SAFE: Similarity-Aware Multi-Modal Fake News Detection

Xinyi Zhou*, Jindi Wu*, and Reza Zafarani 雪城大學

Data Lab, EECS Department, Syracuse University, NY 13244, U.S.A.

{zhouxinyi,reza}@data.syr.edu, jwu172@syr.edu

PAKDD 2020 (Pacific-Asia Conference on Knowledge Discovery and Data Mining)

211209 2021 Data Mining Project Report (Group 5)

Outline
of Paper Explanation

Introduction

Related Work

Methodology

Experiments

Conclusions

Comments

Q & A

Introduction

Fake News Detection

- As "a news article that is intentionally and verifiably false", fake news content often contains textual and visual information.
- Existing content-based fake news detection methods either solely consider textual information or combine both types of data ignoring the relationship (similarity).
- The values in understanding such relationship (similarity) for predicting fake news are two-fold.

Introduction

Relationship (similarity) for predicting fake news

- To attract public attention, some fake news stories (or news stories with low-credibility) prefer to use dramatic, humorous (facetious), and tempting images whose content is **far from the actual content within the news text**.
- When a fake news article tells a story with fictional scenarios or statement, it's difficult to find both pertinent and non-manipulated images to match these fictions
  - Hence a "gap" exists between the textual and visual information of fake news when creators use non-manipulated images to support non-factual scenarios or statements.

Introduction

Mis-captioned Example

Snopes

Search Snopes.com

Become a Member   Submit a Topic   Shop   Latest   Top   Fact Checks   Collect

Home › Fact Checks › Miscaptioned

Miscaptioned

This rating is used with photographs and videos that are “real” (i.e., not the product, partially or wholly, of digital manipulation) but are nonetheless misleading because they are accompanied by explanatory material that falsely describes their origin, context, and/or meaning.

Learn more about our rating system here.

Examples at https://www.snopes.com/fact-check/rating/miscaptioned/.

Is Viral Heart-Shaped Sunset Photo Real?

9 June 2021

This is a realistic work from an artist known for digitally edited images.

Is This Photo of a Baby Albino Bat Real?

27 May 2021

It's time for another round of "Toy or Animal?"

No, This Is Not a Photo of Biden’s ATF Nominee David Chipman at Waco

26 May 2021

A photograph from the deadly siege on the Branch Davidian compound has been shared with...

Did Kit Kat Make ‘No Straight Lines’ Bar for Pride Month?

23 May 2021

A diagonal line is a straight line set on an angle.

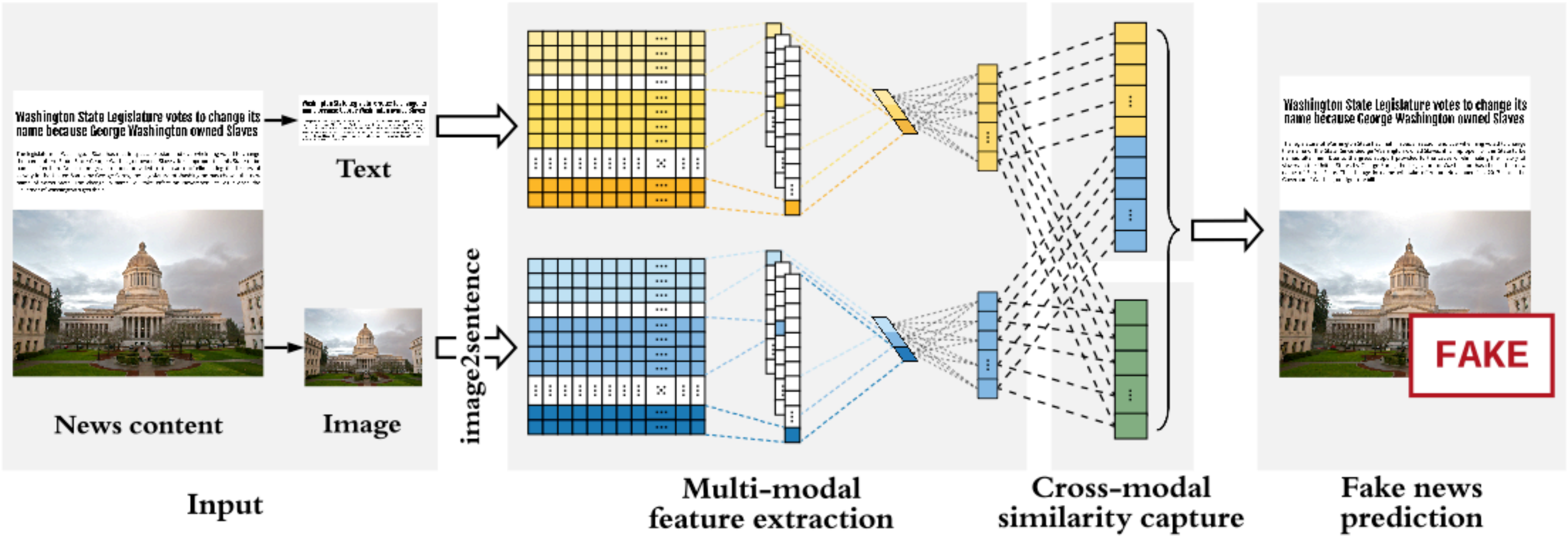
Introduction

Similarity-Aware FakE news detection method (SAFE)

- SAFE consists of three modules:
  - Multi-modal (textual & visual) feature extraction
  - Within-modal (or say, modal-independent) fake news prediction
  - Cross-modal similarity extraction

Introduction

Similarity-Aware FakE news detection method (SAFE)



Introduction
Contributions

- First approach that investigates the role of the relationship (similarity) between news textual & visual information in predicting fake news.
- Proposed a new method to jointly exploit multi-modal (textual & visual) and relational information to learn the representation of news articles and predict fake news.

Related Works

Content-based Fake News Detection

- Various hand-crafted features (linguistic/writing styles)
  - SVM, Random forest
- Rhetorical structures among sentences or phrases within news content
  - Vector space model, Bi-LSTM
- Some researches explored the political bias and homogeneity of news publishers by mining news content that they have published.

Related Works

Social-context-based Fake News Detection

- Identifying the differences in propagation patterns between fake news and the truth.
- Such contributions have also focused on how user profiles and opinions
    - Using feature engineering and neural networks

