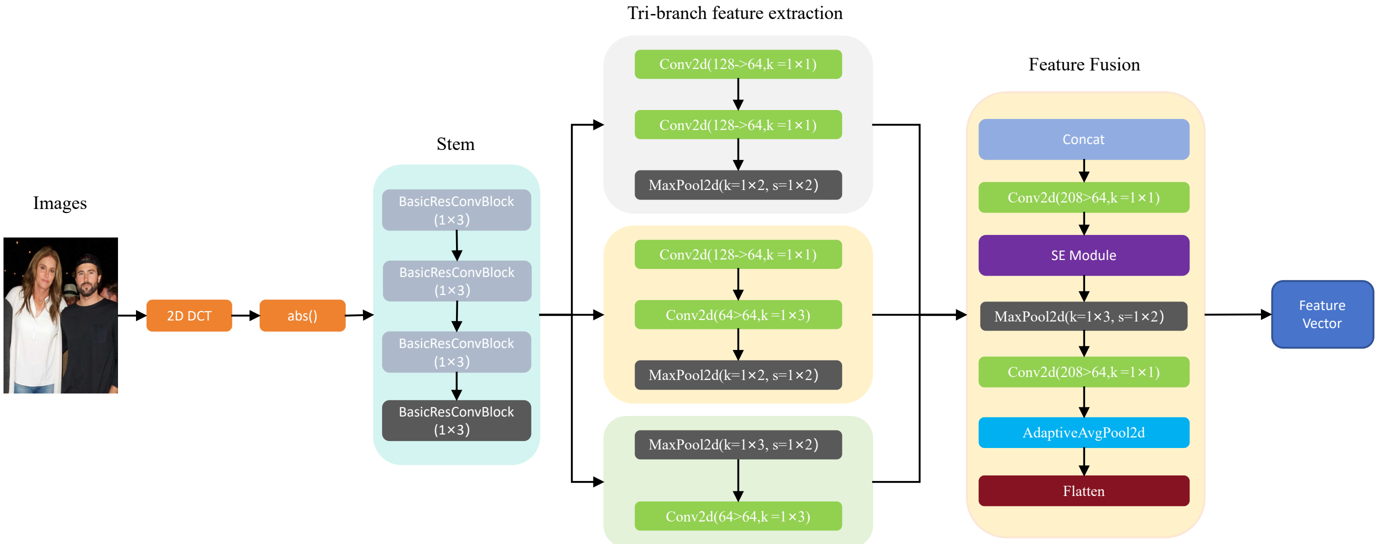


Fig. 2. DCT-CNN pipeline extracts image dct features. First, BasicResConvBlocks preliminarily extract features. Then, the features are sent to a tri-branch networks to extract features at different scales. Finally, a network with se module processes features further and yield feature vectors.

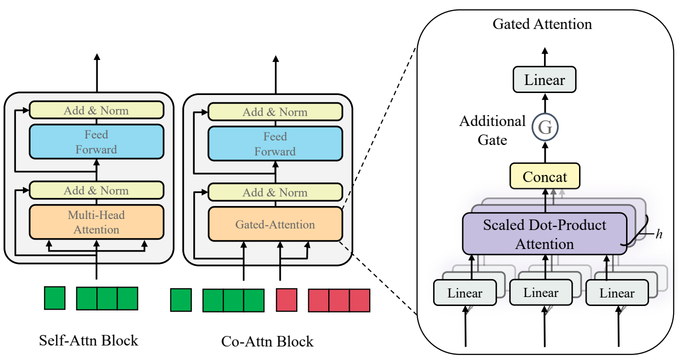


Fig. 3. Fusion block based on attention mechanism. We added a gating mechanism in the cross-attention mechanism.

both spatial content and frequency domain details. In Image-Text Fusion Stream, the Self-Attn Block first strengthens the logical connections within the text. Subsequently, the spatial-frequency fusion features are input into the Co-Attn Block to perform cross-modal attention matching between image and text. After multiple rounds of interaction, a cross-modal feature $R$ that fuses image spatial-frequency features and text semantic features is produced. This feature is used to determine the truthfulness of the news.

