

SceneFND: Multimodal fake news detection by modelling scene context information

Journal of Information Science
2024, Vol. 50(2) 355–367
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/01655515221087683
journals.sagepub.com/home/jis


Guobiao Zhang 

School of Information Management, Wuhan University, China; Department of Computer Systems and Computation, Universitat Politècnica de València, Spain

Anastasia Giachanou

Department of Methodology and Statistics, Utrecht University, The Netherlands

Paolo Rosso

Department of Computer Systems and Computation, Universitat Politècnica de València, Spain

Abstract

Fake news is a threat for the society and can create a lot of confusion to people regarding what is true and what not. Fake news usually contain manipulated content, such as text or images that attract the interest of the readers with the aim to convince them on their truthfulness. In this article, we propose SceneFND (Scene Fake News Detection), a system that combines textual, contextual scene and visual representation to address the problem of multimodal fake news detection. The textual representation is based on word embeddings that are passed into a bidirectional long short-term memory network. Both the contextual scene and the visual representations are based on the images contained in the news post. The place, weather and season scenes are extracted from the image. Our statistical analysis on the scenes showed that there are statistically significant differences regarding their frequency in fake and real news. In addition, our experimental results on two real world datasets show that the integration of the contextual scenes is effective for fake news detection. In particular, SceneFND improved the performance of the textual baseline by 3.48% in PolitiFact and by 3.32% in GossipCop datasets. Finally, we show the suitability of the scene information for the task and present some examples to explain its effectiveness in capturing the relevance between images and text.

Keywords

Fake news detection; multimodal feature fusion; social media; visual scene information

1. Introduction

Social media have drastically changed the way that users read and follow news. People use social media platforms to read and share information about any topic ranging from political events, such as elections and referendums to natural disasters and pandemics [1,2]. For example, during COVID-19 pandemic, there was a huge increase in social media use [3], which were used to share information about topics mainly related to COVID-19. However, the rise in usage of social media platforms together with the anonymity they provide in sharing content converted them to one of the major ways of disseminating inaccurate and false information. In particular, fake news started gaining not only scientists' but also public's attention after two political events, the US 2016 Presidential campaign and Brexit.¹

The automated detection of fake news is not a trivial task. Early works on fake news detection included methods trained on information that was extracted from the textual content of the document, such as statistical text features and

Corresponding author:

Guobiao Zhang, School of Information Management, Wuhan University, Bayi Road 299, Wuhan, China.
Email: zgb0537@gmail.com

emotional signals [4]. More recent works have tried to incorporate information that represents the profile of the users [5–7], the credibility of the source [8] and contextual embeddings [9].

Although different researchers have explored the effectiveness of visual information on fake news detection [10–12], images contain additional visual information, such as contextual scenes that are still under-explored regarding their effectiveness on fake news detection. As statements of news events, both text and images are important components for completing narratives. Likewise, news text and images have a common narrative relationship that complements and echoes each other in the process of stating a particular reality, and follow the basic logic that images are attached to text, images and text imitate each other. News images usually show the core information of an event at the time of its occurrence, while news text provides a linguistic description of the event, both expressing the same semantic meaning from different sides. In terms of information presentation, text representations unfold along a fixed linear logic, while image representations are extended in a spatial dimension [13]. However, news events and spatial scenes tend to co-occur in the real world, for instance, traffic accidents usually occur on the street and not in a room. The regularity of event-scene co-occurrences helps us to give meaning quickly to our visual world. And, this regularity can also provide clues for fake news detection.

To this end, in this article, we propose Scene Fake News Detection (SceneFND), a multimodal system for fake news detection, which different to previous works, fuses information regarding the scenes that are captured in the images together with textual and visual representations of text and images. Our model extracts place (e.g. bedroom, mountain), weather (e.g. rain, snow) and season scenes (e.g. autumn, summer) from the images that are contained in fake and real news and evaluates their effectiveness on the problem of fake news detection.

In particular, in this study we are interested in addressing the following research questions:

- *RQ1*. Which are the top-N scenes contained in the fake and real news?
- *RQ2*. Can we use visual scenes to differentiate between fake and real news?
- *RQ3*. Which of the scenes are statistically significant different between fake and real news?

The rest of this article is organised as follows. Section 2 presents and discusses prior work on fake news detection. Section 3 presents a preliminary analysis regarding the visual scenes that appear in fake and real news. In Section 4, we present the SceneFND model that fuses textual, visual and visual scenes context information for fake news detection. Section 5 presents the data collection and the experimental setup we followed to conduct our experiments. Finally, in Section 6, we present and discuss the results, and we conclude the study in Section 7.

2. Related work

Fake news is an interdisciplinary problem and has attracted the research interest of different fields, including political and computer science [14]. Several studies have been published on the topic of fake news that are summarised in different surveys [14–17]. One of the most well-known studies in the field is by Vosoughi et al. [18] who collected and analysed a dataset of around 126,000 stories. Vosoughi et al. showed that false stories spread significantly farther, faster, deeper and more broadly than the truth in all categories of information.

Early works on fake news detection included methods that used information extracted or inferred from the text of the article, such as statistical text features [4] and semantic information [19]. The extracted information was then used to train machine learning and deep learning algorithms. Many studies have also proposed models that incorporate information, such as sentiment and emotion to address problems, such as credibility detection and misinformation-type classification [20,21]. In addition, others incorporated information inferred from users, such as their credibility or personality traits [5,6,22].

Due to the fact that the articles usually contain text and images, multimodal models can be also effective for fake news detection. Wang et al. [10] proposed the event adversarial neural networks (EANN) model that consists of the textual component represented by word embeddings and the visual that was extracted using the VGG-19 model pre-trained on ImageNet. Khattar et al. [11] proposed the multimodal variational autoencoder (MVAE) model based on bi-directional long short-term memory (bi-LSTMs) and VGG-19 for the text and image representation, respectively, whereas Singhal et al. [23] proposed a model based on bi-directional encoder representations from transformers (BERT) for the textual representation, and VGG-19 to learn the image features.