Methodology

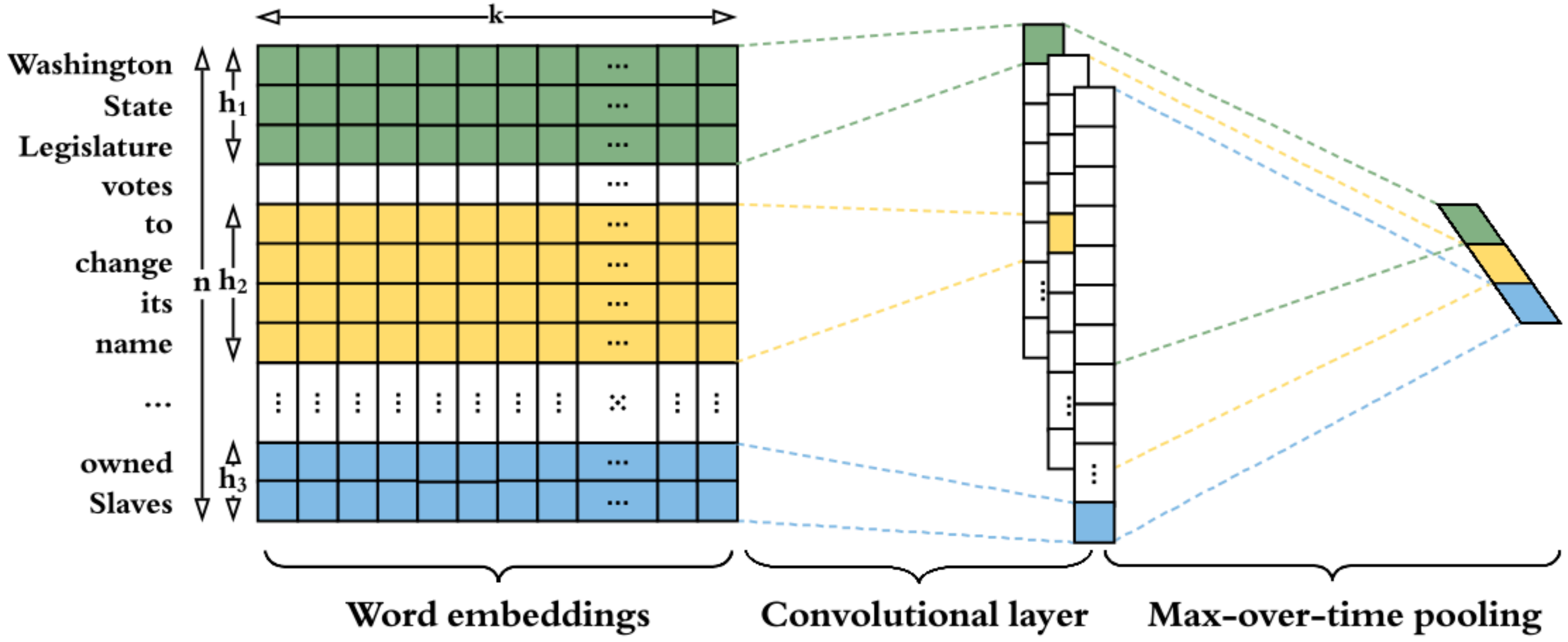
Problem Definition and Key Notation

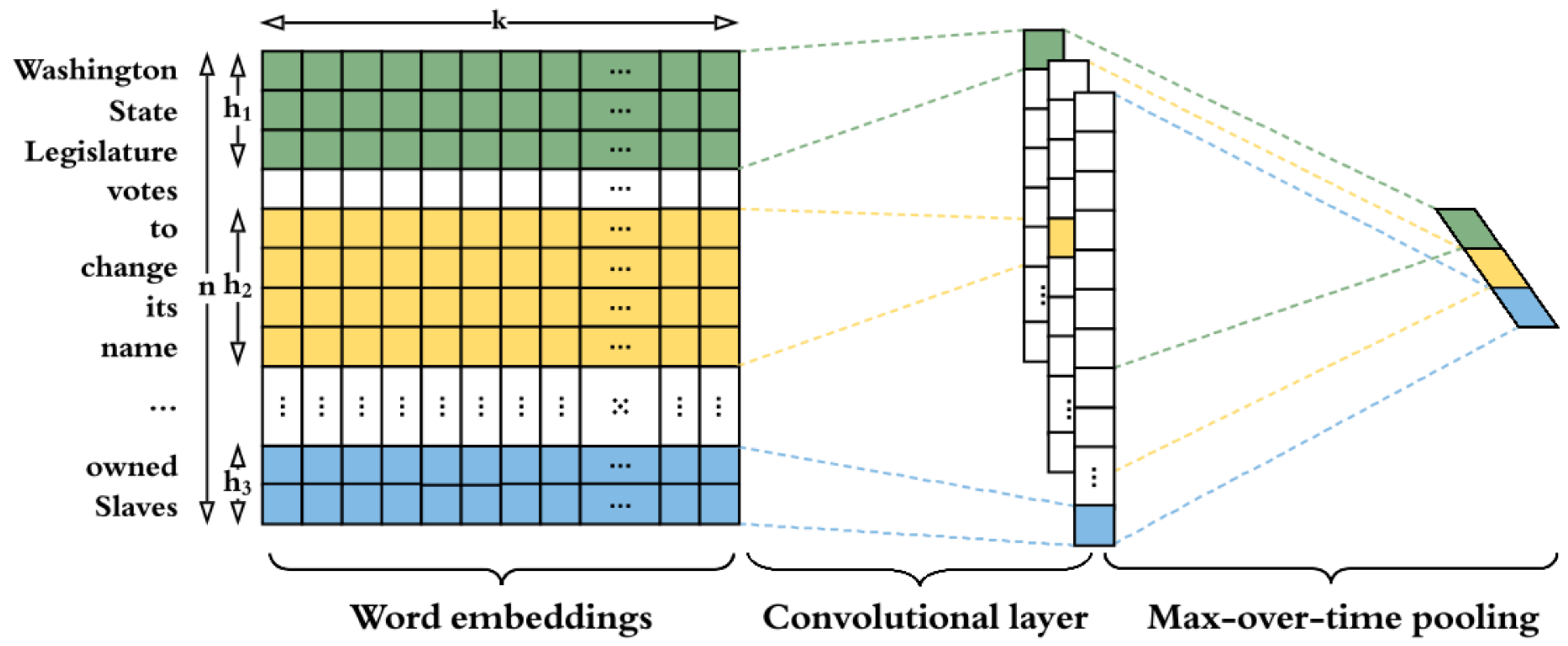
- Given a news article $A = \{T, V\}$ ($T =$ text information, $V =$ visual information)
- Denote $t, v \in \mathbb{R}^d$ as corresponding representations, $t = M_t(T, \theta_t), v = M_v(V, \theta_v)$
- Let $s = M_s(t, v)$ denote the similarity between t and v, where $s \in [0,1]$
- Goal: $M_p : (M_t, M_v, M_s) \xrightarrow{(\theta_t, \theta_v, \theta_p)} \hat{y} \in [0,1]$, where θ_* are parameters to be learned
  - Determine whether A is fake news ($\hat{y} = 1$) or true one ($\hat{y} = 0$).
  - By investigating its textual, visual information, and their relationship.

Methodology

Multi-modal Feature Extraction - Text

- Extend Text-CNN by introducing an additional fully connected layer to automatically extract textual features for each news article.





Methodology

Multi-modal Feature Extraction - Image

- Also use Text-CNN with an additional fully connected layer while first process visual information within news content using a pre-trained **`image2sentence`** model.

a red double decker bus driving down a street.

a man is sitting at a table with a cake.

a man riding a surfboard on a wave.

https://github.com/nikhilmaram/Show_and_Tell

Methodology
Multi-modal Feature Extraction - Image

- Compare to existing multi-modal fake news detection studies that often directly apply a pre-trained CNN model (e.g., VGG) to obtain the representation of news images

- Use image2sentence for consistency and to increase insights when computing the similarity across modalities.

a red double decker bus driving down a street.

Methodology

Modal-independent Fake News Detection

- To properly represent news textual and visual information in predicting fake news, we aim to correctly map the extracted textual and visual features of news content to their possibilities of being fake, and further to their actual labels.

  - Possibilities can be computed by $M_p(t, v) = 1 \cdot \mathrm{softmax}(W_p(t \oplus v) + b_p)$

    - $1 = [1,0]^T$, $W_p \in \mathbb{R}^{2\times 2d}$ and $b_p \in \mathbb{R}^2$ are parameters to be trained.

- Cross-entropy-based loss function:

  - $L_p(\theta_t, \theta_v, \theta_p) = -\mathbb{E}_{(a,y)\sim(A,Y)}(y \log M_p(t, v) + (1-y)\log(1 - M_p(t, v)))$

Methodology

Cross-modal Similarity Extraction

- Most method are considered two different modal features (t, v) separately
  - Just concatenating them with no relation between them explored
- However, besides that, the falsity of a news article can be also detected by assessing how (ir)relevant the textual information is compared to its visual information
- Fake news creators sometimes actively use irrelevant image for false statements to attract readers' attention, or passively use them due to the difficulty in finding a supportive non-manipulated image.

Methodology

Cross-modal Similarity Extraction

- Compared to news articles delivering relevant textual and visual information, those with disparate statements and images are more likely to be fake.

- Define the relevance between news textual and visual information as follows by slightly modifying cosine similarity:

- $M_s(t, v) = \dfrac{t \cdot v + \|t\| \|v\|}{2\|t\| \|v\|}$ v.s. $\cos(t, v) = \dfrac{t \cdot v}{\|t\| \|v\|}$

- In such a way, it's guaranteed that $M_s(t, v)$ is positive and $\in [0,1]$

  - $M_s(t, v) \to 0$: t, v are far from being similar, $\to 1$: t, v are exactly the same

Methodology

Cross-modal Similarity Extraction

- Then defined the loss function based on cross-entropy as below, which assumes that news articles formed with mismatched textual and visual information are more likely to be fake compared to those with matching textual statements and images, when analyzing from a pure similarity perspective:

- $L_S(\theta_t, \theta_v) = -\mathbb{E}_{(a,y)\sim(A,Y)}(y \log(1 - M_s(t, v)) + (1 - y)\log M_s(t, v))$

Methodology

Model Integration and Joint Learning

- When detecting fake new, we aim to correctly recognize fake news stories whose falsity is in their textual and/or visual information, or their relationship.
- Final loss function as
  - $L(\theta_t, \theta_v, \theta_p) = \alpha L_p(\theta_t, \theta_v, \theta_p) + \beta L_s(\theta_t, \theta_v)$
    - $L_p(\theta_t, \theta_v, \theta_p) = -\mathbb{E}_{(a,y)\sim(A,Y)}(y \log M_p(t, v) + (1-y)\log(1-M_p(t, v)))$
    - $L_S(\theta_t, \theta_v) = -\mathbb{E}_{(a,y)\sim(A,Y)}(y \log(1-M_s(t, v)) + (1-y)\log M_s(t, v))$

Experiments
Setup: Dataset

	PolitiFact			GossipCop		
	Fake	**True**	**Overall**	**Fake**	**True**	**Overall**
# News articles	432	624	1,056	5,323	16,817	22,140
– with textual information	420	528	948	4,947	16,694	21,641
– with visual information	336	447	783	1,650	16,767	18,417

https://github.com/KaiDMML/FakeNewsNet

- Experiments are conducted on two well-established public benchmark datasets of fake news detection.
- PolitiFact (politifact.com) (2002.05 ~ 2018.07)
  - non-profit fact-checking website of political statements and reports in the U.S.
- GossipCop (gossipcop.com) (2000.07 ~ 2018.12)
  - fact-checks celebrity reports and entertainment stories published in magazines and newspapers.

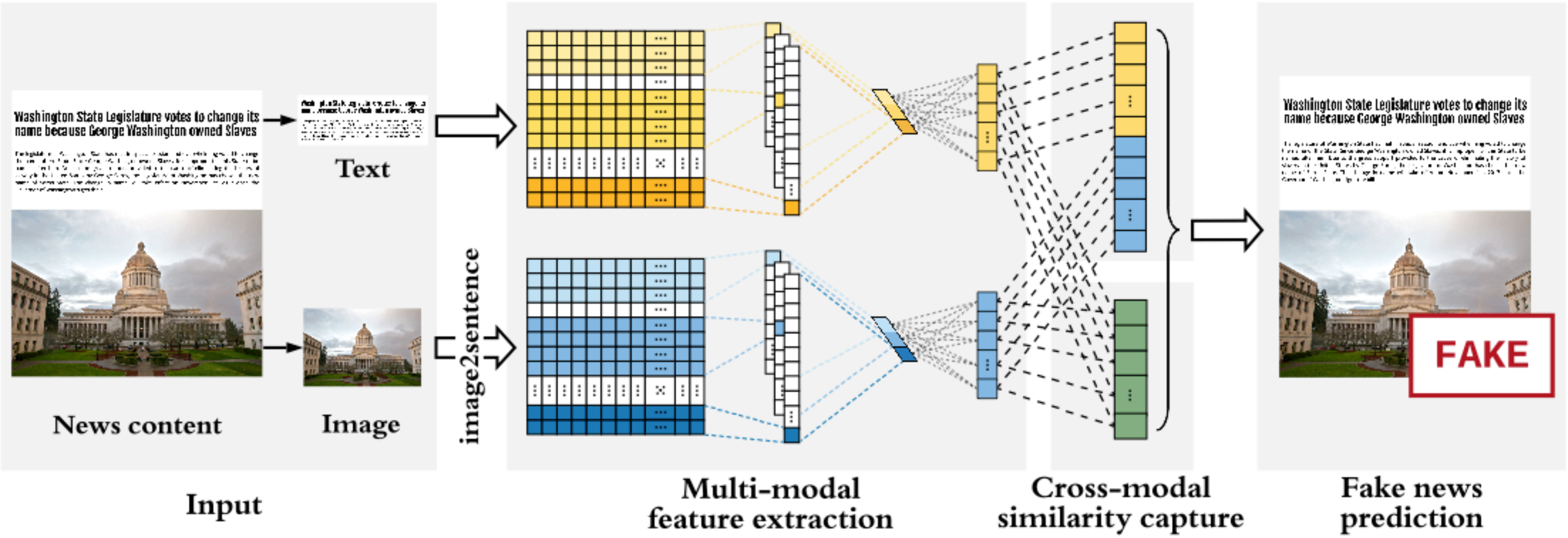
Experiments

Setup: Baselines

- Compare to the following baselines, which detect fake news using
  - (i) textual (LIWC): widely-accepted psycho-linguistics lexicon
  - (ii) visual (VGG-19): use fine-tuned VGG-19 as one of the baselines
  - (iii) multi-modal information (att-RNN):
    - Employ LSTM & VGG-19 with attention mechanism to fuse textual, visual and social-context features of news articles. (exclude social-context feature for fair)

Experiments

Baselines



- Variants of the proposed SAFE method:
- SAFE\T: without using textual information
- SAFE\V: without using visual information
- SAFE\S: without capturing the relationship (similarity) between textual and visual features. In this case, features of each news are fused by concatenating them
- SAFE\W: only assessed the relationship between textual and visual information. In this case, the classifier is directly connected with the output of the cross-modal similarity extraction module.

Experiments

Performance Analysis

		LIWC†	VGG-19≀	att-RNN‡	SAFE\T≀	SAFE\V†	SAFE\S‡	SAFE\W‡	SAFE‡
Politi-Fact	Acc.	0.822	0.649	0.769	0.674	0.721	0.796	0.738	**0.874**
	Pre.	0.785	0.668	0.735	0.680	0.740	0.826	0.752	**0.889**
	Rec.	0.846	0.787	**0.942**	0.873	0.831	0.801	0.844	0.903
	F_1	0.815	0.720	0.826	0.761	0.782	0.813	0.795	**0.896**
Gossip-Cop	Acc.	0.836	0.775	0.743	0.721	0.802	0.814	0.812	**0.838**
	Pre.	**0.878**	0.775	0.788	0.734	0.853	0.875	0.853	0.857
	Rec.	0.317	0.970	0.913	**0.974**	0.883	0.872	0.901	0.937
	F_1	0.466	0.862	0.846	0.837	0.868	0.874	0.876	**0.895**

†: Text-based methods ≀: Image-based methods ‡: Multi-modal methods

- SAFE can outperform all baselines based on the accuracy values and F1 scores for both datasets.

