Cross-modal Ambiguity Learning for Multimodal Fake News Detection

Yixuan Chen
School of Computer Science
Shanghai Key Laboratory of Data Science
Fudan University
Shanghai, China
yixuanchen20@fudan.edu.cn

Dongsheng Li
Microsoft Research Asia
Shanghai, China
dongsli@microsoft.com

Peng Zhang
School of Computer Science
Shanghai Key Laboratory of Data Science
Fudan University
Shanghai, China
zhangpeng_@fudan.edu.cn

Jie Sui
University of Chinese Academy of Sciences
Beijing, China
suijie@ucas.ac.cn

Qin Lv
University of Colorado Boulder
Boulder, United States
qin.lv@colorado.edu

Tun Lu, Li Shang*
School of Computer Science
Shanghai Key Laboratory of Data Science
Fudan University
Shanghai, China
{lutun,lishang}@fudan.edu.cn

WWW_2022

Reported by Xiaoke Li

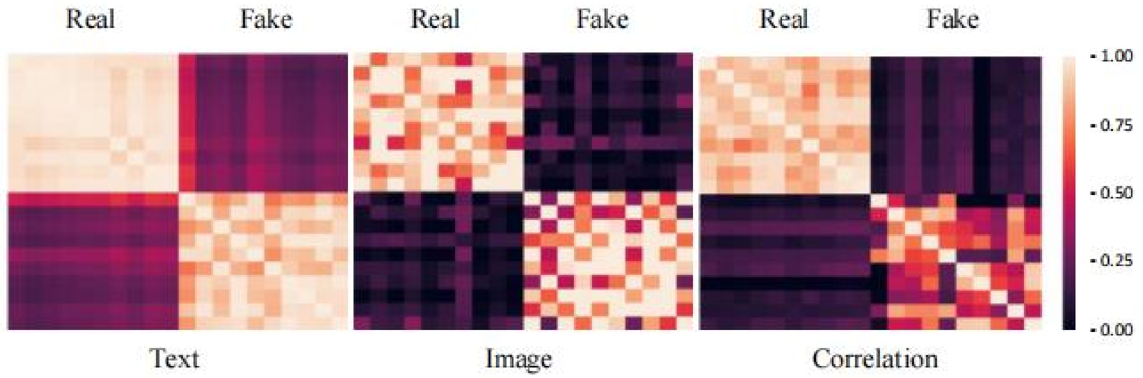
Introduction

Fake news: "An employee of the Jefferson County morgue died this morning after being accidentally cremated by one of his coworkers."

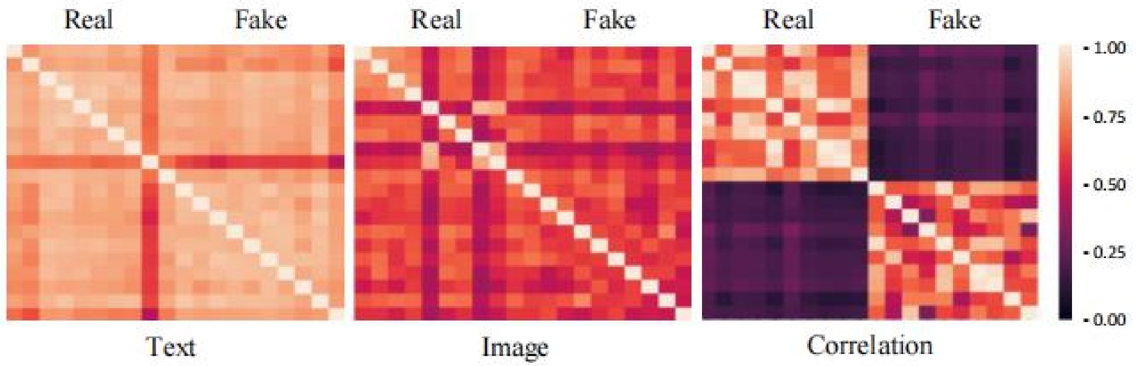
Real news: "You left in peace, left me in pieces."