Given that the textual information of fake news may be not relevant compared with the images they contain, some researchers have tried to incorporate in their models this information represented by the similarity between the image and the text. To this end, Zlatkova et al. [24] explored the effectiveness of text-image similarity in addition to other visual information on the task of claim factuality prediction with respect to an image. In another study, Giachanou et al. [25]

combined textual, visual and information to detect fake news, where semantic information was represented by text–image similarity. In addition, Zhou et al. [26] proposed a similarity-aware fake news detection system based on a neural network that first obtains the latent representation of both its textual and visual information, based on which a similarity measure was defined. Then, the textual, visual and similarity representations were jointly learned and used to predict fake news. Different to the previous studies, we are interested in exploring the impact of the visual scenes that are in the images and if they can be used as an indicator of fake news. To the best of our knowledge, we are the first to explore the effectiveness of scene information for multimodal fake news detection.

3. Visual scene representation

In this section, first, we explain how we performed the visual scene extraction, and then, we conduct a preliminary data analysis to explore whether there are statistical significant differences in the usage of the different visual scenes contained in real and fake news.

3.1. Visual scene extraction

Scene is a term originally used in film and television to refer to the action that takes place in a specific time and space. Also, it can refer to the specific picture formed by the relationship between the characters, which is a specific process to express the plot through their actions. In research, the concept of scene is used in visual perception, and is typically defined as a semantically coherent view of a real world environment comprising background elements and multiple discrete objects arranged in a spatially licensed manner [27].

For the images that are contained in news articles, scene is not only a representation of a spatial location, but also contains features related to a specific space or behaviour, as well as human behaviour patterns and interaction patterns in this environment. A scene can contain various information, such as information for the weather (e.g. rain, snow) or locations (e.g. bedroom, suburb, industrial, kitchen). Information extracted from the scene data can be used to solve some of the ambiguity at the image semantic level [28]. For example, two images that visualise the same persons may express different meanings if they are captured in different scenes. Identifying the scenes from the images can provide additional context information necessary to understand the semantic content. In this article, we propose to explore three types of scenes: *place scenes*, *season scenes* and *weather scenes*. In particular, we extract and analyse the following scenes:

- *Place scenes*. Bedroom, suburb, industrial, kitchen, living room, coast, forest, inside city, highway, mountain, open country, street, tall building, office, and store.
- *Season scenes*. Autumn, summer, spring, and winter.
- *Weather scenes*. Mostly cloudy, clear, mainly clear, cloudy, rain showers, drizzle, moderate rain, rain, fog, rain fog, and snow.

3.2. Scene extraction results

3.2.1. Place scenes. To extract the place scenes, we used the Wilson's method² [29] that is based on DAISY [30] and a support vector machine (SVM). Figure 1 shows the places scenes together with the average probability of their presence in fake and real news in both PolitiFact and GossipCop datasets. From the figure, we observe that in GossipCop most of the scenes have similar presence probability for fake and real news. On the contrary, in PolitiFact, we see that the *office* scene appears with a higher probability in real news compared to fake ones.

3.2.2. Weather scenes. To detect the weather scenes, we used the weather image recognition model.³ Figure 2 shows the probabilities of the different weather scenes in fake and real news in PolitiFact and GossipCop. From the figure, we observe that there are no big differences regarding the presence probability of weather scenes in the datasets.

3.2.3. Season scenes. To extract the season scenes, we applied the SeasonsNNetwork model⁴ that has been proposed to predict the visual season scene. Figure 3 shows the probabilities of the different aspects of season scenes in fake and real news contained in PolitiFact and GossipCop. From the Figure 3(a), we observe that there are certain differences between fake and real news in spring, autumn and winter scenes, but there is almost no difference in summer scene. For GossipCop dataset, there is a few difference in the spring scene, and few difference in other season scenes.

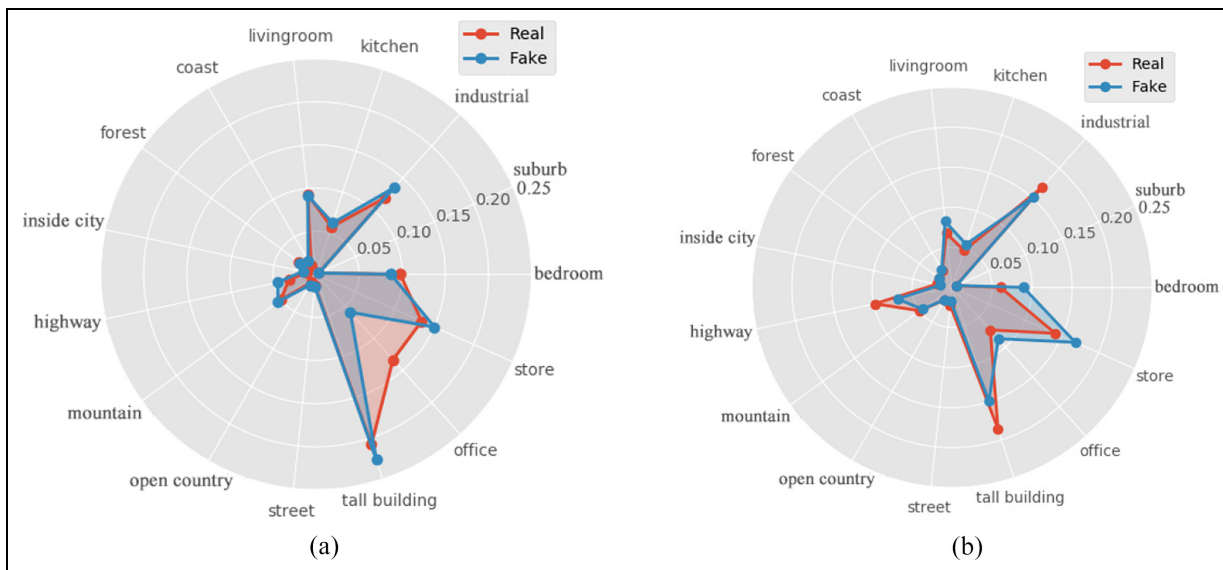


Figure 1. Average place scene scores for real and fake news of (a) PolitiFact and (b) GossipCop datasets.