Experiments

Performance Analysis

		LIWC†	VGG-19≀	att-RNN‡	SAFE\T≀	SAFE\V†	SAFE\S‡	SAFE\W‡	SAFE‡
Politi-Fact	Acc.	0.822	0.649	0.769	0.674	0.721	0.796	0.738	**0.874**
	Pre.	0.785	0.668	0.735	0.680	0.740	0.826	0.752	**0.889**
	Rec.	0.846	0.787	**0.942**	0.873	0.831	0.801	0.844	0.903
	F_1	0.815	0.720	0.826	0.761	0.782	0.813	0.795	**0.896**
Gossip-Cop	Acc.	0.836	0.775	0.743	0.721	0.802	0.814	0.812	**0.838**
	Pre.	**0.878**	0.775	0.788	0.734	0.853	0.875	0.853	0.857
	Rec.	0.317	0.970	0.913	**0.974**	0.883	0.872	0.901	0.937
	F_1	0.466	0.862	0.846	0.837	0.868	0.874	0.876	**0.895**

†: Text-based methods ≀: Image-based methods ‡: Multi-modal methods

- Based on PolitiFact data, the general performance of methods is SAFE (multi-modal) > att-RNN (multi-modal) ≈ LIWC (text) > VGG-19 (visual)

Experiments
Performance Analysis

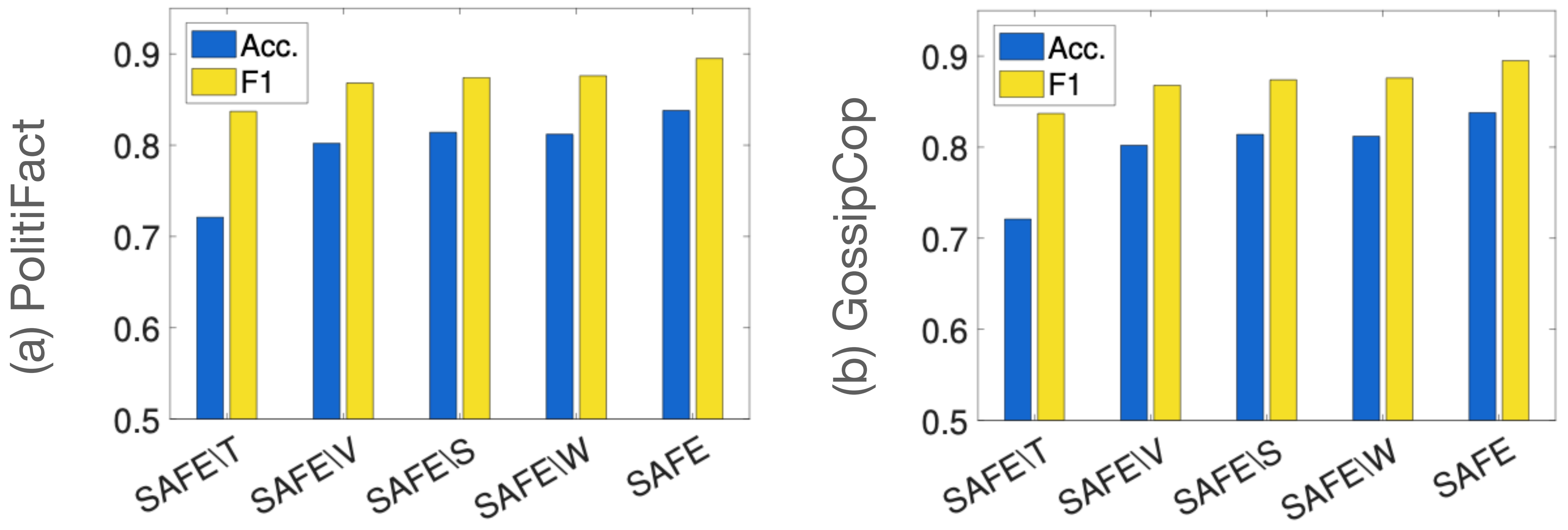
		LIWC†	VGG-19≀	att-RNN‡	SAFE\T≀	SAFE\V†	SAFE\S‡	SAFE\W‡	SAFE‡
PolitiFact	**Acc.**	0.822	0.649	0.769	0.674	0.721	0.796	0.738	**0.874**
	Pre.	0.785	0.668	0.735	0.680	0.740	0.826	0.752	**0.889**
	Rec.	0.846	0.787	**0.942**	0.873	0.831	0.801	0.844	0.903
	F_1	0.815	0.720	0.826	0.761	0.782	0.813	0.795	**0.896**
GossipCop	**Acc.**	0.836	0.775	0.743	0.721	0.802	0.814	0.812	**0.838**
	Pre.	**0.878**	0.775	0.788	0.734	0.853	0.875	0.853	0.857
	Rec.	0.317	0.970	0.913	**0.974**	0.883	0.872	0.901	0.937
	F_1	0.466	0.862	0.846	0.837	0.868	0.874	0.876	**0.895**

†: Text-based methods ≀: Image-based methods ‡: Multi-modal methods

- While for GossipCop data, such performance is
  SAFE (multi-modal) > VGG-19 (visual) > att-RNN (multi-modal) > LIWC (text)

Experiments

Module Analysis



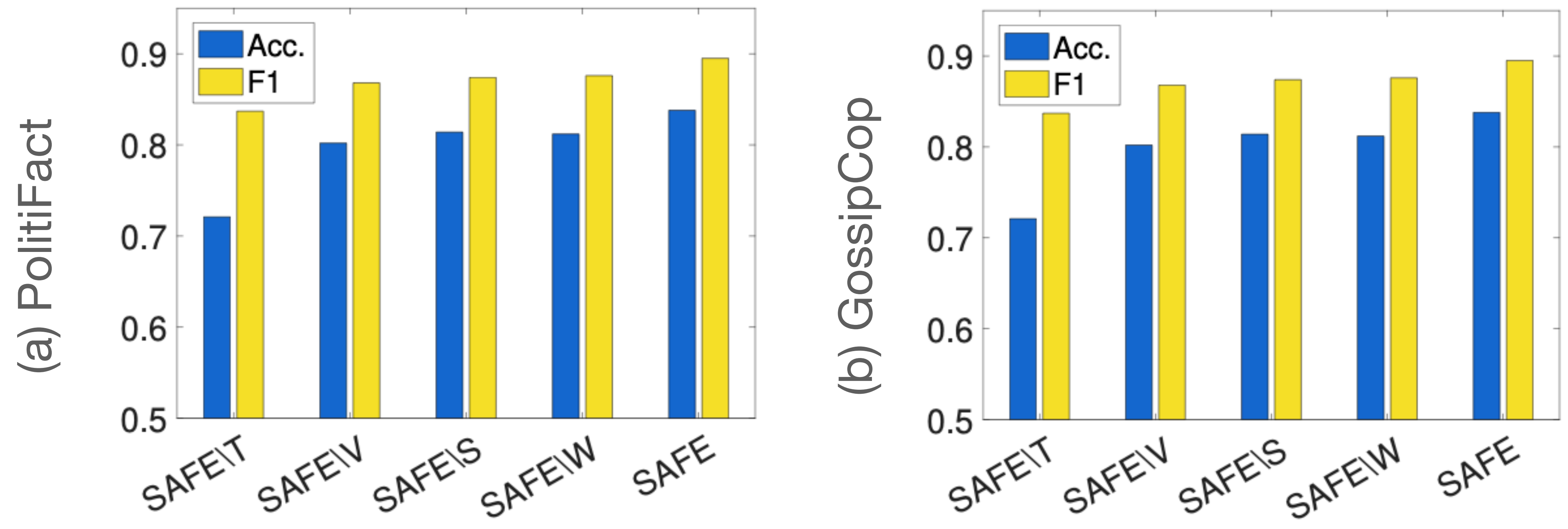
		LIWC†	VGG-19≀	att-RNN‡	SAFE\T≀	SAFE\V†	SAFE\S‡	SAFE\W‡	SAFE‡
Politi-Fact	**Acc.**	0.822	0.649	0.769	0.674	0.721	0.796	0.738	**0.874**
	Pre.	0.785	0.668	0.735	0.680	0.740	0.826	0.752	**0.889**
	Rec.	0.846	0.787	**0.942**	0.873	0.831	0.801	0.844	0.903
	F_1	0.815	0.720	0.826	0.761	0.782	0.813	0.795	**0.896**
Gossip-Cop	**Acc.**	0.836	0.775	0.743	0.721	0.802	0.814	0.812	**0.838**
	Pre.	**0.878**	0.775	0.788	0.734	0.853	0.875	0.853	0.857
	Rec.	0.317	0.970	0.913	**0.974**	0.883	0.872	0.901	0.937
	F_1	0.466	0.862	0.846	0.837	0.868	0.874	0.876	**0.895**

†: Text-based methods ≀: Image-based methods ‡: Multi-modal methods

- (1) integrating news textual information, visual information, and their relationship (SAFE) performs best among all variants,