### D. Detection Head and Loss Function

We define a simple Multi-Layer Perceptron network $M$ as our prediction head, which consists of two linear layers and an activation function. We use the output of $M$ to represent the probability that a news image-text pair is fake news, this probability can be denoted as $\hat{y}$. The loss function is devised to minimize the cross-entropy value [7]:

$$L = -ylog * (\hat{y}) - (l - y) * log(1 - \hat{y}), \tag{4}$$

where y is the label. For datasets that have imbalanced labels, we use FocalLoss [8] for optimization.

$$FL = -\alpha_t * (1 - p_t)^{\gamma} * log(p_t), \tag{5}$$

where $p_t$ is the probability that the model correctly predicts the class. $(1 - p_t)^{\gamma}$ is modulating factor, $\alpha_t$ is balance factor.

## IV. Experiment

### A. Experiment Datasets and Settings

*a) Experiment Datasets:* We use three public real-world social media datasets: Weibo [12], Twitter [11], and Gossipcop [13]. These three datasets are commonly used in content-based rumor detection tasks. Weibo is a widely used Chinese dataset for rumor detection. The real news comes from China's authoritative news agency, Xinhua News. The Twitter dataset is also a well-known multimodal dataset used for rumor detection. The data inside comes from real users' tweets on the Twitter website. The Gossipcop dataset is a subset of FakeNewsNet, primarily focusing on fact-checking news in the entertainment sector. Before the experiments, we filtered out unimodal news without images or textual descriptions, news with videos, and non-English/non-Chinese news. Then we performed word segmentation and removed stop words from the text. Some metadata for the three datasets are shown in the Table I.

TABLE I
Metadata of Three Datasets.

| Head | *weibo* | *twitter* | *gossipcop* |
|---|---|---|---|
| rumor number | 4174 | 5371 | 2715 |
| non-rumor number | 3632 | 7258 | 9615 |
| train set number | 6118 | 11635 | 9863 |
| test set number | 1688 | 994 | 2467 |

*b) Experiment Settings:* Our parameter design is divided into general parameters and parameters set for specific datasets. In terms of general parameters, Our experiments are conducted on server with Linux operating system and single 4090. We set batch size to 32, dropout rate to 0.3, max text sequence to 128 and random seed to 777. In attention mechanism, the attention heads number and dropout rate are set to 12 and 0.1. The parameters set for the specific dataset include learning rate, backbone freezing epochs, training epochs and warmup number. All these parameters are shown in the Table II.

TABLE II
PARAMETERS SET FOR SPECIFIC DATASET.

| **Head** | ***weibo*** | ***twitter*** | ***gossipcop*** |
|---|---|---|---|
| epoch number | 30 | 15 | 50 |
| learning rate | 3e-6 | 5e-7 | 3e-6 |
| backbone frozen number | 3 | 15 | 3 |
| warmup number | 6 | 2 | 6 |

## B. Comparison Experiment

*a) Baseline and Evaluation Metrics:* To better validate the effectiveness of FS-DARD, we selected different baseline models for each dataset. These baseline methods are sourced from top conferences such as SIGIR, WWW, AAAI. These methods use cross-modal learning, fourier spectrum, and external information enhancement to achieve rumor detection. We evaluate these methods using four metrics: accuracy, recall, precision, and Micro-F1 score. As not all baseline models had been tested on all three datasets, we try our best reproduced the experiments and reported our reproducing results. The names, sources and experimental results of all models are presented in Table III.