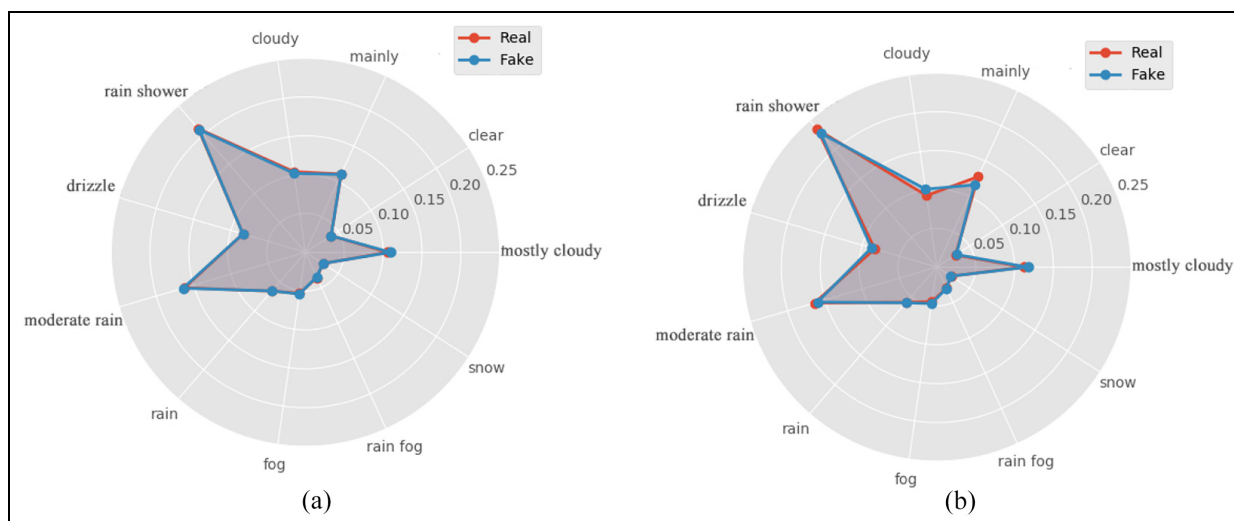


Figure 2. Average weather scene probability for real and fake news of (a) PolitiFact and (b) GossipCop datasets.

To address RQ1 (*Which are the top-N scenes contained in the fake and real news?*), we explore which are the scenes that co-occur more frequently in fake and real news. To find the top scene combinations, first we detect the scenes with the highest probability for each image and for each scene type. Then, we count the frequency of all the difference co-occurrence combinations in fake and real news. Table 1 shows the top three combination of scenes that co-occur in fake and real news in the two datasets. An interesting result from the table is that *store, rain showers, autumn* is a combination that occurs a lot in fake news of both datasets.

4. SceneFND model

In this section, we present our multimodal *SceneFND* model that aims to differentiate between fake and real news. Our model is based on a neural network and combines the following three different types of feature: textual, visual and scene representations. The architecture of *SceneFND* is depicted in Figure 4.

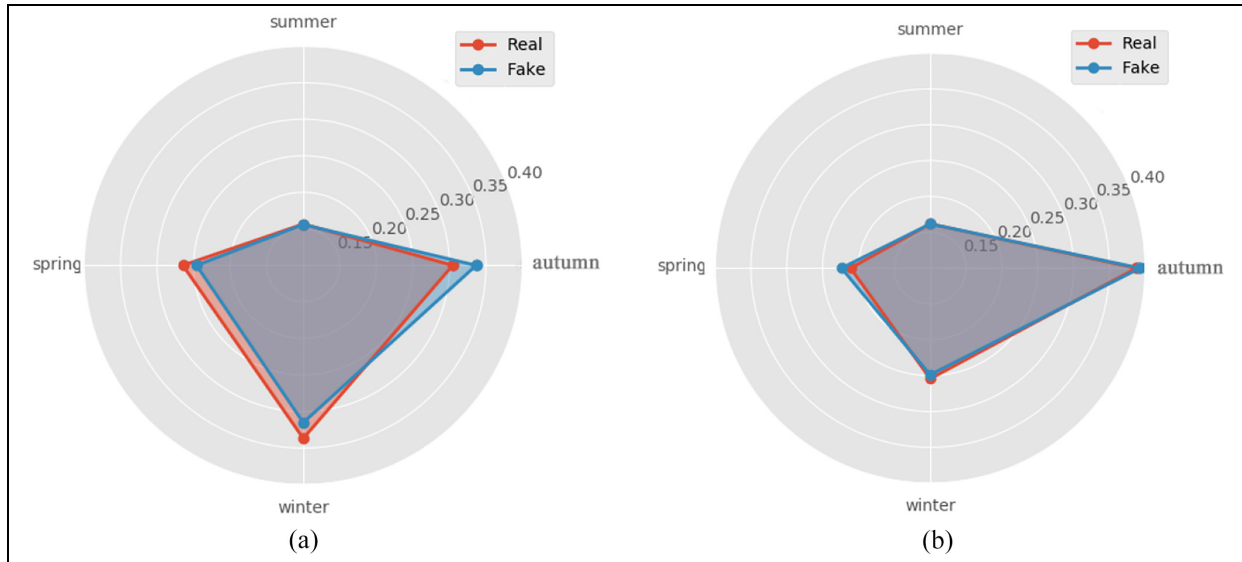


Figure 3. Average season scene scores for real and fake news of (a) PolitiFact and (b) GossipCop datasets.