Experiments

Module Analysis



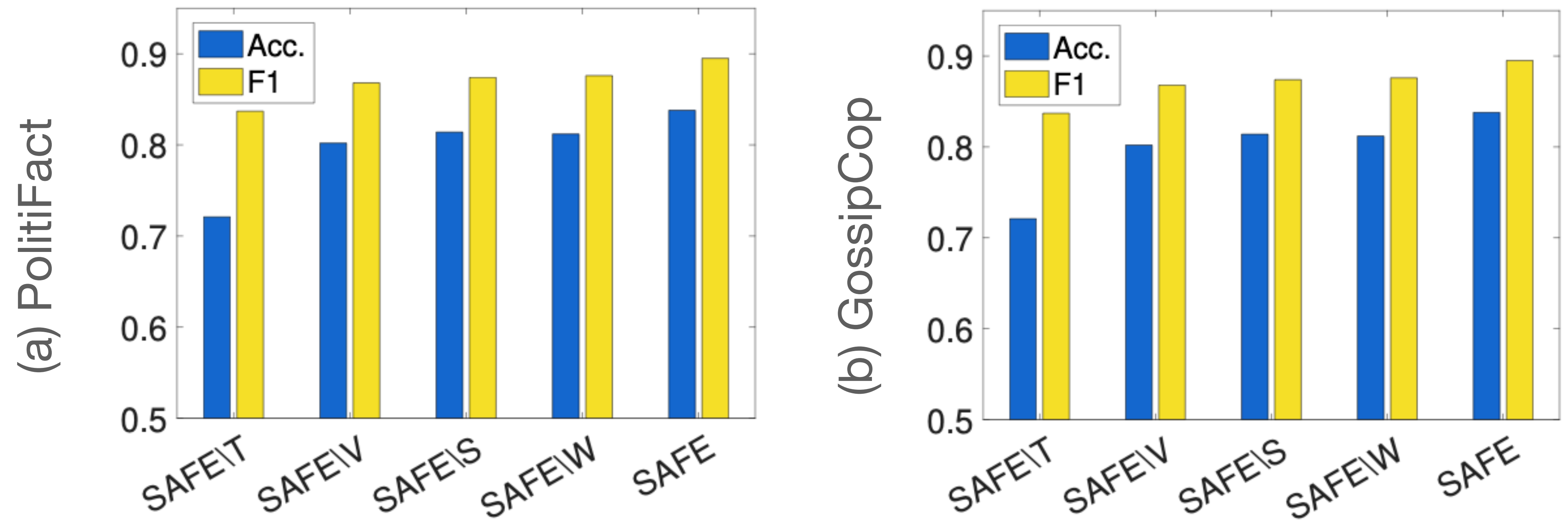
		LIWC†	VGG-19≀	att-RNN‡	SAFE\T≀	SAFE\V†	SAFE\S‡	SAFE\W‡	SAFE‡
Politi-Fact	**Acc.**	0.822	0.649	0.769	0.674	0.721	0.796	0.738	**0.874**
	Pre.	0.785	0.668	0.735	0.680	0.740	0.826	0.752	**0.889**
	Rec.	0.846	0.787	**0.942**	0.873	0.831	0.801	0.844	0.903
	$\mathbf{F_1}$	0.815	0.720	0.826	0.761	0.782	0.813	0.795	**0.896**
Gossip-Cop	**Acc.**	0.836	0.775	0.743	0.721	0.802	0.814	0.812	**0.838**
	Pre.	**0.878**	0.775	0.788	0.734	0.853	0.875	0.853	0.857
	Rec.	0.317	0.970	0.913	**0.974**	0.883	0.872	0.901	0.937
	$\mathbf{F_1}$	0.466	0.862	0.846	0.837	0.868	0.874	0.876	**0.895**

†: Text-based methods ≀: Image-based methods ‡: Multi-modal methods

- (2) using multi-modal information (SAFE\S or SAFE\W) performs better compared to using single-modal information (SAFE\T or SAFE\V)

Experiments

Module Analysis



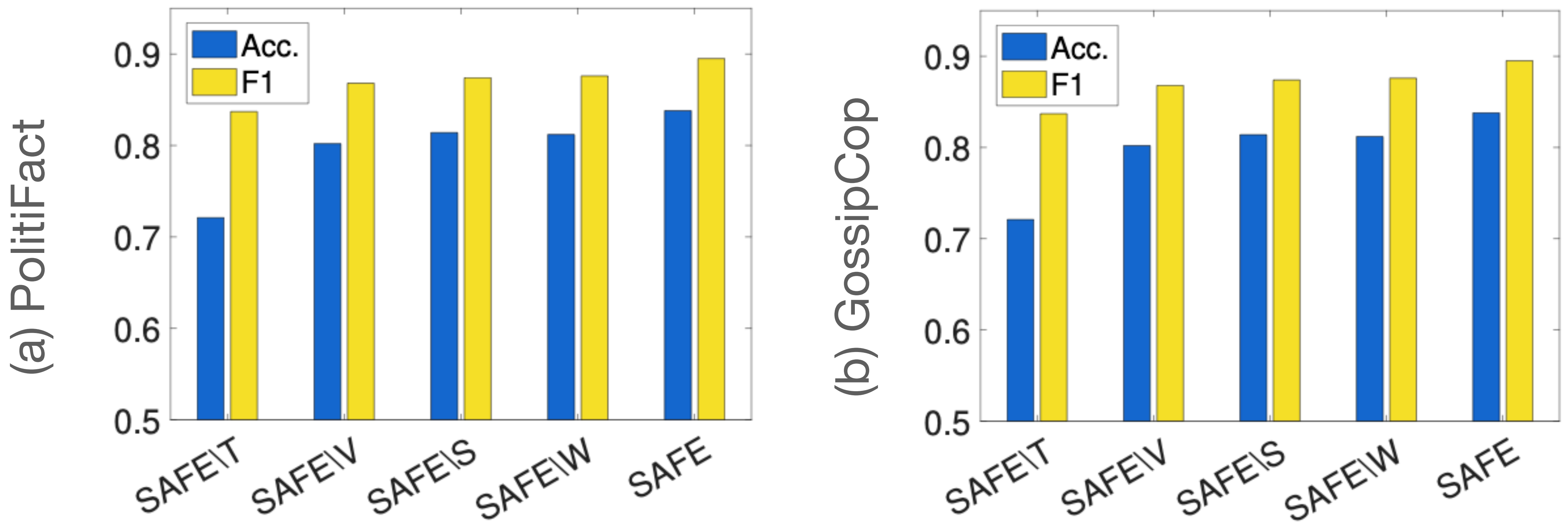
		LIWC†	VGG-19≀	att-RNN‡	SAFE\T≀	SAFE\V†	SAFE\S‡	SAFE\W‡	SAFE‡
Politi-Fact	**Acc.**	0.822	0.649	0.769	0.674	0.721	0.796	0.738	**0.874**
	Pre.	0.785	0.668	0.735	0.680	0.740	0.826	0.752	**0.889**
	Rec.	0.846	0.787	**0.942**	0.873	0.831	0.801	0.844	0.903
	F_1	0.815	0.720	0.826	0.761	0.782	0.813	0.795	**0.896**
Gossip-Cop	**Acc.**	0.836	0.775	0.743	0.721	0.802	0.814	0.812	**0.838**
	Pre.	**0.878**	0.775	0.788	0.734	0.853	0.875	0.853	0.857
	Rec.	0.317	0.970	0.913	**0.974**	0.883	0.872	0.901	0.937
	F_1	0.466	0.862	0.846	0.837	0.868	0.874	0.876	**0.895**

†: Text-based methods ≀: Image-based methods ‡: Multi-modal methods

- (3) it is comparable to detect fake news by either independently using multi-modal information (SAFE\S) or mining their relationship (SAFE\W)

Experiments

Module Analysis



		LIWC†	VGG-19≀	att-RNN‡	SAFE\T≀	SAFE\V†	SAFE\S‡	SAFE\W‡	SAFE‡
Politi-Fact	**Acc.**	0.822	0.649	0.769	0.674	0.721	0.796	0.738	**0.874**
	Pre.	0.785	0.668	0.735	0.680	0.740	0.826	0.752	**0.889**
	Rec.	0.846	0.787	**0.942**	0.873	0.831	0.801	0.844	0.903
	F_1	0.815	0.720	0.826	0.761	0.782	0.813	0.795	**0.896**
Gossip-Cop	**Acc.**	0.836	0.775	0.743	0.721	0.802	0.814	0.812	**0.838**
	Pre.	**0.878**	0.775	0.788	0.734	0.853	0.875	0.853	0.857
	Rec.	0.317	0.970	0.913	**0.974**	0.883	0.872	0.901	0.937
	F_1	0.466	0.862	0.846	0.837	0.868	0.874	0.876	**0.895**

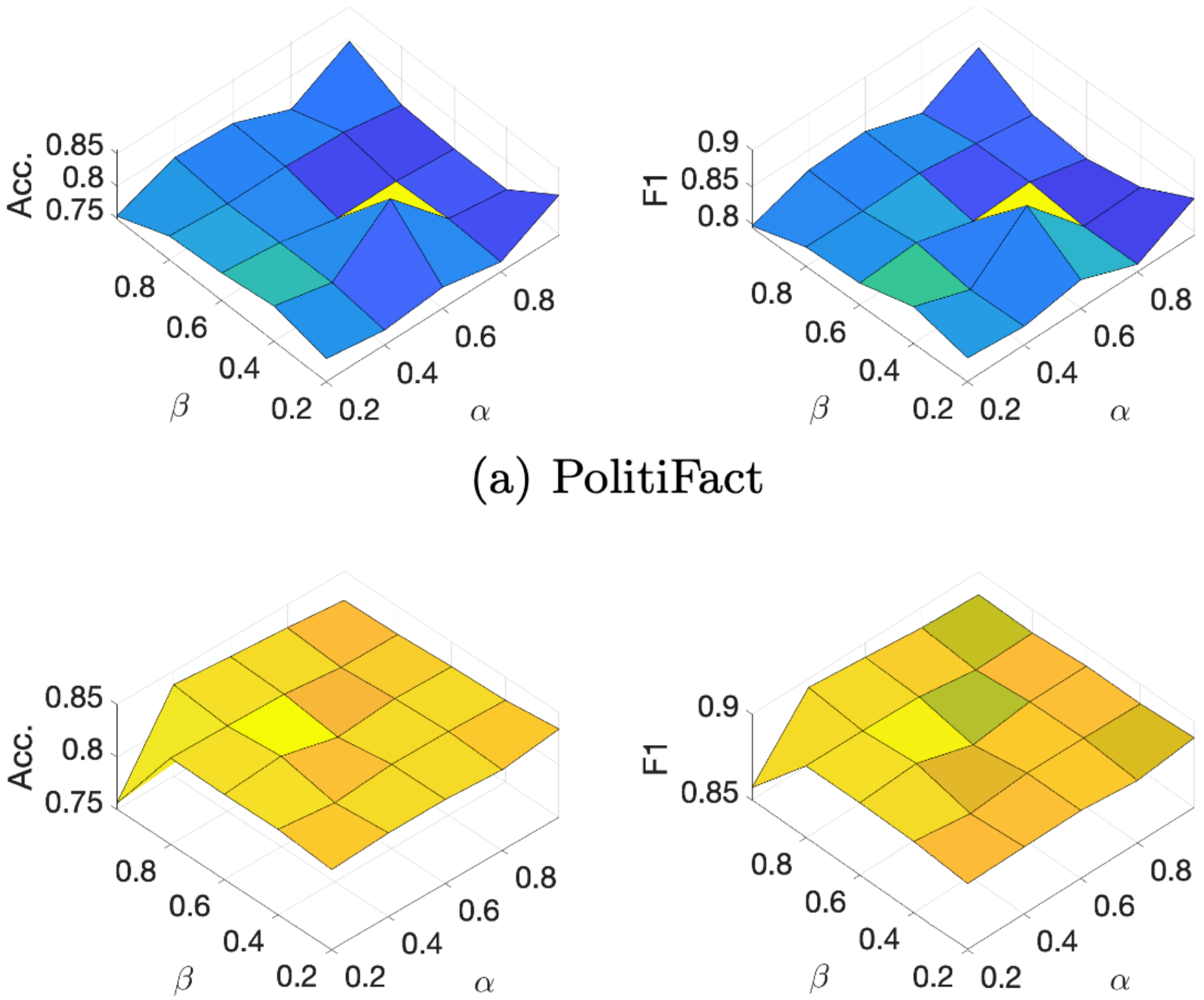
†: Text-based methods ≀: Image-based methods ‡: Multi-modal methods

- (4) textual information (SAFE\V) is more important compared to visual information (SAFE\T)

