# MVFNN: Multi-Vision Fusion Neural Network for Fake News Picture Detection

Junxiao Xue[1], Yabo Wang[1(✉)], Shuning Xu[1], Lei Shi[1(✉)], Lin Wei[1], and Huawei Song[2]

[1] School of Software, Zhengzhou University, Zhengzhou 450002, China
wangyb@stu.zzu.edu.cn, shilei@zzu.edu.cn
[2] Zhongyuan Network Security Research Institute, Zhengzhou 450002, China

**Abstract.** During this year's Novel Coronavirus (2019-nCoV) outbreak, the spread of fake news has caused serious social panic. This fact necessitates a focus on fake news detection. Pictures could be viewed as fake news indicators and hence could be used to identify fake news effectively. However, fake news pictures detection is more challenging since fake news picture identification is more difficult than the fake picture recognition. This paper proposes a multi-vision fusion neural network (MVFNN) which consists of four main components: the visual modal module, the visual feature fusion module, the physical feature module and the ensemble module. The visual modal module is responsible for extracting image features from images pixel domain, frequency domain, and tamper detection. It cooperates with the visual features fusion module to detect fake news images from multi-vision fusion. And the ensemble module combines visual features and physical features to detect the fake news pictures. Experimental results show that our model could achieve better detection performance by at least 4.29% than the existing methods in benchmark datasets.

**Keywords:** Fake news pictures · Deep learning · Multi-vision domain

## 1 Introduction

The rise of social platforms such as Weibo and Twitter not only brings convenience to users, but also provides soil for the breeding and dissemination of fake news. The frantic spread of fake news has had many negative effects. Take the Novel Coronavirus (2019-nCoV) in 2020 as an example, the spread of various fake news caused serious social panic during the virus outbreak. Fake news seriously harms the harmony and stability of society [1,2], which necessitates the effective automated fake news detection [3–5].

The work was supported by the National Key Research and Development Program of China: No. 2018******400, and the training plan of young backbone teachers in colleges and universities of Henan Province.


F. Tian et al. (Eds.): CASA 2020, CCIS 1300, pp. 112–119, 2020.
https://doi.org/10.1007/978-3-030-63426-1_12

Pictures are always important parts of the news. Studies have shown that the spread range of the news containing pictures is wider than the one without pictures by 11 times [6]. Fake news always use provocative pictures to attract and mislead readers as well. Therefore, an effective way to identify fake news pictures would help to detect the fake news. Actually, one important potential remedy for fake news recognition is to make use of the visual modal content of the news. Jin et al. found that the fake news pictures were statistically different from those of real news [6]. For example, the number of pictures illustrated in the news, the proportion of news containing hot pictures, and the proportion of special pictures (such as long pictures, chat screenshots, etc.) are also frequently used as statistical features for detection. Compared with fake pictures detection, fake news pictures detection is a more challenging task. This is because fake news pictures are more diverse. As shown in the Fig. 1, fake news pictures mainly have the following types: **1.** The tampered news picture: the picture is maliciously tampered to mislead the readers. **2.** The misleading picture: the picture itself is real but it is misinterpreted with the text description.

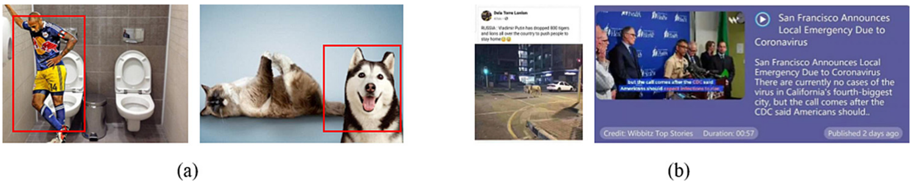


**Fig. 1.** Some fake news pictures, (a) are tampered news pictures. (b) are misleading pictures that pictures don't match text.

The main contribution of this paper is:

- A new fake news image detection model is proposed, which can effectively identify fake news images by combining image tampering information, semantic information, frequency domain information and statistical characteristics.
- The validity of the proposed model is verified by a large number of experiments on two real datasets.