Table 1. Scene co-occurrence ranking.

	Real	Fake
Rank	PolitiFact	
1	Tall building, rain showers, winter	Store, rain showers, autumn
2	Tall building, rain showers, autumn	Industrial, rain showers, winter
3	Industrial, rain showers, winter	Tall building, rain showers, autumn
Rank	GossipCop	
1	Tall building, rain showers, autumn	Tall building, rain showers, autumn
2	Office, rain showers, autumn	Store, rain showers, autumn
3	Store, rain showers, autumn	Tall building, rain showers, spring

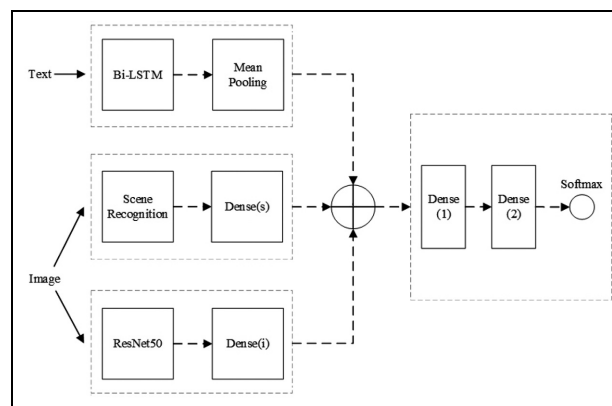


Figure 4. Architecture of the SceneFND multimodal fake news detection model.

4.1. Textual feature representation

The first component of our system is the textual representation. The textual content of the post or the article is the most important information for the detection of fake news. Textual information has been shown to be useful for a wide range

of text classification tasks, from reputation analysis to irony detection and from sentiment analysis to credibility detection [20,31]. Previous fake news detection approaches have used representations, such as bag-of-words and word embeddings [20,24]. However, the main limitation of those representations is that they cannot capture the contextual information. On the contrary, our system uses a more sophisticated representation that is based on recurrent neural network (RNN) with bi-LSTM units. Bi-LSTM has been proven to be effective in many text classification tasks [32]. For our textual feature extractor, first, we represent each word in the text as a word embedding vector. The embedding vector for each word is initialised with the pre-trained word embedding on the Google news dataset. Then, the sequence of hidden outputs is passed to a mean pooling layer over the time steps to produce a single temporal component as the textual feature.

4.2. Scenes feature representation

In the second component, we aim to capture the scene representations that are present in the image. To do this, we first calculate the probabilities on each scene category with different scene recognition methods. In particular, we use Wilson's method [29] for the place scene, the weather image recognition model for the weather scenes and the SeasonNetwork for the season scene extraction, as already mentioned before. We also seek a consistent representation to get the same (small) dimension for every tweet since it will ultimately be used in the *feature fusion* component. To this end, we decided to concatenate the probabilities of the different scene types in one vector and apply a fully connected layer to extract scene vector representation of each tweet.

4.3. Visual feature representation

In general, news articles contain one or more images, and the semantic expression of the image is often closely related to the content of the news. Based on this, we also take into account the visual representation of the images contained in the news. For the visual representation, we extract the output of the second last layer of ResNet-50 network pre-trained on the ImageNet dataset and pass it through a fully connected layer to represent final visual feature.

4.4. Feature fusion

The last step of the model is to fuse the three different representations into a single one. The three feature vectors obtained via different modalities are fused together using a simple concatenation technique to obtain the desired representation. This news representation is then passed through two fully connected neural networks for fake news classification. The softmax function is applied to the output layer to derive a probability representation for each category.

5. Experimental setup

In this section, we present the details of the experimental setup including a description of the dataset, the parameter settings and baselines.

5.1. Dataset

To perform our experiments and evaluate the effectiveness of the SceneFND, a multimodal fake news dataset had to be used. To conduct the initial scene analysis, we use FakeNewsNet⁵ [33], which is a multimodal fake news detection dataset. FakeNewsNet contains articles that were collected from two different fact-checking websites: PolitiFact⁶ and GossipCop⁷ that share news contents annotated as fake or not by professional journalists and experts. In both PolitiFact and GossipCop datasets, each news entity contains news content, including text, images, users' retweets/replies and spatio-temporal information. To create our data collection, we used the tweet ids provided as part of the FakeNewsNet and the Twitter API to collect the tweets (text and image) that were available. To remove duplicate images due to news retweets, we applied a hashing method.⁸ In total, we collected 1480 and 2731 images that contained fake and real posts, respectively, in PolitiFact. For GossipCop, we collected 12,696 and 12,628 images contained fake and true posts, respectively, as shown in Table 2.

For our experiments on fake news detection, we also used FakeNewsNet's collection that we split into training, validation and test sets using the ratio of 8:1:1. Then, we used the tweet ids provided as part of FakeNewsNet and the Twitter API to collect the tweets (text and image) that were available. In total, we managed to collect 7999 PolitiFact and 21,542 GossipCop tweets posts. Table 3 shows the statistics regarding the final collection and its labels.

Table 2. Number of tweets with images.

	Fake	True
PolitiFact	1480	2731
GossipCop	12,696	12,628

Table 3. Statistics of the collections and the labels.

	PolitiFact		GossipCop	
	Fake	Real	Fake	Real
Training set	3827	2828	16,953	16,810
Validation set	345	231	2011	2006
Testing set	440	328	2130	1632
Total	4612	3387	21,094	20,448

Table 4. Experimental parameter settings after optimisation.