Experiments

Parameter Analysis

- α and β are used to allocate the relative importance between
  - multi-modal features (α)
  - similarity across modalities (β)
- Acc: 0.75~0.85
- F1: 0.8~0.9



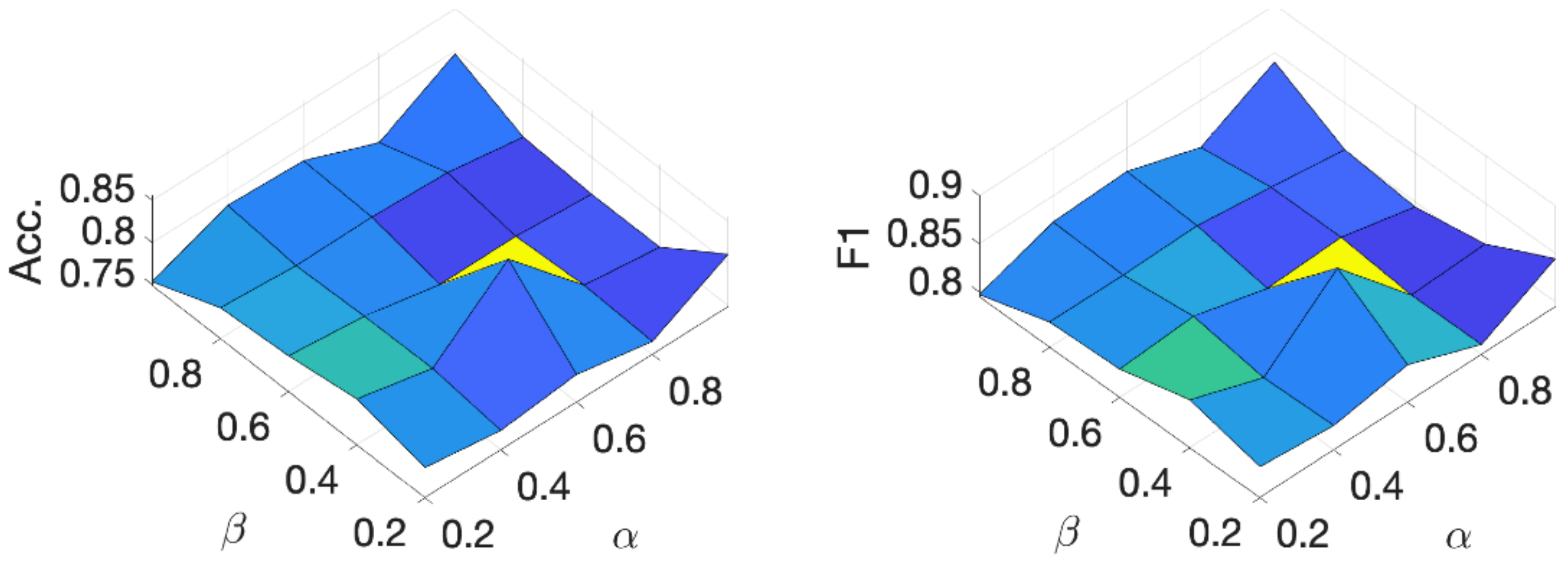
(a) PolitiFact

(b) GossipCop

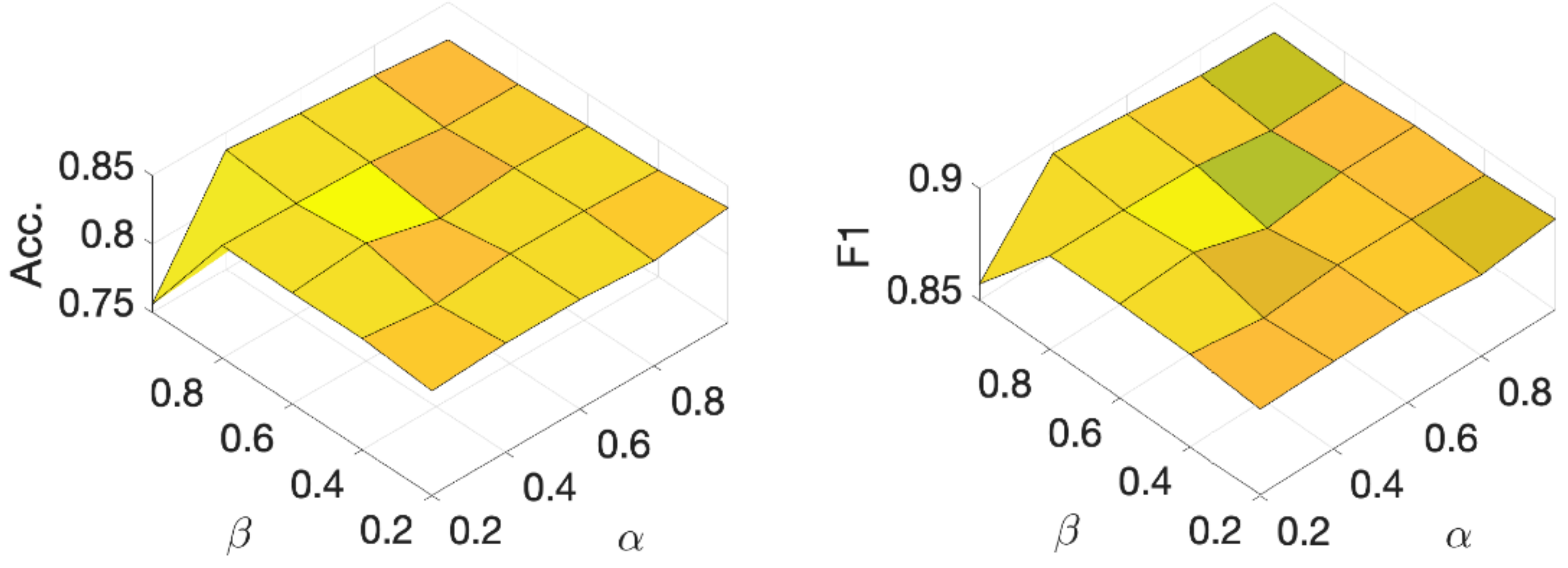
Experiments

Parameter Analysis

- The proposed method performs best
  - α : β = 0.4 : 0.6 in PolitiFact
  - α : β = 0.6 : 0.4 in GossipCop
- which again validates the importance of both multi-modal information and cross-modal relationship in predicting fake news.



(a) PolitiFact



(b) GossipCop

Experiments

Case Study

- Aim to answer the following questions:
  - Is there any real-world fake news story whose textual and visual information are not closely related to each other?
  - If there is, can SAFE correctly recognize such irrelevance and further recognize its falsity?
- For this purpose, author went through the news articles in the two datasets, and compared their ground truth labels with their similarity scores computed by SAFE.

Experiments

Case Study

- Gap between textual and visual information exist for some fictitious stories for (but not limited to) two reasons:

  1. Such stories are difficult to be supported by non-manipulated images

  - In Fig(a), where no voting- and bill-related image is actually available.

(a) $s = 0.024$

Experiments

Case Study

- Gap between textual and visual information exist for some fictitious stories for (but not limited to) two reasons:

  1. Such stories are difficult to be supported by non-manipulated images

     - Compared to the couples having a real intimate relationship, the fake ones often have rare group photos or use collages.

(c) $s = 0.001$

(c) $s = 0.983$

Experiments

Case Study

- Gap between textual and visual information exist for some fictitious stories for (but not limited to) two reasons:

  2. Using “attractive” though not closely relevant images can help increase the news traffic

  - the fake news in Fig. (b) includes an image with a smiling individual that conflicts with the death story

(b) $s = 0.044$

Experiments

Case Study

(a) $s = 0.024$ (b) $s = 0.044$

Fig. 5. Fake News

(a) $s = 0.966$ (b) $s = 0.975$ (c) $s = 0.983$

Fig. 6. True News

- SAFE helps correctly assess the relationship (similarity) between news textual and visual information.
- For fake news stories in Fig. 5, their corresponding similarity scores are all low and SAFE correctly labels them as fake news. Similarly, SAFE assigns all true news stories in Fig. 6 a high similarity score, and predicts them as true news

Conclusion

- Proposed a similarity-aware multi-modal method, named SAFE, for FakeNews Detection

- The method extracts both textual and visual features of news content, and investigates their relationship.

- Experimental results indicate multi-modal features and the cross-modal relationship (similarity) are valuable with a comparable importance in fake news detection.

Experiments

By ourselves

Outline
of Our Experiments

Preprocessing

Rebuild Model Training / Testing

Optimization

Preprocessing

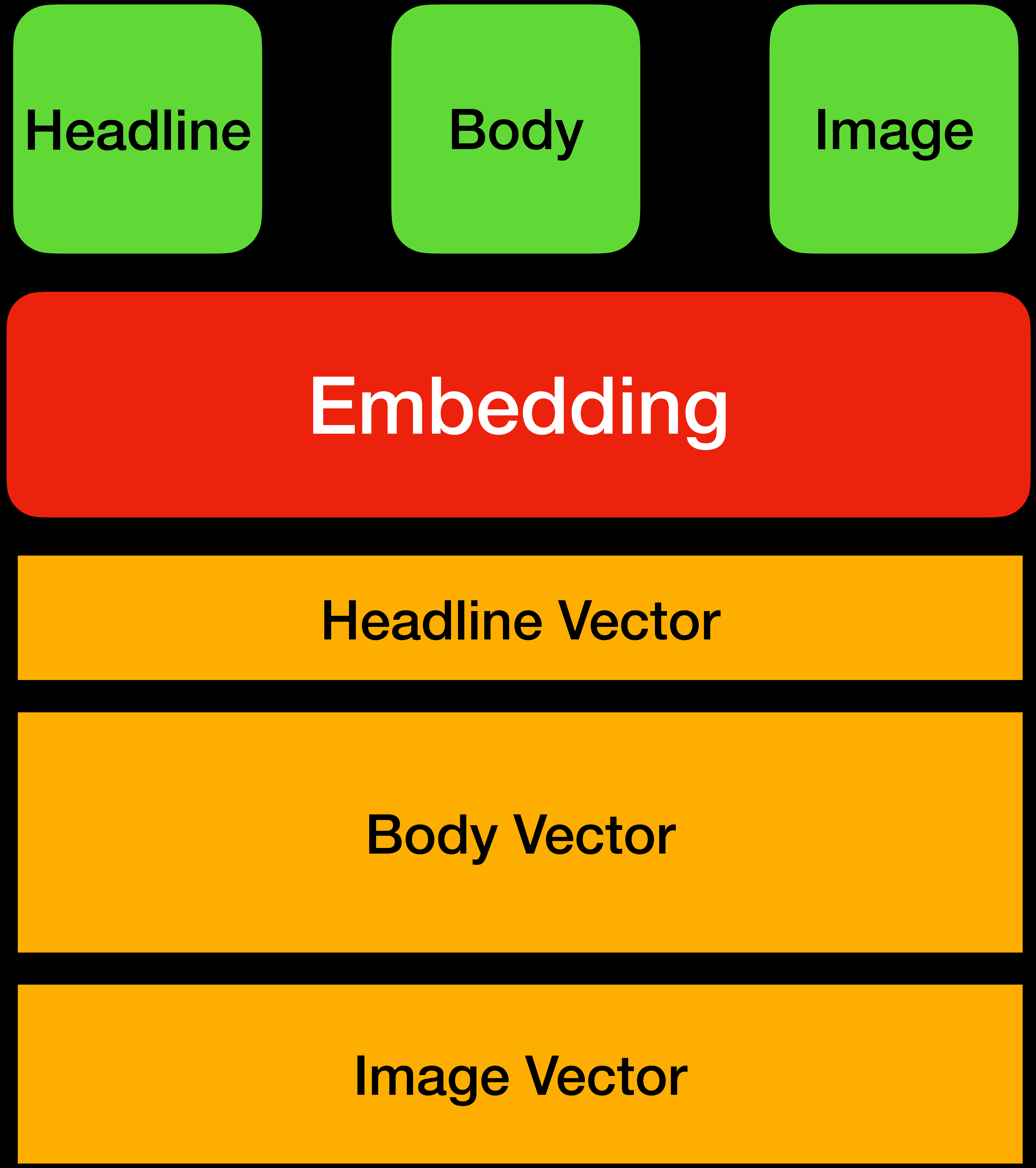
Show and Tell (image2sentence)