Experimental results demonstrate that FS-DARD consistently outperforms all competing methods in terms of overall Accuracy across all datasets, achieving 0.954, 0.953, and 0.903 respectively. Notably, on the Gossipcop dataset, FS-DARD shows a significant improvement in the rumors F1-score (0.781) compared to the second-best model (0.680), representing a substantial gain in detecting minority class samples in imbalanced scenarios. While some baselines like ERIC-FND achieve competitive Precision in specific metrics, FS-DARD maintains the most robust and balanced performance across the majority of evaluation dimensions (indicated in red).

From the experiments, we can also see that the Twitter dataset, due to its small number of images and one image corresponding to multiple texts, allows all models to achieve good results. However, when the test scenario is switched to GossipCop dataset, most of these models struggle to finish their task. Our model not only got the highest accuracy, but also performed evenly, achieving the best results for both rumor cases and non-rumor cases.

## C. Ablation Experiment

To better evaluate the role of each module in the model, we remove each module from the entire model for comparison. "ALL" indicates that the model FS-DARD includes all modules, containing the frequency domain features (F), the spatial-frequency fusion stream module (SF), the image-text fusion stream module (IT), and the gating mechanism of the co-attn blocks (GA). We report the ablation experiment results in Table IV.

Table IV compares the performance of the FS-DARD method and its ablated variants across three datasets: Weibo, Twitter, and Gossipcop. FS-DARD achieves the highest accuracy and F1 on Weibo (0.954/0.953) and Twitter (0.953/0.950), and leads on Gossipcop (0.903/0.935). Removing any component—whether F, SF, IT, or GA—causes noticeable drops, especially on Gossipcop, where accuracy falls from 0.903 to as low as 0.618 without IT. These results confirm that every module contributes to the overall robustness of FS-DARD. Because nearly 50% of the images in the GossipCop dataset are manipulated celebrity faces, face swapping often leaves forgery traces in the frequency domain.

FS-DARD's superior accuracy stems from the synergistic interplay of its four components. In F, frequency features derived from DCT offer a novel perspective for detecting rumors. These features help us understand the details and structure of images, thereby detecting artifacts that are difficult to detect in news images. The SF effectively combines image spatial semantics with frequency features, allowing frequency features to truly play their role. The IT (Fusion) layer compresses text, images, and propagation structure information into a unified semantic space, mitigating single-modal noise. GA can filter noise to some extent, but the model is close to complete even without GA, so eliminating GA has the least impact.

## D. Visualization and Case Studies

Now we use the t-SNE tool [9] to visualize our classification results. We can observe from Figure 4 that after classification by our model, the isolation area between the two types of samples is more distinct, thanks to the effective multimodal feature extraction and attention-based fusion in our model. The application of these modules in our method is based on our in-depth understanding of the characteristics of multimodal rumors, and ultimately facilitates the distinction between fake news and real news.

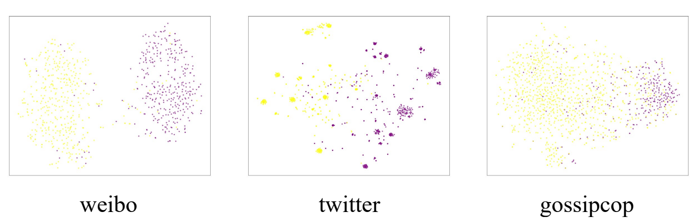


Fig. 4. Visualization of the classification results for the three datasets shows that our model can successfully identify rumors.

TABLE III
PERFORMANCE COMPARISON ON DIFFERENT MODELS AND DATASETS.