Parameter	Value
Epoch	10
Dropout	0.4
Batch size	32
Activation function	relu
Learning rate	0.0001
Optimiser	adam

For the text modality, we tokenised the English tweets using the tokeniser from Tensorflow⁹ and we used the emoji package to translate emotion icons into text strings. We also normalised the tweets by converting digit, user mentions and web/url links into special tokens DIGIT, USER and URL, respectively. All the sentences that exceeded the decided length are trimmed, whereas those below the decided length are padded with zeroes. The final length value is 64. For the image component of the model, all the images are resized to $224 \times 224 \times 3$.

5.2. Experimental settings

For our experiments we used Keras with Tensorflow backend.¹⁰ We initialised our embedding layer with the pre-trained GoogleNews-vectors-negative300¹¹ words and phrase vectors. Table 4 shows the optimised parameters that we used for the neural network. We have experimented with other hyper parameters, such as different hidden layer numbers, hidden units, learning rates and dropout. The dropout is applied to each layer. We used the same parameters for both PolitiFact and GossipCop datasets.

5.3. Evaluation

To evaluate the SceneFND model, we report accuracy, macro precision, macro recall and macro *F1*-score. In addition, we compare the performance of the proposed model with several baselines. Our baselines refer to single modality models and previously proposed multimodal models. We also run the experiments using variants of the SceneFND model.

1. *Text*. This is a single modality model that is based on textual representations learned with Bi-LSTM.
2. *Image*. The second single modality model is based on image feature learned with ResNet50 pre-trained on ImageNet dataset.

Table 5. Experimental results on fake new detection on PolitiFact.

Model	Acc.	P	R	F1
Text	0.8047	0.8068	0.8132	0.8040
Image	0.6133	0.6705	0.6458	0.6067
Text + image	0.8229	0.8446	0.8424	0.8229
Text + image + place	0.8190	0.8476	0.8409	0.8190
Text + image + season	0.8164	0.8427	0.8374	0.8163
Text + image + weather	0.8268	0.8482	0.8461	0.8268
SceneFND	0.8320	0.8479	0.8491	0.8320
SpotFake	0.8074	0.8263	0.8073	0.7965

The bold font indicates best result obtained.

Table 6. Experimental results on fake new detection on GossipCop.

Model	Acc.	P	R	F1
Text	0.7241	0.7249	0.7289	0.7231
Image	0.6813	0.6843	0.6873	0.6806
Text + image	0.7339	0.7317	0.7355	0.7320
Text + image + place	0.7547	0.7566	0.7610	0.7540
Text + image + season	0.7249	0.7371	0.7374	0.7249
Text + image + weather	0.7302	0.7431	0.7432	0.7302
SceneFND	0.7480	0.7487	0.7531	0.7471
SpotFake	0.6991	0.7334	0.6991	0.6978

The bold font indicates best result obtained.

3. *Text + image*. This is multimodal model that is based on text and image features learned with Bi-LSTM and ResNet50.
4. *SpotFake* [23]. This is variant of the proposed model that is based on text and image features learned with BERT and VGG-19 pre-trained on ImageNet dataset, respectively.

6. Results and discussion

In this section, we present the experimental results of the SceneFND and baselines. First, we present the quantitative results and then the qualitative analysis.

6.1. Quantitative results

To address RQ2 (*Can we use scene cues to differentiate between fake news and real news?*), we evaluate and compare the performance of the SceneFND and the baselines. Table 5 shows the results of our proposed model and of the baselines on PolitiFact. From the results, we observe that SceneFND achieves the highest performance with regards to *F1*-score. In particular, our model outperforms the text and image baselines by 3.42% and 31.32%, respectively. In addition, our model manages to obtain a higher score compared with SpotFake over which the improvement is 4.35%.

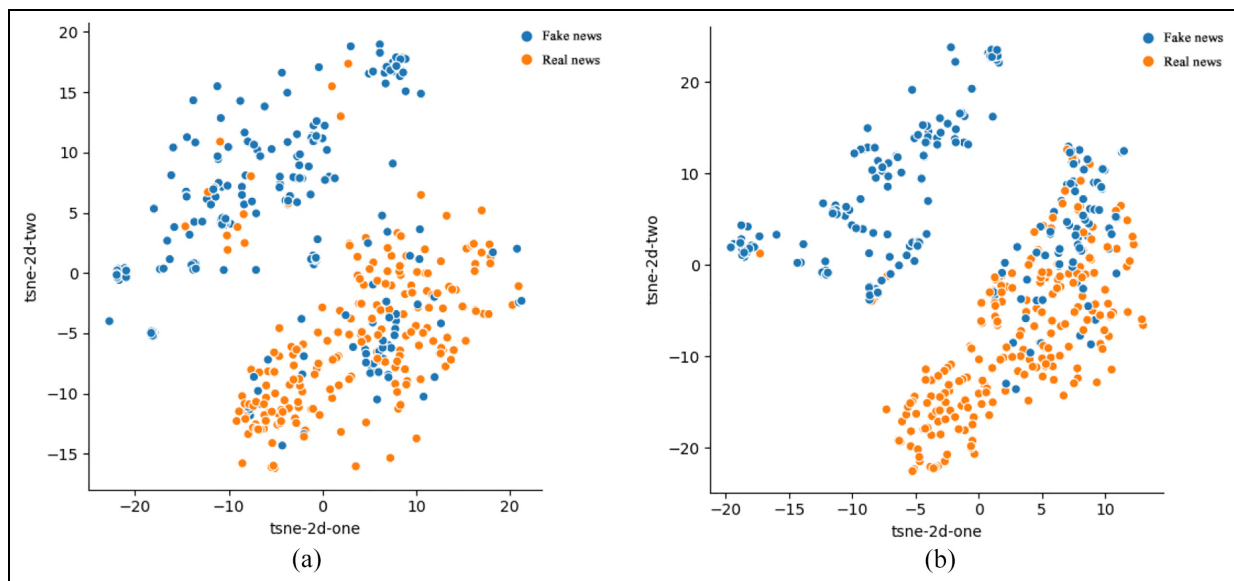
Table 6 shows the results of the SceneFND model and of the baselines on GossipCop. In this case, we observe that the best results are obtained for the SceneFND variant *text + image + place*. On GossipCop, SceneFND model outperforms SpotFake by 7.74%. Regarding text and image baselines, SceneFND manages improvements of 4.18% and 10.23%, respectively.