- Remove null image file (0 bytes file)
- Use MSCOCO pretrain weight
- Image fed into show and tell accuracy is bad
- Caption to SIF embedding

a man is holding a tennis

Preprocessing
Embedding

- Use SIF Embedding
- From Paper : A Simple but Tough-to-Beat Baseline for Sentence Embeddings
- Github : https://github.com/PrincetonML/SIF



Rebuild Model Training / Testing

Tensorflow / fix code bug

- npy load into code will encounter TypeError

Requirements

- Python 3.7
- TensorFlow 2.2
- xlwt
- nltk

```
 ----------- epoch No.1 -----------
fold No. 1
 step 50, training accuracy: clickbait 0.5625,  loss 0.425017
 step 100, training accuracy: clickbait 0.90625,  loss 0.288553
 step 150, training accuracy: clickbait 0.984375,  loss 0.0957951
 step 200, training accuracy: clickbait 1,  loss 0.026231
tf version: 2.2.0
Traceback (most recent call last):
  File "train.py", line 171, in <module>
    tf.compat.v1.app.run()
```

```
  File "/home/user/miniconda/envs/py36/lib/python3.6/site-packages/tensorflow/python/client/session.py", line 263, in for_fetch
    (fetch, type(fetch)))
TypeError: Fetch argument None has invalid type <class 'NoneType'>
(py36) user ➜ ~/code/SAFE (master ✗) $
```

Rebuild Model Training / Testing

Tensorflow / fix code bug

```
with tf.Graph().as_default():
    session_conf = tf.compat.v1.ConfigProto(
        allow_soft_placement=FLAGS.allow_soft_placement,
        log_device_placement=FLAGS.log_device_placement)
    session_conf.gpu_options.allow_growth = True
    sess = tf.compat.v1.Session(config=session_conf)

    with sess.as_default():
        layer = Graph(head_size=FLAGS.head_size,
                      body_size=FLAGS.body_size,
                      image_size=FLAGS.image_size,
                      input_len=32,
                      embedding_len=300,
                      l2_reg_lambda=FLAGS.l2_reg_lambda,
                      lr=FLAGS.learning_rate,
                      num_filters=FLAGS.num_filters,
                      final_len=FLAGS.final_len)

        # Initialize all variables
        sess.run(tf.compat.v1.global_variables_initializer())
        writer1 = tf.compat.v1.summary.FileWriter(boarddir + '/plot_1', sess.graph)
```

Rebuild Model Training / Testing

Tensorflow / fix code bug

tf.compat.v1.summary.merge_all

View source on GitHub

Merges all summaries collected in the default graph.

```
tf.compat.v1.summary.merge_all(
    key=tf.GraphKeys.SUMMARIES, scope=None, name=None
)
```

Migrate to TF2

Caution: This API was designed for TensorFlow v1. Continue reading for details on how to migrate from this API to a native TensorFlow v2 equivalent. See the TensorFlow v1 to TensorFlow v2 migration guide for instructions on how to migrate the rest of your code.

```
self.merged = tf.compat.v1.summary.merge_all()
```

```
# training set
summary1 = sess.run(layer.merged, feed_dict={layer.input_headline_: x_batch_head,
                                             layer.input_body_: x_batch_body,
                                             layer.input_image_: x_batch_image,
                                             layer.input_y: y_batch,
                                             layer.dropout_keep_prob: FLAGS.dropout_keep_prob,
                                             layer.batch_size: FLAGS.batch_size})
```

Preprocessing

NULL & Questions

- Article、Image Unpair
- PolitiFact -> 316
- Take 75 as Testing
- Author : jwu21@email.wm.edu

	PolitiFact		
	Fake	True	Overall
# News articles	432	624	1,056
– with textual information	420	528	948
– with visual information	336	447	783

```
ACC. = 0.8157894736842105
precision = 0.7954545454545454
recall = 0.875
f1_score = 0.8333333333333334
tp:35.0
tn: 27.0
fp:9.0
fn: 5.0
```

Rebuild Model Training / Testing

Tensorflow / fix code bug

- Given HeadLine、Body、Image
- Predict Real or Fake
- For example : -> Real

"About the Show

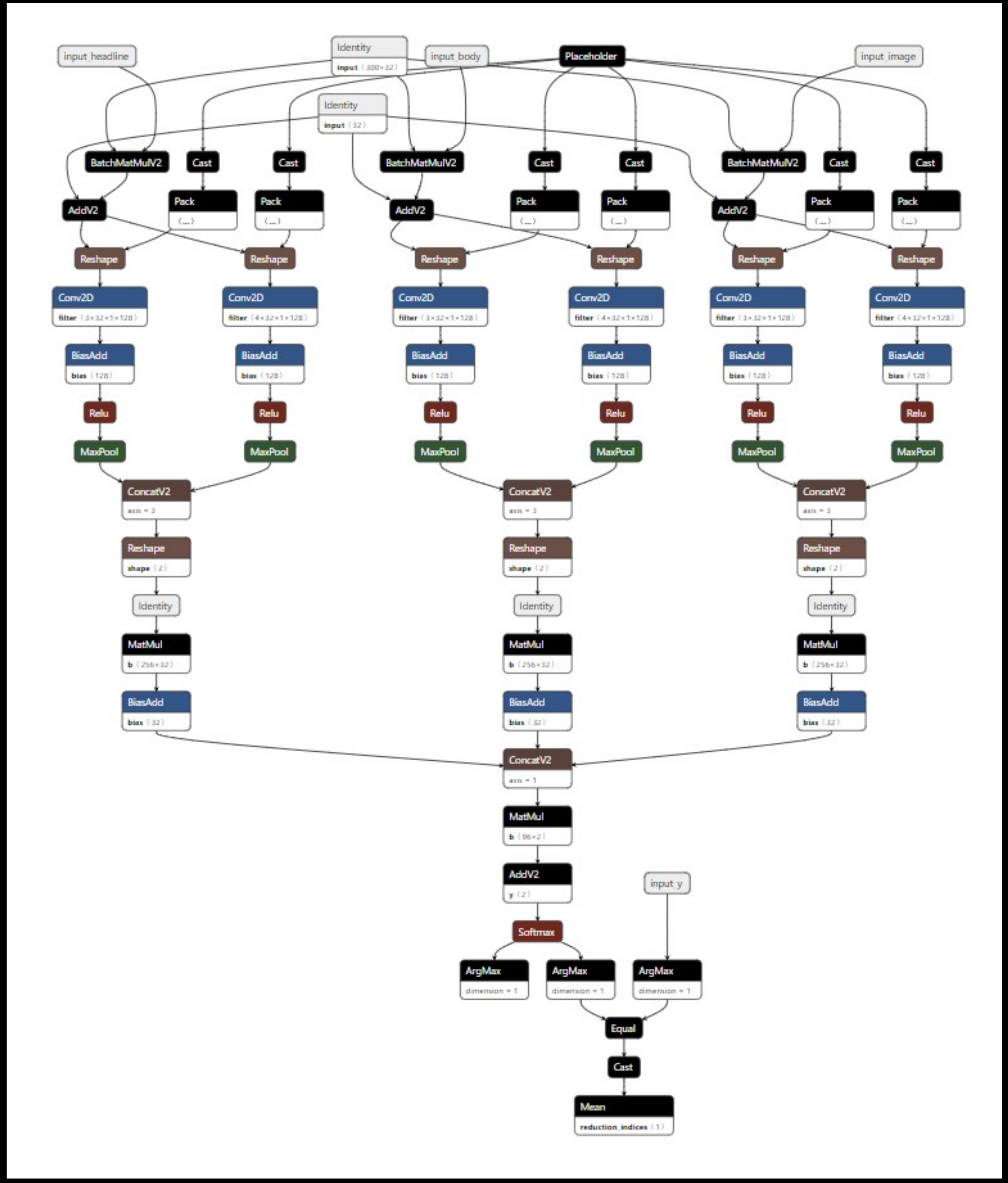
The Colbert Report

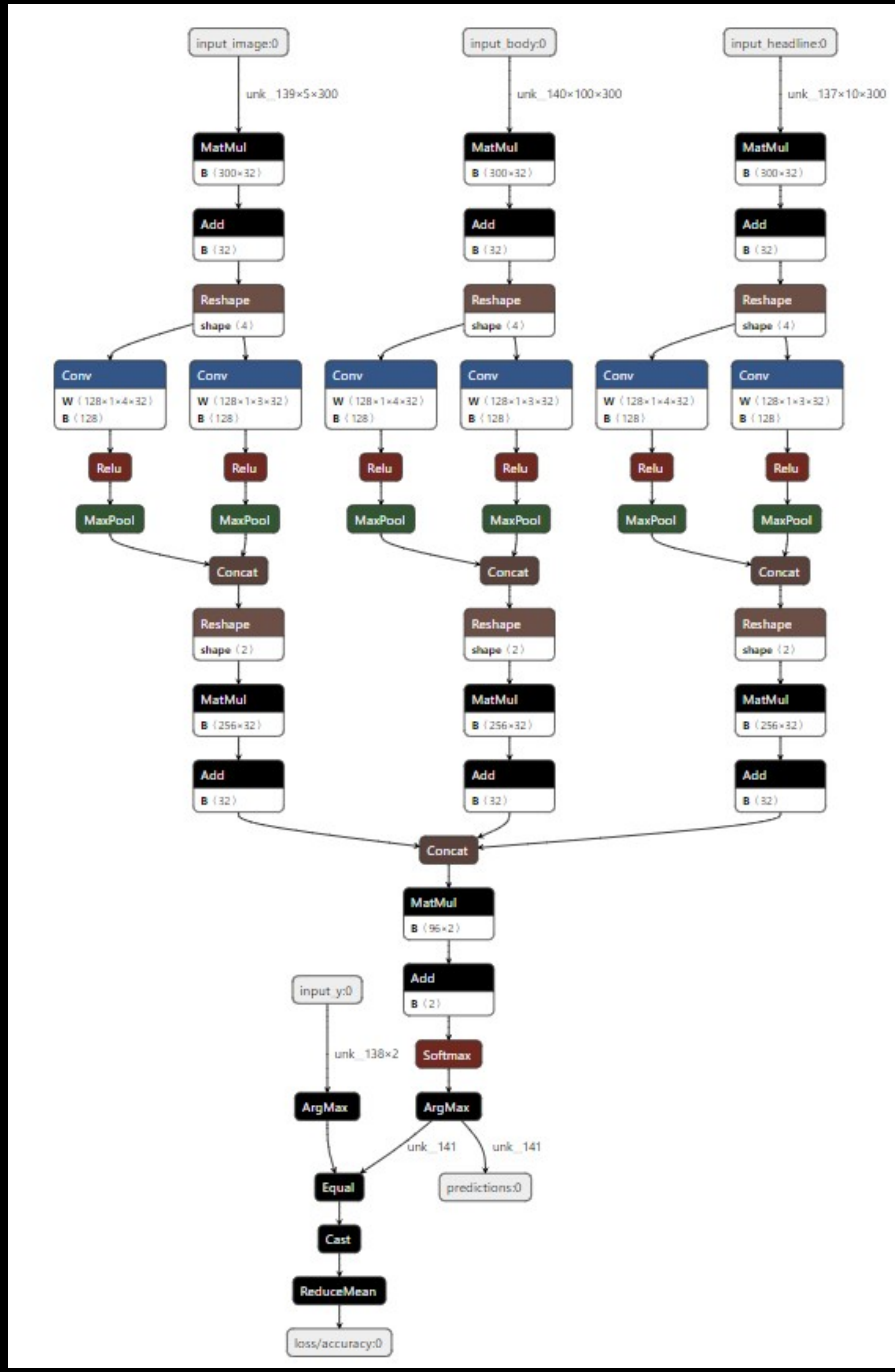
As America's most fearless purveyor of \”
truthiness,\" Stephen Colbert shines a light on
ego-driven punditry, moral hypocrisy and
government incompetence, raising the bar for
political satire."