| Dataset | Method | Source | Accuracy | Rumor | | | Non-Rumor | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Precision | Recall | F1-score | Precision | Recall | F1-score |
| Weibo | CAFE [16] | WWW 2022 | 0.840 | 0.855 | 0.830 | 0.842 | 0.825 | 0.851 | 0.837 |
| | FSRU [15] | AAAI 2024 | 0.901 | 0.922 | 0.892 | 0.906 | 0.879 | 0.913 | 0.895 |
| | ERIC-FND [17] | AAAI 2025 | 0.946 | 0.985 | 0.914 | 0.948 | 0.908 | 0.984 | 0.944 |
| | RaCMC [18] | AAAI 2025 | 0.915 | 0.910 | 0.924 | 0.917 | 0.921 | 0.906 | 0.914 |
| | FS-DARD | Ours | 0.954 | 0.942 | 0.958 | 0.949 | 0.964 | 0.951 | 0.958 |
| Twitter | HMCAN [18] | SIGIR 2021 | 0.897 | 0.971 | 0.801 | 0.878 | 0.853 | 0.979 | 0.912 |
| | CAFE [16] | WWW 2022 | 0.806 | 0.807 | 0.799 | 0.803 | 0.805 | 0.813 | 0.809 |
| | FSRU [15] | AAAI 2024 | 0.952 | 0.983 | 0.938 | 0.960 | 0.901 | 0.984 | 0.940 |
| | ERIC-FND [17] | AAAI 2025 | 0.945 | 0.987 | 0.910 | 0.947 | 0.905 | 0.986 | 0.944 |
| | FS-DARD | Ours | 0.953 | 0.966 | 0.944 | 0.955 | 0.933 | 0.959 | 0.946 |
| Gossipcop | DistilBert [21] | ARXIV 2021 | 0.857 | 0.805 | 0.527 | 0.637 | 0.866 | 0.960 | 0.911 |
| | CAFE [16] | WWW 2022 | 0.867 | 0.732 | 0.490 | 0.587 | 0.887 | 0.957 | 0.921 |
| | FSRU [15] | AAAI 2024 | 0.849 | 0.729 | 0.636 | 0.680 | 0.898 | 0.913 | 0.905 |
| | RaCMC [18] | AAAI 2025 | 0.879 | 0.745 | 0.563 | 0.641 | 0.902 | 0.954 | 0.927 |
| | FS-DARD | Ours | 0.903 | 0.750 | 0.814 | 0.781 | 0.947 | 0.924 | 0.935 |

TABLE IV
ABLATION EXPERIMENT RESULTS.

| Method | Weibo | | Twitter | | Gossipcop | |
|---|---|---|---|---|---|---|
| | Accuracy | F1 | Accuracy | F1 | Accuracy | F1 |
| FS-DARD | 0.954 | 0.953 | 0.953 | 0.950 | 0.903 | 0.935 |
| -w/o F | 0.889 | 0.889 | 0.918 | 0.917 | 0.738 | 0.685 |
| -w/o SF | 0.894 | 0.893 | 0.945 | 0.944 | 0.745 | 0.697 |
| -w/o IT | 0.858 | 0.855 | 0.937 | 0.936 | 0.618 | 0.584 |
| -w/o GA | 0.909 | 0.908 | 0.958 | 0.956 | 0.858 | 0.799 |

This is real? Shark in NewJersey #HurricaneSandy#Sandy

In Nepal. Two and a half year old sister protected by four year old brother in Nepal

Fig. 5. Two typical examples of rumors.

Figure 5 shows two classic rumor cases. In the left image, a picture of a shark has been spliced into the water, and such manipulation techniques can easily be detected through spatial and frequency domains. The image on the right illustrates a classic case of semantic inconsistency between the text and the image. In fact, the two main people in this photo are Vietnamese and did not experience an earthquake.

## V. CONCLUSION

This paper proposes a novel Frequency-Semantic Dual-Attention Network to address the challenge of fusing text and image features for rumor detection. We utilize three different networks to extract features from frequency domain, spatial domain, and text domain, respectively. Then, two different fusion modules are designed based on attention mechanisms to fully leverage the different types of features. Experiments on three public rumor detection datasets validate the effectiveness of the proposed method, showing that FS-DARD is better than current state-of-the-art methods. In the future, we will continue to research content-based rumor detection methods and attempt to uncover more fine-grained forgery clues.