To address the third research question (*Which of the scenes are statistically significant different between fake and real news?*), we run the Mann–Whitney’s U test on the frequency of the visual scenes as those appear in fake and real news in PolitiFact and GossipCop. Table 7 presents the results of this analysis. The results show that for most of the visual scenes the differences are statistically significant. This result supports our claim that visual scenes follow different frequency patterns in fake and real news. However, there are few visual scenes for which the differences are not statistically significant. Regarding PolitiFact, some of the weather scenes do not show statistical differences, including rain showers, rain, fog and snow. Finally, regarding GossipCop, the only scenes that have no statistical differences between fake and real news are autumn, spring, winter and rain.

Table 7. Statistical significant differences between fake and real news on GossipCop and PolitiFact datasets using Mann–Whitney’s U test.

Scene	PolitiFact	GossipCop	Scene	PolitiFact	GossipCop
Bedroom	**	**	Autumn	**	–
Suburb	**	**	Summer	–	**
Industrial	**	**	Spring	**	–
Kitchen	**	**	Winter	**	–
Living room	**	**	Mostly cloudy	**	**
Coast	**	**	Clear	**	**
Forest	**	**	Mainly clear	**	**
Inside city	**	**	Cloudy	**	**
Highway	**	**	Rain showers	–	**
Mountain	**	**	Drizzle	–	**
Open country	**	**	Moderate rain	**	**
Street	**	**	Rain	–	–
Tall building	**	**	Fog	–	**
Office	**	**	Rain fog	–	**
Store	**	**	Snow	–	**

** and * refer to statistically significant differences of < 0.01 and < 0.05 , respectively, whereas – means that the difference is not statistically significant.

**Figure 5.** Visualisations of learned latent news feature representations on the test set of PolitiFact (a) visualisation of text + image model and (b) visualisation of SceneFND model.

6.2. Qualitative analysis

To further analyse the effectiveness of the proposed model, we qualitatively visualise the news features learned by SceneFND and in particular by the *text + image* variant for PolitiFact and *text + image + place* for GossipCop test sets with t-SNE [34]. The visualisations of the learned features are depicted in Figures 5 and 6 for PolitiFact and GossipCop, respectively. For PolitiFact, we observe that the feature representations learned by the proposed model SceneFND are more discriminable, and especially in fake news posts. We obtain similar observations for GossipCop where the SceneFND variant (*text + image + place*) learned feature representations that are more discriminable

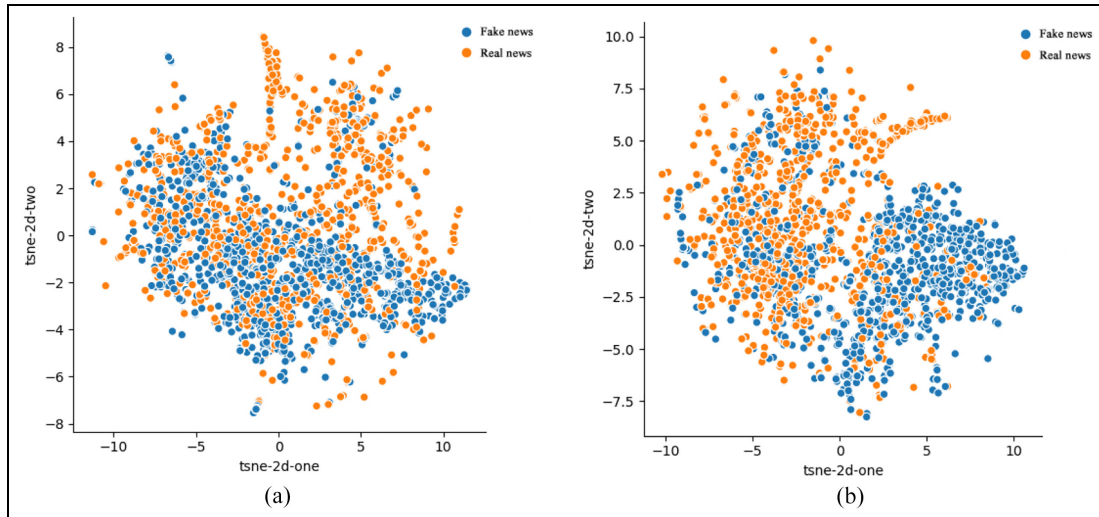


Figure 6. Visualisations of learned latent news feature representations on the test set of GossipCop (a) visualisation of text + image model and (b) visualisation of SceneFND model.



Figure 7. Some fake news detected by SceneFND but missed by text + image model on the PolitiFact dataset. (a) Real news and (b) fake news. (a) Maryk = land League of Conservation Voters 2018 Legislative Wrap Up (place scene: office; season scene: autumn; weather scene: rain showers) and (b) United Airlines Flight Attendant Slaps Crying Baby During Flight (place scene: industrial; season scene: winter; weather scene: clear).