Optimization

Checkpoint to PB to ONNX

- Use constant graph
- Convert to model.pb
- Use tf2Onnx to convert ONNX

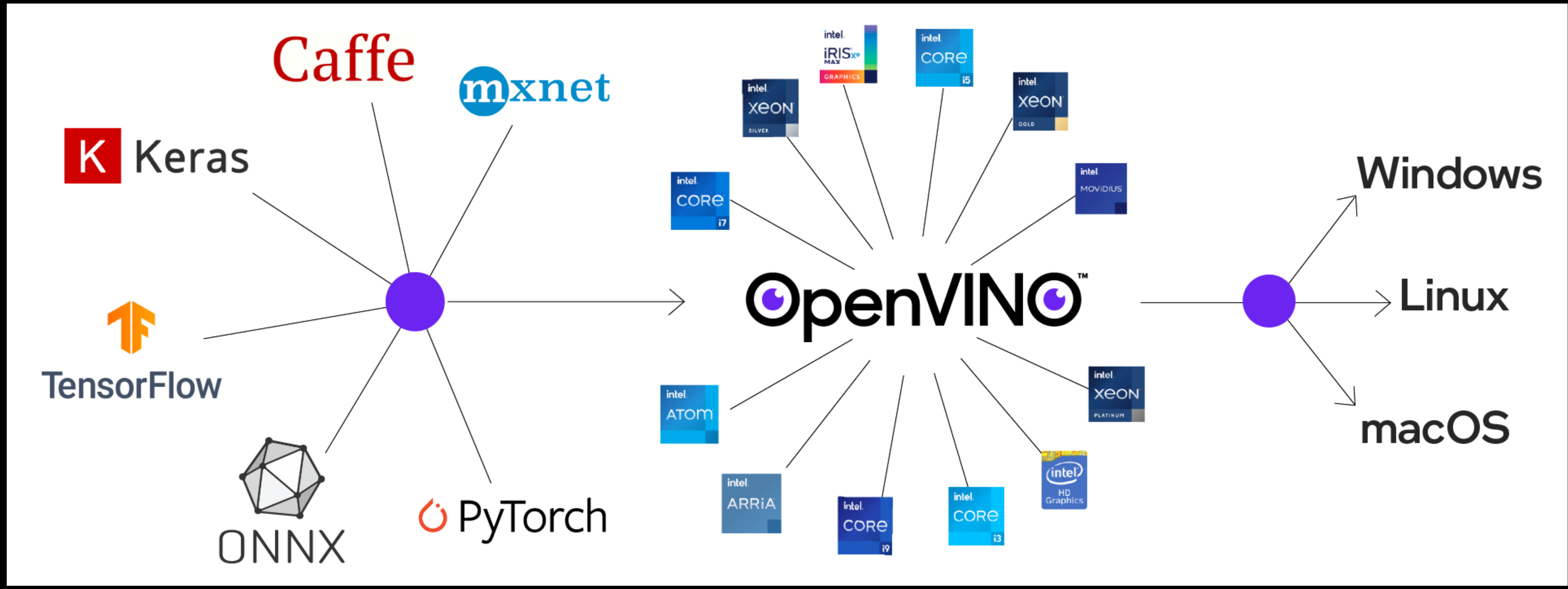


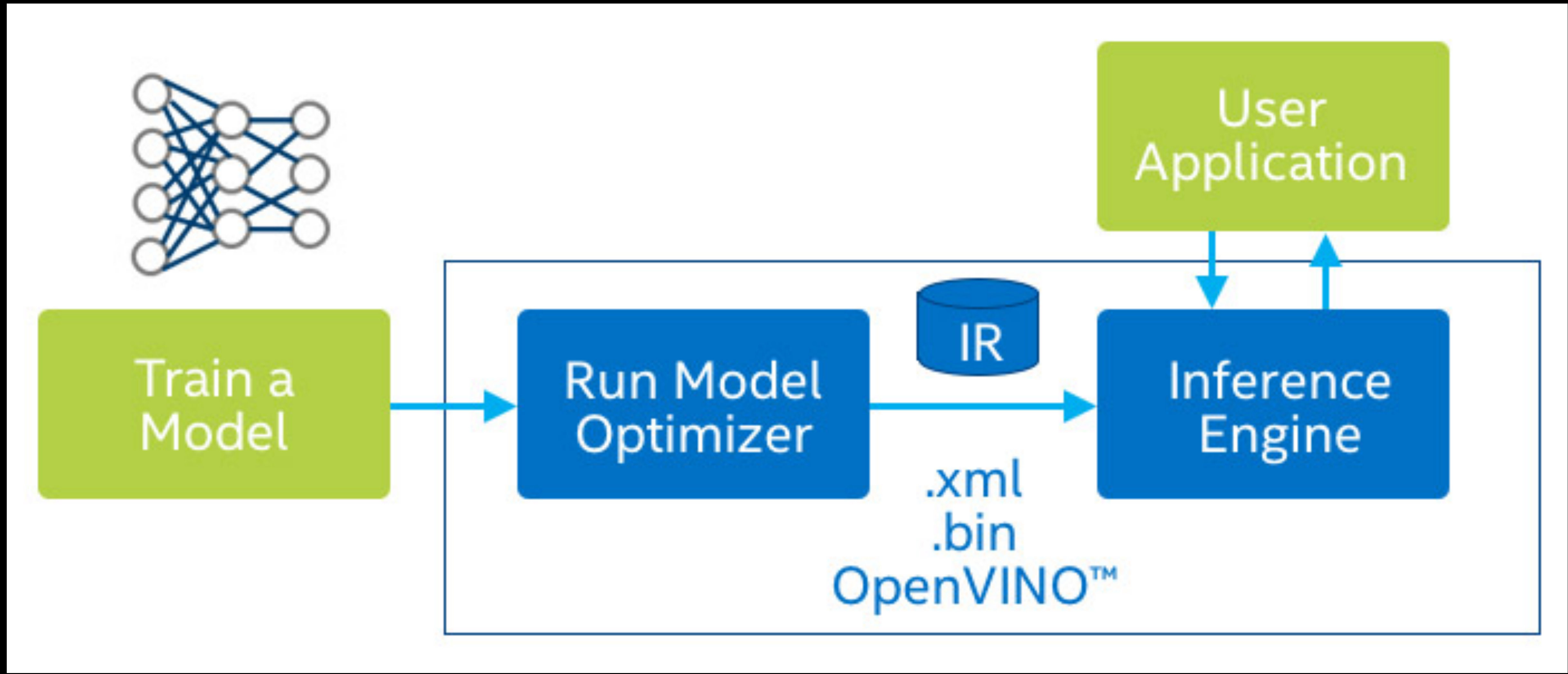


Optimization

OpenVINO optimaztion

- Intel OpenVINO Tool Kit
- Deep Learning Framework Supported
- Reduce Resource Demands
- Efficiently Deploy On A Range of Intel® Platforms
- Optimize -> Inference

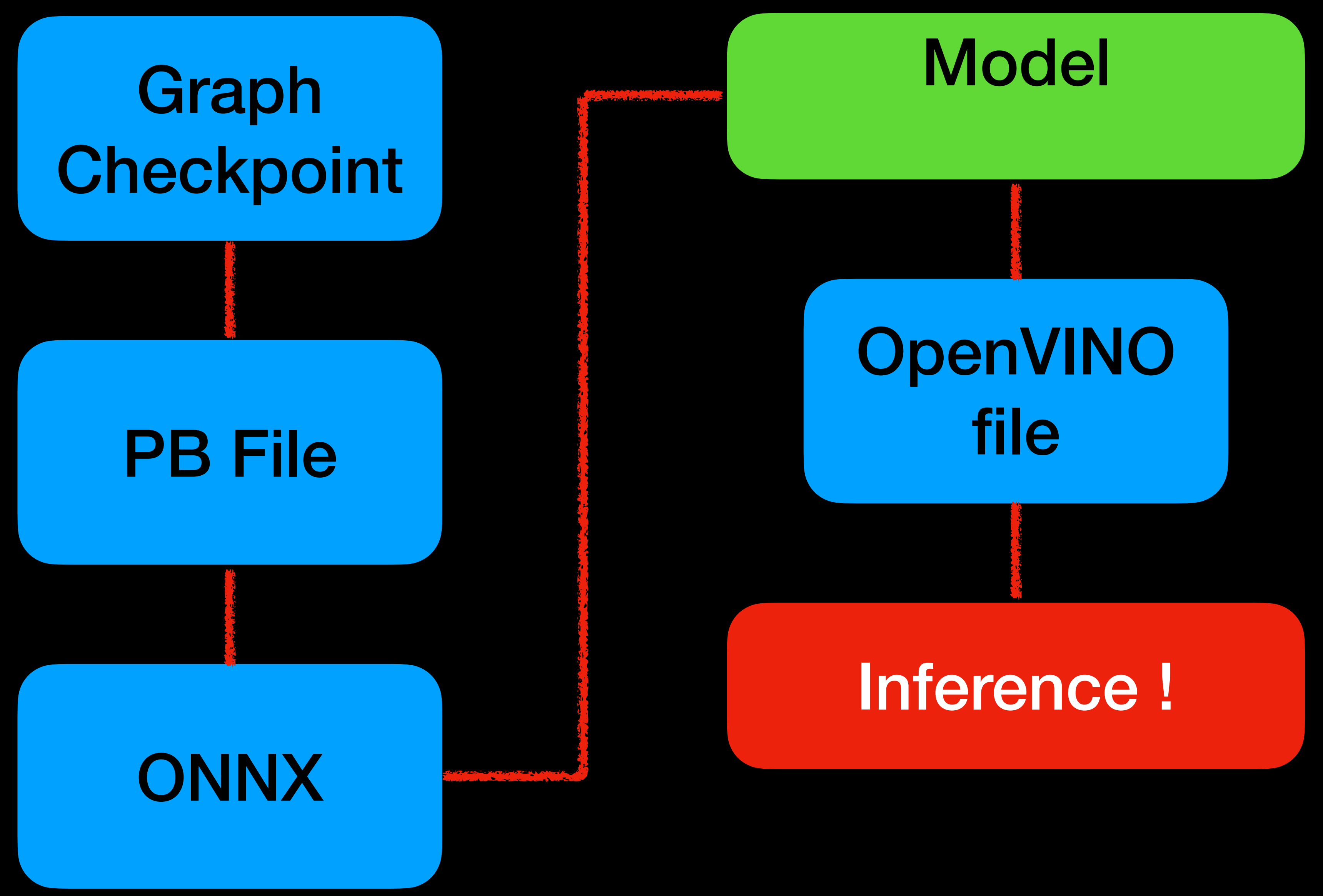




Optimization

Optimaztion Flow

- TensorFlow Checkpoint
- TensorFlow PB file
- ONNX file
- OpenVINO file(.xml 、.bin)

