Fake News Detection with Multimodal Machine Learning

Team Member: Danni Ma, Kaijun Feng, Chuanqi Chen

Mentor: Akshay Smiti

Problem summary

Fake news has been a serious social problem with huge negative influence on both politics and culture. We use multimodal methods to detect fake news based on image and text data, which improves the performance compared to that of a unimodal approach.

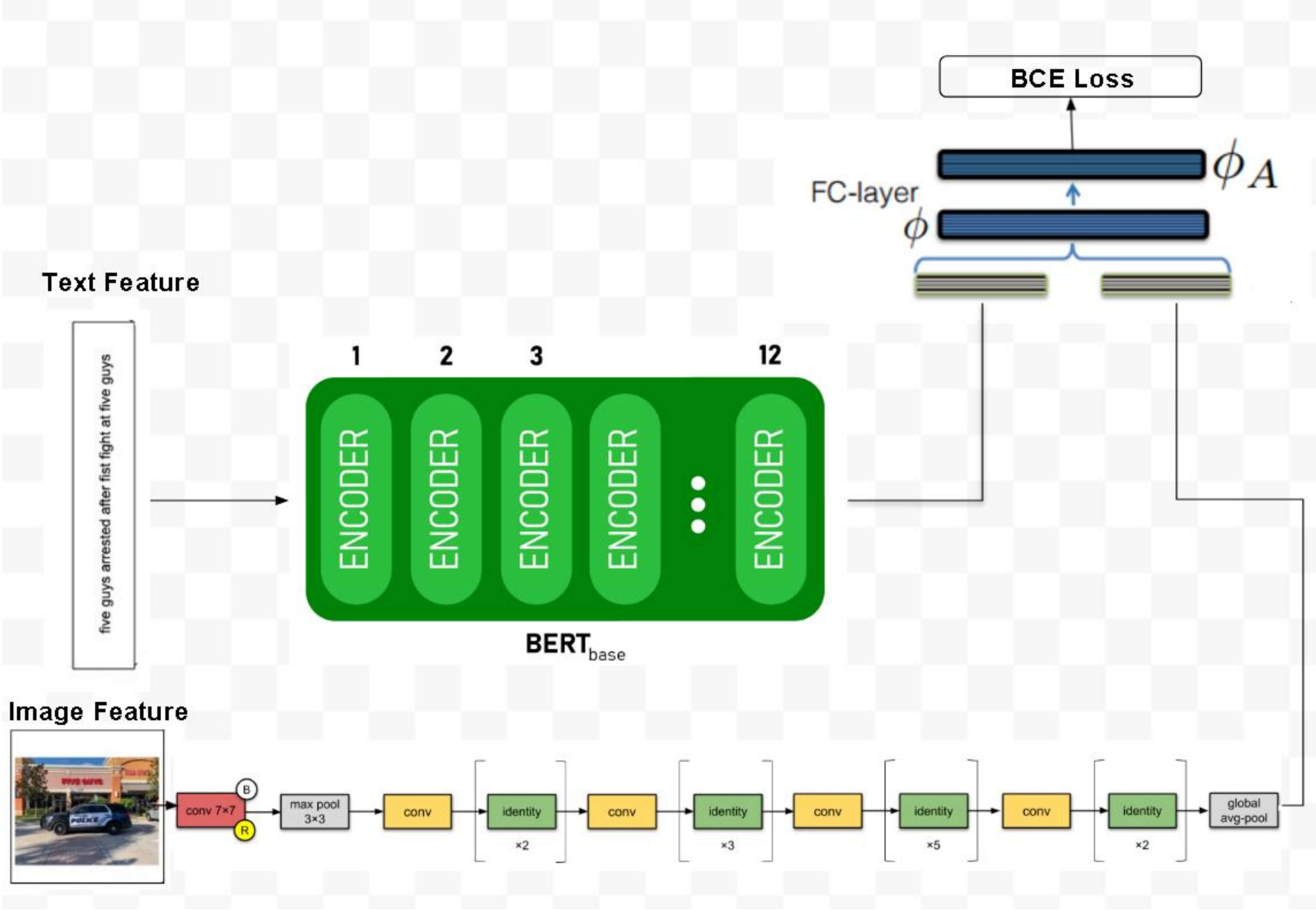
Dataset

In this project, we uses a new dataset called Fakeddit, introduced by Nakamura et al. (2020). Fakeddit consists of over 1 million samples from multiple categories of fake news and over 60% of which are multimodal with both images and text data. The authors originally demonstrated multimodal detection of up to 89% accuracy.

Approach

Similar to the approach of Nakamura et al, we adopt a late fusion architecture where text and image features are concatenated and passed through a fully connected layer with 1 output feature for 2-way classification. Here our main innovations are:

- Fully retrain the language and vision models for feature extraction
- Single output with BCE logit loss for efficiency and robustness
- Concatenate full language and vision features for completeness



Results

Method (NLP Task)	Parameters	Metrics Validation	 mcc
bert-base-uncased[1]	110M	0.89	0.767
bert-large-uncased	340M	0.501	
distilbert-base-uncased	66M	0.50	
gpt2-xl	1558M	0.73	0.482
Transformer-XL	257M	0.80	0.599

[1] Best NLP model to detect fake news.

Method (CNN Task)	Parameters	Metrics training	 Validation
resnet50 (Fully Retrain) [1]	23M	0.8475	0.8418
resnet50 (Fine Tune)	23M	0.6846	0.7058
resnet18 (Fine Tune)	11M	0.6434	0.6637
VGG16 (Fine Tune)	138M	0.6464	0.6590

[1] Best CNN model to detect fake images.

Method (Multimodal Task)	training	validation	Metrics test	mcc	F1
Bert-base-uncased + ResNet	0.9403	0.9206	0.9217	0.8374	0.9003

Conclusion

In conclusion, we have successfully implemented multiple NLP and CNN models to detect fake news, and fake images. We then combined two best performing models BERT base and ResNet50 for multimodal fake news detection with a late fusion architecture. We achieved state of the art performance with 0.9217 test accuracy and 0.9003 F1 score which is superior to the previous published work on the same dataset.

References

- Nakamura, K.; Levy, S.; Wang, W. Y. r/Fakeddit: A New Multimodal Benchmark Dataset for Fine-grained Fake News Detection. https://github.com/entitize/Fakeddit.
- Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Trans-formers for Language Understanding. 2019.
- Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. International Conference on Learning Representations. 2015.