The organization of the rest paper is as follows: the related works are reviewed in the next section. In Sect. 3, we provide a detailed description of our proposed model. The experimental results and analysis is presented in Sect. 4. Finally, our conclusions are summarized in Sect. 5.

## 2 Related Works

In the fake news detection task, in addition to text content, visual information is an important part of fake news detection. With the spread of multimedia content, researchers have begun to include visual information in the detection of fake news.

Some early methods based on machine learning used basic statistical characteristics [8], such as the number of pictures in the news, the proportion of news with popular pictures. However, these handmade characteristics time-consuming, laborious and limited to study complex patterns, lead to fake news detection task of poor generalization performance.

Visual forensic features are usually used in image processing detection. Some works extract visual forensic features to assess the authority of the attached images [7]. However, these forensic features are mostly hand-made and used to detect specific traces of manipulation, which is not applicable for fake news of real pictures [9].

Since the Convolutional Neural Network (CNN) has been verified to be effective in image classification [10], most existing multimedia content-based works use pre-trained deep CNNs, such as VGG19, to obtain general visual representations and fuse them with text information [10–13]. The first work incorporating multi-modal contents on the social networks via deep neural networks in fake news detection was reported in Ref [12], in which Wang et al. proposed an end-to-end event adversarial neural network to detect newly emerged fake news based on multi-modal features. Khattar et al. [13] proposed a new method to learn multi-modal information sharing representation for fake news detection. But due to the lack of task-related information, the visual features they adopted were too general to reflect the inherent characteristics of fake news pictures.The MVNN model proposed by Qi combined the information of pictures in both physical and semantic levels, however, it did not perform deep feature mining and was dedicated for particular field of fake new pictures [14].

In order to overcome these limitations, a new deep network combining visual modal features and physical features is proposed to detect fake news images.

## 3 The Proposed Model

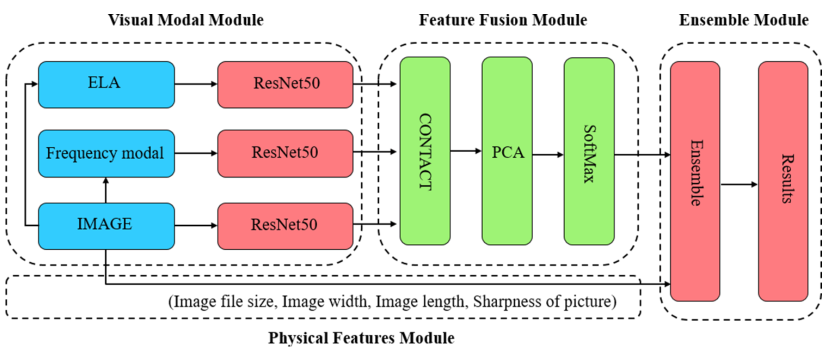


**Fig. 2.** The framework of our proposed MVFNN.

### 3.1 Model Overview

In this work, we designed a method which could conduct image tampering detection, image semantic detection and frequency domain detection simultaneously to explore the different visual patterns of fake news images in various visual modalities and extract effective features. At the same time, it combines the physical features such as the clarity, big or small and size of images to detect the fake news pictures. The model is shown in Fig. 2.

### 3.2 Model Derivation

In this part, we show and derive modules of the model, the details are as follows:

In the tamper detection part, we first apply the error-level analysis (ELA) algorithm on the input images. The main idea of ELA is based on the fact that the tampered region of the image is significantly different from the original one after a fixed-quality compression, and the location of the tampered region is obtained accordingly, as shown in Fig. 3. For the ELA processed images, we used the pre-trained ResNet50 to extract features. In addition, we added a 2048-neuron fully-connected layer to the ResNet50 in order to obtain the feature vectors denoted as $F_t = [v_1^t,\ v_2^t,\ v_3^t,\ \dots\ ,v_{2048}^t\ ]^T$.

In the semantic detection part, we directly use the pre-trained ResNet50 to extract the features of the input images. The extracted features can be expressed as $F_s = [v_1^s,\ v_2^s,\ v_3^s,\ \dots\ ,v_{2048}^s\ ]^T$.