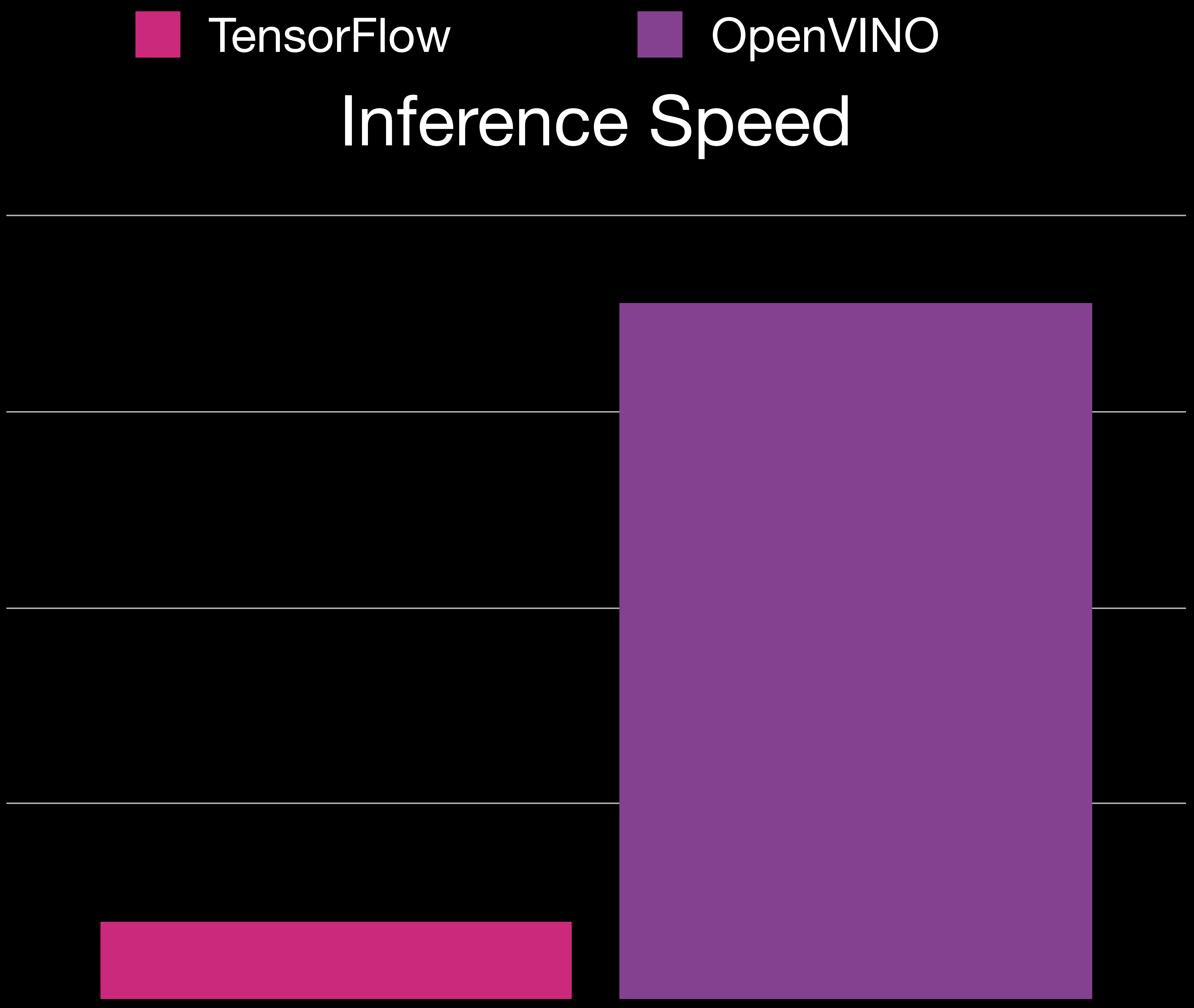


Optimization

Benchmark

- TensorFlow Inference Time : 61.54 ms
- OpenVINO Inference Time : 7.32 ms
- Speed Up 8.4X

```
Count:      8192 iterations
Duration:   60030.23 ms
Latency:    28.55 ms
Throughput: 29476.35 FPS
```



Comments

of Similarity-Aware FakE news detection (SAFE)

- Add across-modal relationship (similarity) to detect fake news detection
- Use image2sentence get image caption
- Baseline:
    - Text feature baseline only one of traditional method
    - Multi-modal baseline only one to compared

Q & A

Optimization

of SAFE

Algorithm 1: SAFE

Input: $A = \{(T_j, V_j)\}_{j=1}^{m}$, $Y = \{y_j\}_{j=1}^{m}$, $H = \{h_k\}_{k=1}^{g}$, γ

Output: $\theta_p = \{\mathbf{W}_p, \mathbf{b}_p\}$, $\theta_t = \{\mathbf{W}_t, \mathbf{b}_t, \mathbf{w}_t, b_t\}$, $\theta_v = \{\mathbf{W}_v, \mathbf{b}_v, \mathbf{w}_v, b_v\}$

Randomly initialize $\mathbf{W}_p, \mathbf{b}_p, \mathbf{W}_t, \mathbf{b}_t, \mathbf{w}_t, b_t, \mathbf{W}_v, \mathbf{b}_v, \mathbf{w}_v, b_v$;

收敛

while *not convergence* **do**

  foreach (T_j, V_j) **do**

    Update θ_p: $\{\mathbf{W}_p, \mathbf{b}_p\} \leftarrow$ Eq. (12);

    foreach h_k **do**

      Update θ_t: $\{\mathbf{W}_t, \mathbf{b}_t, \mathbf{w}_t, b_t\} \leftarrow$ Eqs. (14-18);

      Update θ_v: similar to updating θ_t;

    end

  end

end

return $\mathbf{W}_p, \mathbf{b}_p, \mathbf{W}_t, \mathbf{b}_t, \mathbf{w}_t, b_t, \mathbf{W}_v, \mathbf{b}_v, \mathbf{w}_v, b_v$

Update θ_p. Let γ be the learning rate, the partial derivative of $\mathcal{L}$ w.r.t. θ_p is:

$$\theta_p \leftarrow \theta_p - \gamma \cdot \alpha \frac{\partial \mathcal{L}_p}{\partial \theta_p}. \tag{11}$$

As $\theta_p = \{\mathbf{W}_p, \mathbf{b}_p\}$, updating θ_p is equivalent to updating both $\mathbf{W}_p$ and $\mathbf{b}_p$ in each iteration, which respectively follow the following rules:

$$\mathbf{W}_p \leftarrow \mathbf{W}_p - \gamma \cdot \alpha \Delta\mathbf{y}(\mathbf{t} \oplus \mathbf{v})^\top, \quad \mathbf{b}_p \leftarrow \mathbf{b}_p - \gamma \cdot \alpha \Delta\mathbf{y}, \tag{12}$$

where $\Delta\mathbf{y} = [\hat{y} - y, y - \hat{y}]^\top$.

Update θ_t. The partial derivative of $\mathcal{L}$ w.r.t. θ_t is generally computed by

$$\theta_t \leftarrow \theta_t - \gamma(\alpha \frac{\partial \mathcal{L}_p}{\partial \mathcal{M}_t} \frac{\partial \mathcal{M}_t}{\partial \theta_t} + \beta \frac{\partial \mathcal{L}_s}{\partial \mathcal{M}_t} \frac{\partial \mathcal{M}_t}{\partial \theta_t}). \tag{13}$$

Let $\nabla \mathcal{L}_*(\mathbf{t}) = \frac{\partial \mathcal{L}_*}{\partial \mathcal{M}_t}$, $\mathbf{t}_0 = \frac{\mathbf{t}}{||\mathbf{t}||}$, $\mathbf{v}_0 = \frac{\mathbf{v}}{||\mathbf{v}||}$, and $\mathbf{W}_{p,L}$ denote the first d columns of $\mathbf{W}_p$, we can have

$$\nabla \mathcal{L}_p(\mathbf{t}) = \mathbf{W}_{p,L}^\top \Delta\mathbf{y}, \tag{14}$$

$$\nabla \mathcal{L}_s(\mathbf{t}) = \frac{1-y}{2s\,\|\mathbf{t}\|}((2s-1)\mathbf{t}_0 - \mathbf{v}_0), \tag{15}$$

based on which the parameters in θ_t are respectively updated as follows:

$$\mathbf{W}_t \leftarrow \mathbf{W}_t - \gamma \cdot \mathbf{D}_t \mathbf{B}_t, \quad \mathbf{b}_t \leftarrow \mathbf{b}_t - \gamma \cdot \mathbf{B}_t, \tag{16}$$

$$\mathbf{w}_t \leftarrow \mathbf{w}_t - \gamma \cdot \mathbf{x}_t^{\hat{i}:(\hat{i}+h-1)} \mathbf{W}_t^\top \mathbf{B}_t, \quad b_t \leftarrow b_t - \gamma \cdot \mathbf{W}_t^\top \mathbf{B}_t, \tag{17}$$

where $\hat{i} = \arg\max_i \{c_t^i\}_{i=1}^{n-h+1}$, $\mathbf{D}_t \in \mathbb{R}^{d \times d}$ is a diagonal matrix with entry value $c_t^{\hat{i}}$, and

$$\mathbf{B}_t = \alpha \nabla \mathcal{L}_p(\mathbf{t}) + \beta \nabla \mathcal{L}_s(\mathbf{t}). \tag{18}$$

Update θ_v. It is similar to updating θ_t; we omit details due to space constraints.