## ACKNOWLEDGMENT

This work is partially supported by the National Natural Science Foundation of China under Grant (No. 62441233, 62472166, U22A2030), and Technology Projects of Hunan Province (No. 2015TP1004)

## REFERENCES

[1] Qi, Peng, et al. “Exploiting multi-domain visual information for fake news detection.” 2019 IEEE international conference on data mining (ICDM). IEEE, 2019.

[2] He, Kaiming, et al. “Deep residual learning for image recognition.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[3] Szegedy, Christian, et al. ”Rethinking the inception architecture for computer vision.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[4] Devlin, Jacob, et al. “Bert: Pre-training of deep bidirectional transformers for language understanding.” Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers). 2019.

[5] Wu, Yang, et al. “Multimodal fusion with co-attention networks for fake news detection.” Findings of the association for computational linguistics: ACL-IJCNLP 2021. 2021.

[6] Qiu, Zihan, et al. “Gated Attention for Large Language Models: Non-linearity, Sparsity, and Attention-Sink-Free.” arXiv preprint arXiv:2505.06708 (2025).

[7] Shannon, Claude E. “A mathematical theory of communication.” The Bell system technical journal 27.3 (1948): 379-423.
[8] Lin, Tsung-Yi, et al. “Focal loss for dense object detection.” Proceedings of the IEEE international conference on computer vision. 2017.
[9] Maaten, Laurens van der, and Geoffrey Hinton. “Visualizing data using t-SNE.” Journal of machine learning research 9.Nov (2008): 2579-2605.
[10] Vaswani, Ashish, et al. “Attention is all you need.” Advances in neural information processing systems 30 (2017).
[11] Boididou, et al. “Verifying Multimedia Use at MediaEval 2016.” In MediaEval 2016 Workshop.
[12] Jin, Zhiwei, et al. “Multimodal fusion with recurrent neural networks for rumor detection on microblogs.” Proceedings of the 25th ACM international conference on Multimedia. 2017.
[13] Shu, Kai, et al. “Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media.” Big data 8.3 (2020): 171-188.
[14] Hu, Jie, Li Shen, and Gang Sun. “Squeeze-and-excitation networks.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
[15] Lao, An, et al. “Frequency spectrum is more effective for multimodal representation and fusion: A multimodal spectrum rumor detector.” Proceedings of the AAAI conference on artificial intelligence. Vol. 38. No. 16. 2024.
[16] Chen, Yixuan, et al. “Cross-modal ambiguity learning for multimodal fake news detection.” Proceedings of the ACM web conference 2022. 2022.
[17] Cao, Biwei, et al. “External Reliable Information-enhanced Multimodal Contrastive Learning for Fake News Detection.” Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 39. No. 1. 2025.
[18] Yu, Xinquan, et al. “Racmc: Residual-aware compensation network with multi-granularity constraints for fake news detection.” Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 39. No. 1. 2025.
[19] Qian, Shengsheng, et al. “Hierarchical multi-modal contextual attention network for fake news detection.” Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval. 2021.
[20] Han, Yi, Shanika Karunasekera, and Christopher Leckie. “Graph neural networks with continual learning for fake news detection from social media.” arXiv preprint arXiv:2007.03316 (2020).
[21] Allein, Liesbeth, Marie-Francine Moens, and Domenico Perrotta. “Like article, like audience: Enforcing multimodal correlations for disinformation detection.” arXiv preprint arXiv:2108.13892 (2021).
[22] Zhu, Xuekang, et al. “Mesoscopic insights: orchestrating multi-scale & hybrid architecture for image manipulation localization.” Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 39. No. 10. 2025.
[23] Kwon, Myung-Joon, et al. “CAT-Net: Compression artifact tracing network for detection and localization of image splicing.” Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2021.