Considering that the image recompression can be well reflected in the frequency domain, we first split it into three channels of RGB, and then conduct Discrete Cosine Transform (DCT) in three channels respectively. After that, they are combined together, and then a Fourier transform is performed on the picture to obtain the feature representation in the frequency domain. The features are subsequently used as the inputs to the ResNet50.

In the physical feature module, we use the size of the image file, the length of the image, the width of the image, and the quantification value of the sharpness of the image as the physical features to detect the fake news image, that is $F_p = [p_1,\ p_2,\ p_3,\ p_4\ ]$. To calculate the quantized value of the image, we first use the $3 \times 3$ Laplace operator to do the convolution operation on the original image, and then take the variance of the convolution operation result as the quantized value of the image.

In the feature fusion module, for the feature $F_t = [v_1^t,\ v_2^t,\ v_3^t,\ \dots\ ,v_{2048}^t\ ]^T$,$F_s = [v_1^s,\ v_2^s,\ v_3^s,\ \dots\ ,v_{2048}^s\ ]^T$ and $F_f = [v_1^f,\ v_2^f,\ v_3^f,\ \dots\ ,v_{2048}^f\ ]^T$ extracted by ResNet50, we firstly concatenate the features, namely get a new feature vector $F_c = [F_t,\ F_s,\ F_f]^T$. Considering the high feature dimension after extraction, we use PCA to reduce the dimension of extracted features by mapping the feature vector of 6144 dimensions to the feature vector of 1024 dimensions in order to obtain more compact features.

In the end, in the ensemble module, we combine the physical features of the image and the visual modal prediction results of the image, that is $F_a = [p,\ p_1,\ p_2,\ p_3,\ p_4\ ]$, and then use XGBoost to identify the final fake news image.

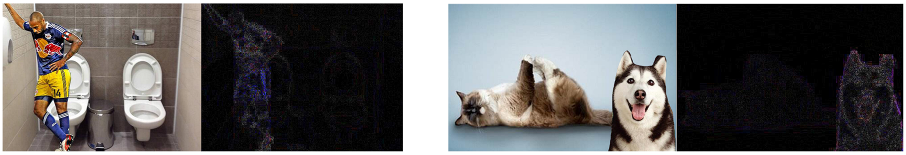

**Fig. 3.** Images processed by ELA. We can clearly see the tampered part.

## 4 Experiments

### 4.1 Datasets

In this paper we use the dataset (D1) [17] provided by the Zhiyuan Fake News Recognition Competition and the dataset (D2) used in the literature [16]. D1 contains 20459 real news pictures, 13636 fake news pictures, all images are from the news on Weibo. D2 contains 3725 real news pictures, 2804 fake news pictures, which are from the well-known authoritative news websites (such as The New York Times, The Washington Post, etc.)

### 4.2 Baselines

In this section, in order to evaluate the effectiveness of our proposed method, we selected several representative fake news picture detection methods as the baseline.

- **Forensics features (FF)+LR:** Ref [13] employed image forensics features in detecting fake news. Logistic regression algorithm was used as the classifier.
- **Pre-trained VGG19:** Pre-trained VGG is widely used as a feature extractor for detection of multi-modal fake news [13].
- **ConvAE:** Autoencoder (AE) is an artificial neural network learning efficient data representation in an unsupervised manner [15].
- **MVNN:** The MVNN model covers the visual content of the physical and semantic levels of pictures [14].

### 4.3 Performance Comparison

The comparison results between the proposed method and the baseline methods on D1 and D2 are shown in Tabel 1. It can be observed that the proposed MVFNN in this paper could significantly improve the ACC, F1 and AUC as compared with the baseline methods. For the D1 dataset, the accuracy was improved by 26.87%, 19.34%, 18.18%, 4.29% as compared with FF+LR, VGG19 (Pre-trained), ConvAE, MVNN, respectively. This shows that our model can capture the inherent characteristics of fake news pictures effectively.