Figure 8. Fake news detected by SceneFND variant of *text + image + place* but misclassified by *text + image* model on the GossipCop dataset. (a) Real news and (b) fake news. (a) Emma Roberts' Lace Top Is the Statement Piece Every Wardrobe Needs (place scene: inside city; season scene: spring; weather scene: rain showers) and (b) @realDonaldTrump Paging All Adult Day Care Staff Someone needs a diaper change and his medication (place scene: forest; season scene: spring; weather scene: rain showers).

among samples with different labels. The comparison between *text + image* and SceneFND proves that the integration of the scene representation component leads to better feature representations and better performance.

Finally, we manually explore samples that were misclassified by the *text + image* baseline but were correctly detected by SceneFND to illustrate the importance of considering the visual scene representation for fake news detection. Figures 7 and 8 show two samples that were correctly classified by SceneFND and misclassified by *text + image* baseline from PolitiFact and GossipCop, respectively.

From the figures, we observe that the text and image contents do not show clear evidence regarding their factuality. Also we observe that the scenes, and especially the place scenes, extracted from real news image are semantically closer to text content compared with the fake news. This shows that the proposed model SceneFND obtains additional clues from the scene information to make an accurate classification. From the images, we can say that the scenes reflect the correlation between images and text. In fake news, scene information helps to capture the semantic incoherence between images and text, whereas in real news, some of the scene information captures the semantic consistency of images and text.

7. Conclusion

In this article, we have proposed SceneFND, a multimodal model for the task of fake news detection. Our model consists of three main components. The first component refers to the textual representation that is obtained using Bi-LSTM. The second component refers to the visual representation obtained using ResNet-50 network. More importantly, we also represent the visual scenes of the images and incorporate this information as the third component of our model. In particular, we extracted and integrated places, weather and season scenes. Although in GossipCop, most of the scenes have similar presence probability for fake and real news, we observed that in PolitiFact the office scenes appears with a higher

probability in real news compared to fake news. Moreover, store, rain showers and autumn is a combination that occurs a lot in fake news in both datasets (RQ1).

We evaluate our proposed model on two datasets and compare its performance with several baselines. Our results showed that the visual scenes that are present in the images can be useful for fake news detection (RQ2). In particular, SceneFND improved the performance of the textual baseline by 3.48% in PolitiFact and by 3.32% in GossipCop. Finally, our analysis showed that the most of the visual scenes have statistical significant different frequencies in fake and real news (RQ3).

As a future work, we plan to carry out a deeper quantitative and qualitative study of the semantic correlations between text and image to precisely identify the type of correlations that really reflect the fake nature of the news. Finally, we plan to compare different type of semantic correlations to reveal the relationship of image and text in event descriptions.


Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: The work of Anastasia Giachanou is funded by the Dutch Research Council (grant VI.Vidi.195.152). The work of Paolo Rosso was in the framework of the Iberian Digital Media Research and Fact-Checking Hub (IBERIFIER) funded by the European Digital Media Observatory (2020-EU-IA0252), and of the XAI-DisInfodemics research project on eXplainable AI for disinformation and conspiracy detection during infodemics, funded by the Spanish Ministry of Science and Innovation (PLEC2021-007681).

ORCID iD

Guobiao Zhang  <https://orcid.org/0000-0002-1568-2492>

Notes

1. <https://www.theguardian.com/world/2017/nov/14/how-400-russia-run-fake-accounts-posted-bogus-brexit-tweets>.
2. <https://github.com/flytxtds/scene-recognition>.
3. <https://github.com/shanerbo/PredictWeatherInImage>.
4. <https://github.com/drewdru/seasonsNNNetwork>.
5. <https://github.com/KaiDMML/FakeNewsNet>.
6. <https://www.politifact.com/>.
7. <https://www.gossipcop.com/>.
8. <https://www.pyimagesearch.com/2020/04/20/detect-and-remove-duplicate-images-from-a-dataset-for-deep-learning/>.
9. https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/text/Tokenizer.
10. <https://tensorflow.google.cn/guide/keras>.
11. <https://github.com/mmihaltz/word2vec-GoogleNews-vectors>.

References

- [1] Zhang C, Fan C, Yao W et al. Social media for intelligent public information and warning in disasters: an interdisciplinary review. *Int J Inform Manage* 2019; 49: 190–207.
- [2] Miller M, Banerjee T, Muppalla R et al. What are people tweeting about Zika? An exploratory study concerning its symptoms, treatment, transmission, and prevention. *JMIR Public Health Surveill* 2017; 3(2): e38.
- [3] Basile V, Cauteruccio F and Terracina G. How dramatic events can affect emotionality in social posting: the impact of covid-19 on reddit. *Future Internet* 2021; 13(2): 29.
- [4] Castillo C, Mendoza M and Poblete B. Information credibility on Twitter. In: *Proceedings of the 20th international conference on World Wide Web (WWW'11)*, Hyderabad, India, 28 March–1 April 2011, pp. 675–684. New York: ACM.
- [5] Shu K, Wang S and Liu H. Understanding user profiles on social media for fake news detection. In: *Proceedings of the 2018 IEEE conference on multimedia information processing and retrieval (MIPR'18)*, Miami, FL, 10–12 April 2018, pp. 430–435. New York: IEEE.
- [6] Giachanou A, Rissola EA, Ghanem B et al. The role of personality and linguistic patterns in discriminating between fake news spreaders and fact checkers. In: E Métais, F Meziane, H Horacek et al. (eds) *Natural language processing and information systems (NLDB'2020)*. Cham: Springer, 2020, pp. 181–192.
- [7] Giachanou A, Ghanem B, Rissola EA et al. The impact of psycholinguistic patterns in discriminating between fake news spreaders and fact checkers. *Data Knowl Eng* 2022; 138: 101960.