[24] Dutta, Anurag, et al. “WaveDIF: Wavelet sub-band based Deepfake Identification in Frequency Domain.” Proceedings of the Computer Vision and Pattern Recognition Conference. 2025.
[25] Liu, Xiaomo, et al. “Real-time rumor debunking on twitter.” Proceedings of the 24th ACM international on conference on information and knowledge management. 2015.
[26] Ma, Jing, Wei Gao, and Kam-Fai Wong. “Detect rumors in microblog posts using propagation structure via kernel learning.” Association for Computational Linguistics, 2017.
[27] Qi, Peng, et al. ”Improving fake news detection by using an entity-enhanced framework to fuse diverse multimodal clues.” Proceedings of the 29th ACM international conference on multimedia. 2021.
[28] Zhang, Qi, et al. “Rumor detection with hierarchical representation on bipartite ad hoc event trees.” IEEE Transactions on Neural Networks and Learning Systems 35.10 (2023): 14112-14124.
[29] Groh, Matthew, et al. “Human detection of political speech deepfakes across transcripts, audio, and video.” Nature communications 15.1 (2024): 7629.
[30] Allen, Jennifer, Duncan J. Watts, and David G. Rand. “Quantifying the impact of misinformation and vaccine-skeptical content on Facebook.” Science 384.6699 (2024): eadk3451.
[31] Titcomb, J., and J. Carson. “What exactly is it–and how can you spot it? Fake news.” 2021,
[32] Tan, Li, et al. “Research status of deep learning methods for rumor detection.” Multimedia Tools and Applications 82.2 (2023): 2941-2982.
[33] Oshikawa, Ray, Jing Qian, and William Yang Wang. “A survey on natural language processing for fake news detection.” Proceedings of the Twelfth Language Resources and Evaluation Conference. 2020.
[34] Guo, Bin, et al. “The future of false information detection on social media: New perspectives and trends.” ACM Computing Surveys (CSUR) 53.4 (2020): 1-36.
[35] Ying, Qichao, et al. “Bootstrapping multi-view representations for fake news detection.” Proceedings of the AAAI conference on Artificial Intelligence. Vol. 37. No. 4. 2023.
[36] MA, Jing, et al. “Detecting rumors from microblogs with recurrent neural networks.(2016).” Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2016).
[37] Cheng, Mingxi, Shahin Nazarian, and Paul Bogdan. “Vroc: Variational autoencoder-aided multi-task rumor classifier based on text.” Proceedings of the web conference 2020. 2020.
[38] Shu, Kai, Suhang Wang, and Huan Liu. “Beyond news contents: The role of social context for fake news detection.” Proceedings of the twelfth ACM international conference on web search and data mining. 2019.
[39] Bian, Tian, et al. “Rumor detection on social media with bi-directional graph convolutional networks.” Proceedings of the AAAI conference on artificial intelligence. Vol. 34. No. 01. 2020.
[40] Thorne, James, et al. “FEVER: a large-scale dataset for fact extraction and VERification.” arXiv preprint arXiv:1803.05355 (2018).
[41] Hu, Linmei, et al. “Compare to the knowledge: Graph neural fake news detection with external knowledge.” Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: long papers). 2021.
[42] Tan, Chuangchuang, et al. ”Frequency-aware deepfake detection: Improving generalizability through frequency space domain learning.” Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 38. No. 5. 2024.
[43] Bian, Tian, et al. ”Rumor detection on social media with bi-directional graph convolutional networks.” Proceedings of the AAAI conference on artificial intelligence. Vol. 34. No. 01. 2020.
[44] Shu, Kai, Suhang Wang, and Huan Liu. ”Beyond news contents: The role of social context for fake news detection.” Proceedings of the twelfth ACM international conference on web search and data mining. 2019.
[45] Zhou, Xinyi, and Reza Zafarani. ”A survey of fake news: Fundamental theories, detection methods, and opportunities.” ACM Computing Surveys (CSUR) 53.5 (2020): 1-40.