**Table 1.** Performance comparison between different methods on the D1 and D2.

| Method | D1 | | | D2 | | |
|---|---|---|---|---|---|---|
| | ACC | F1 | AUC | ACC | F1 | AUC |
| FF+LR | 0.6654 | 0.6892 | 0.6603 | 0.6832 | 0.6763 | 0.6831 |
| VGG19(Pre-trained) | 0.7407 | 0.7911 | 0.7324 | 0.7083 | 0.7001 | 0.7034 |
| ConvAE | 0.7523 | 0.7674 | 0.7498 | 0.7328 | 0.7263 | 0.7315 |
| MVNN | 0.8912 | 0.9085 | 0.8902 | 0.8387 | 0.8172 | 0.8327 |
| **MVFNN(Ours)** | **0.9341** | **0.9453** | **0.9338** | **0.8853** | **0.8786** | **0.8832** |

### 4.4 Ablation Study

To visually demonstrate the effectiveness of different network components, we designed some internal model comparisons to simplify the MVFNN variant by removing some components:

- **Tamper:** features of the picture after ELA to detect fake news pictures.
- **Physical Features:** physical features are used to determine the authenticity of the image.
- **Tamper+Semantics:** features of tamper detection and semantic detection are integrated to identify fake news pictures.
- **Tamper+Semantics+Frequency Domain:** the features of the tamper detection are fused with the features of the semantic detection and the features of the frequency domain detection to identify fake news pictures.
- **PCA:** PCA is added to reduce the dimension of the fused feature.

The results of the ablation study are reported in Table 2 and we can find that each of network components is a contribution. Meanwhile, we compared the performance of Random Forests and XGBoost in the integration module,

**Table 2.** Individual performance of each module on D1 and D2.

| Method | D1 | | | D2 | | |
|---|---|---|---|---|---|---|
| | ACC | F1 | AUC | ACC | F1 | AUC |
| Tamper(ResNet50) | 0.8519 | 0.8800 | 0.8523 | 0.7687 | 0.7235 | 0.7638 |
| Physical Features(XGBoost) | 0.8910 | 0.9077 | 0.8858 | 0.8734 | 0.8724 | 0.8716 |
| Tamper+Semantics | 0.8960 | 0.9122 | 0.8913 | 0.8307 | 0.8120 | 0.8307 |
| Tamper+Semantics+ Frequency | 0.9012 | 0.9189 | 0.8998 | 0.8314 | 0.8138 | 0.8324 |
| PCA | 0.9083 | 0.9253 | 0.9023 | 0.8368 | 0.8214 | 0.8354 |
| MVFNN(RF) | 0.9298 | 0.9355 | 0.9301 | 0.8802 | 0.8654 | 0.8792 |
| **MVFNN(XGBoost)** | **0.9341** | **0.9453** | **0.9338** | **0.8853** | **0.8786** | **0.8832** |

and the experiments proved that XGBoost performs better on both datasets, which means that XGBoost is more suitable for fake news image detection.

### 4.5 Multi-feature Fusion

In this section, we evaluated different feature fusion schemes. Since some existing models used attention mechanism to fuse features, we employed PCA, Attention, PCA+Attention, and Attention+PCA to conduct the feature fusion. PCA represents that only PCA is used to reduce the dimension of features in the feature fusion part. Attention represents the use of an attentional mechanism to assign weight to the feature; PCA+Attention represents the one that applies PCA dimensionality reduction followed by the attention layer while Attention+PCA means applying attention first then conduct dimensionality reduction with PCA. The results are shown in Fig. 4.

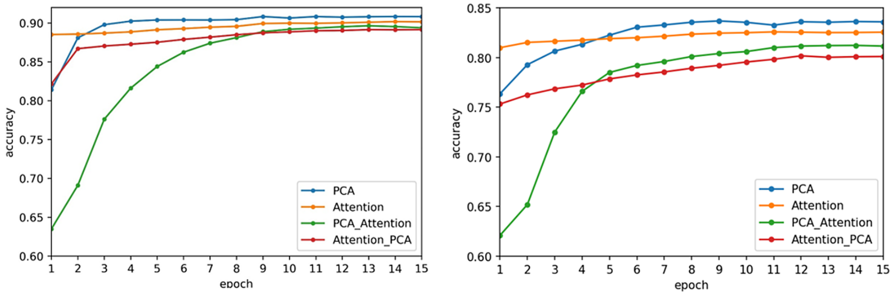


**Fig. 4.** Performance using different methods in feature fusion.

According to Fig. 4, it can be seen that PCA is superior to attention mechanism and other two combination methods in the feature fusion task. The possible reason is that the features extracted by each module are of equal importance, and hence attention mechanism does not work well in such case.