- [8] Popat K, Mukherjee S, Yates A et al. DeClarE: debunking fake news and false claims using evidence-aware deep learning. In: *Proceedings of the 2018 conference on empirical methods in natural language processing (EMNLP'18)*, Brussels, 31 October–4 November 2018, pp. 22–32. Stroudsburg: Association for Computational Linguistics (ACL).
- [9] Jain V, Kaliyar RK, Goswami A et al. AENeT: an attention-enabled neural architecture for fake news detection using contextual features. *Neural Comput Appl* 2022; 34: 771–782.
- [10] Wang Y, Ma F, Jin Z et al. 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 (KDD'18)*, London, 19–23 August 2018, pp. 849–857. New York: ACM.
- [11] Khattar D, Goud JS, Gupta M et al. MVAE: multimodal variational autoencoder for fake news detection. In: *Proceedings of the World Wide Web conference (WWW'19)*, San Francisco, CA, 13–17 May 2019, pp. 2915–2921. New York: Association for Computing Machinery (ACM).
- [12] Giachanou A, Zhang G and Rosso P. Multimodal multi-image fake news detection. In: *Proceedings of the 2020 IEEE 7th international conference on data science and advanced analytics (DSAA)*, Sydney, NSW, Australia, 6–9 October 2020, pp. 647–654. New York: IEEE.
- [13] Gombrich EH. *The image and the eye: further studies in the psychology of pictorial representation*. Oxford: Phaidon Press, 1982.
- [14] Ruffo G, Semeraro A, Giachanou A et al. Surveying the research on fake news in social media: a tale of networks and language, 2021, <https://arxiv.org/abs/2109.07909?context=cs.SI>
- [15] Shu K, Sliva A, Wang S et al. Fake news detection on social media: a data mining perspective. *ACM SIGKDD Explor News* 2017; 19(1): 22–36.
- [16] Zhou X and Zafarani R. A survey of fake news: fundamental theories, detection methods, and opportunities. *ACM Comput Surv* 2021; 53(5): 109.
- [17] Zubiaga A, Aker A, Bontcheva K et al. Detection and resolution of rumours in social media: a survey. *ACM Comput Surv* 2019; 51(2): 32.
- [18] Vosoughi S, Roy D and Aral S. The spread of true and false news online. *Science* 2018; 359(6380): 1146–1151.
- [19] Pérez-Rosas V, Mihalcea R. and Experiments in open domain deception detection. In: *Proceedings of the 2015 conference on empirical methods in natural language processing*, Lisbon, 17–21 September 2015, pp. 1120–1125. Stroudsburg, PA: Association for Computational Linguistics (ACL).
- [20] Giachanou A, Rosso P and Crestani F. Leveraging emotional signals for credibility detection. In: *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval (SIGIR'19)*, Paris, 21–25 July 2019, pp. 877–880. New York: ACM.
- [21] Guo C, Cao J, Zhang X et al. Exploiting emotions for fake news detection on social media, 2019, https://www.researchgate.net/publication/331543936_Exploiting_Emotions_for_Fake_News_Detection_on_Social_Media
- [22] Giachanou A, Rosso P and Crestani F. The impact of emotional signals on credibility assessment. *J Assoc Inf Sci Tech* 2021; 72(9): 1117–1132.
- [23] Singhal S, Shah RR, Chakraborty T et al. SpotFake: a multi-modal framework for fake news detection. In: *Proceedings of the 2019 IEEE 5th international conference on multimedia big data (BigMM)*, Singapore, 11–13 September 2019, pp. 39–47. New York: IEEE.
- [24] Zlatkova D, Nakov P and Koychev I. Fact-checking meets fauxtography: verifying claims about images. In: *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP'19)*, Hong Kong, China, 3–7 November 2019, pp. 2099–2108. Stroudsburg: Association for Computational Linguistics (ACL).
- [25] Giachanou A, Zhang G and Rosso P. Multimodal fake news detection with textual, visual and semantic information. In: Sojka P, Kopeček I, Pala K, et al. (eds) *Text, speech, and dialogue*. Cham: Springer, 2020, pp. 30–38.
- [26] Zhou X, Wu J and Zafarani R. SAFE: similarity-aware multi-modal fake news detection. In: H Lauw, RW Wong, A Ntoulas et al. (eds) *Advances in knowledge discovery and data mining*, vol. 12085. Cham: Springer, 2020, pp. 354–367.
- [27] Henderson JM and Hollingworth A. High-level scene perception. *Annu Rev Psychol* 1999; 50(1): 243–271.
- [28] Xu L and Wang X. Semantic description of cultural digital images: using a hierarchical model and controlled vocabulary. *D-Lib Magazine*, 2015, vol. 21, issue no. 5–6, pp. 5–11.
- [29] Wilson J and Arif M. Scene recognition by combining local and global image descriptors, 2017, <https://arxiv.org/abs/1702.06850>
- [30] Tola E, Lepetit V and Fua P. A fast local descriptor for dense matching. In: *Proceedings of the 2008 IEEE conference on computer vision and pattern recognition (CVPR'08)*, Anchorage, AK, 23–28 June 2008, pp. 430–438. New York: IEEE.
- [31] Hernández Fariás DI, Patti V and Rosso P. Irony detection in Twitter: the role of affective content. *ACM T Internet Techn* 2016; 16(3): 19.
- [32] Jang B, Kim M, Harerimana G et al. Bi-LSTM model to increase accuracy in text classification: combining word2vec CNN and attention mechanism. *Appl Sci* 2020; 10(17): 5841.
- [33] Shu K, Mahudeswaran D, Wang S et al. FakeNewsNet: a data repository with news content, social context and spatialtemporal information for studying fake news on social media, 2018, <https://arxiv.org/abs/1809.01286>
- [34] Van der Maaten L and Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008; 9(17): 2579–2605.