## 5 Conclusion

This paper proposes a fake news picture detection framework that combines multiple visual modalities and physical features. The proposed MVFNN consists of four main components: The visual modal module, the visual features fusion module, the physical features module and the ensemble module. The visual modal module is responsible for extracting image features from images pixel domain, frequency domain, and tamper detection. It cooperates with the visual features fusion module to detect fake news images from multi-vision fusion. The physical features module extract the physical feature such as the sharpness of images, and the ensemble module combines visual features and physical features to detect the fake news pictures. Experimental results show that our model could achieve better performance in benchmark datasets.

## References


1. Allcott, H., Gentzkow, M.: Social media and fake news in the 2016 election. J. Econ. Perspect. **31**(2), 211–36 (2017)
2. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake news detection on social media: A data mining perspective. ACM SIGKDD Explor. Newsl. **19**(1), 22–36 (2017)
3. Kumar, S., Shah, N.: False information on web and social media: a survey. In: Social Media Analytics: Advances and Applications (2018)
4. Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., Procter, R.: Detection and resolution of rumours in social media: a survey. ACM Comput. Surv. (CSUR) **51**(2), 1–36 (2018)
5. Wu, L., Morstatter, F., Hu, X., Liu, H.: Mining misinformation in social media. In: Big Data in Complex and Social Networks, pp. 135–162. Chapman and Hall/CRC), UK (2016)
6. Jin, Z., Cao, J., Zhang, Y., Zhou, J., Tian, Q.: Novel visual and statistical image features for microblogs news verification. IEEE Trans. Multimedia **19**(3), 598–608 (2016)
7. Boididou, C., et al.: Verifying multimedia use at MediaEval 2015. MediaEval **3**(3), 7 (2015)
8. Wu, K., Yang, S., Zhu, K. Q.: False rumors detection on sina weibo by propagation structures. In: 2015 IEEE 31st International Conference on Data Engineering, South Korea, pp. 651–662. IEEE (2015)
9. Ma, J., Gao, W., Wong, K.F.: Detect rumors on Twitter by promoting information campaigns with generative adversarial learning. In: The World Wide Web Conference, USA, pp. 3049–3055. ACM (2019)
10. Shu, K., Wang, S., Liu, H.: Beyond news contents: the role of social context for fake news detection. In: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, USA, pp. 312–320. ACM (2019)
11. Jin, Z., Cao, J., Guo, H., Zhang, Y., Luo, J.: Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In: Proceedings of the 25th ACM international conference on Multimedia, USA, pp. 795–816. ACM (2017)
12. Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Gao, J.: Eann: Event adversarial neural networks for multi-modal fake news detection. In: Proceedings of the 24th acm sigkdd International Conference on Knowledge Discovery and Data Mining, UK, pp. 849–857. ACM (2018)
13. Khattar, D., Goud, J.S., Gupta, M., Varma, V.: Mvae: multimodal variational autoencoder for fake news detection. In: The World Wide Web Conference, USA, pp. 2915–2921. ACM (2019)
14. Qi, P., Cao, J., Yang, T., Guo, J., Li, J. : Exploiting multidomain visual information for fake news detection. In: 19th IEEE International Conference on Data Mining, China, IEEE (2019)
15. Masci, J., Meier, U., Cireşan, D., Schmidhuber, J.: Stacked convolutional auto-encoders for hierarchical feature extraction. In: Honkela, T., Duch, W., Girolami, M., Kaski, S. (eds.) ICANN 2011. LNCS, vol. 6791, pp. 52–59. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21735-7_7
16. Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., TI-CNN, P. Y. : Convolutional Neural Networks for Fake News Detection. arXiv preprint arXiv:1806.00749(2018)
17. Datasets1. https://biendata.com/competition/falsenews. Accessed 23 Oct 2019