Figure 1: Illustrations of cross-modal ambiguity.

Introduction



(a) Cross-modal correlation may be unhelpful or even harmful when text and image alone are sufficient.



(b) Cross-modal correlation can present extra insights when text and image alone are insufficient.

Figure 2: Illustration of the importance of ambiguity-aware cross-modal correlation using the Weibo dataset [15]. Each cell of the heat map represents the cosine similarity between the representations of each text or image pair.

Method

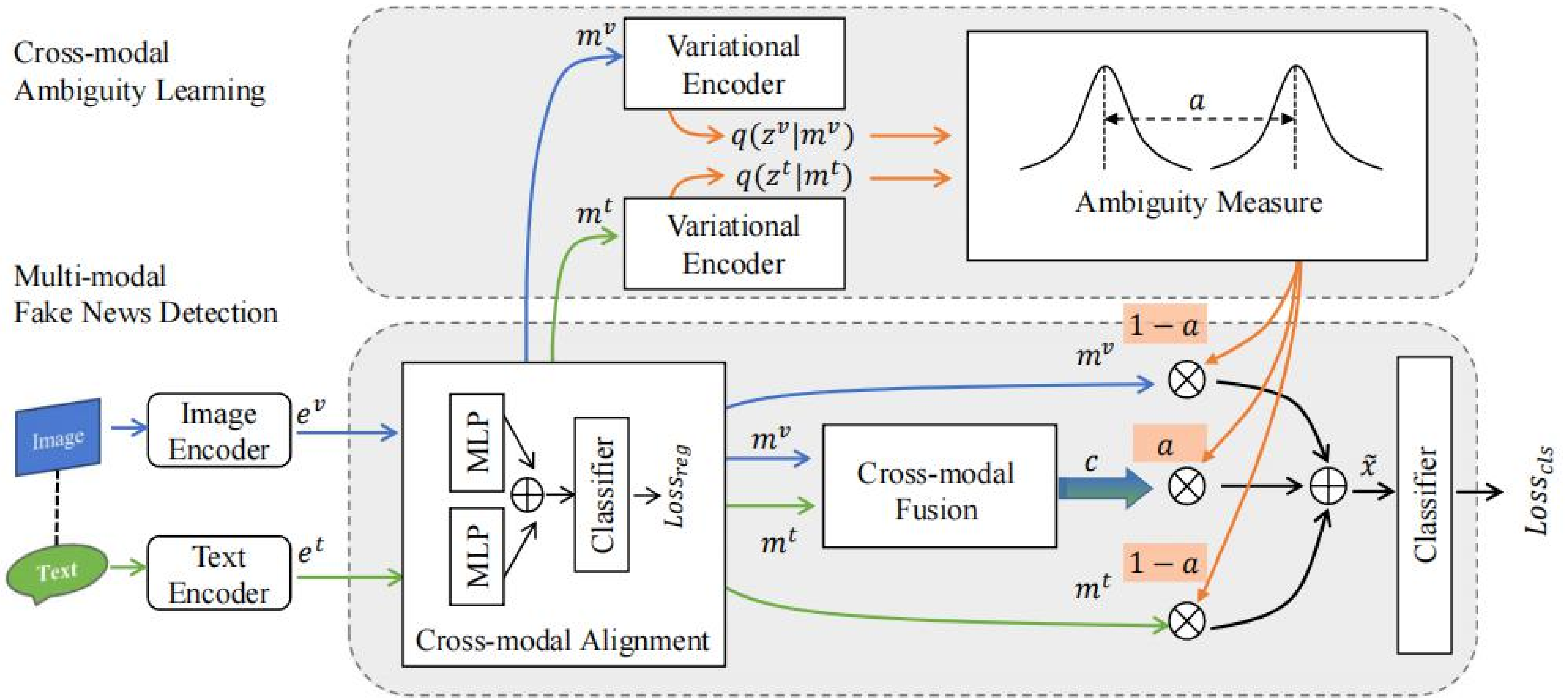
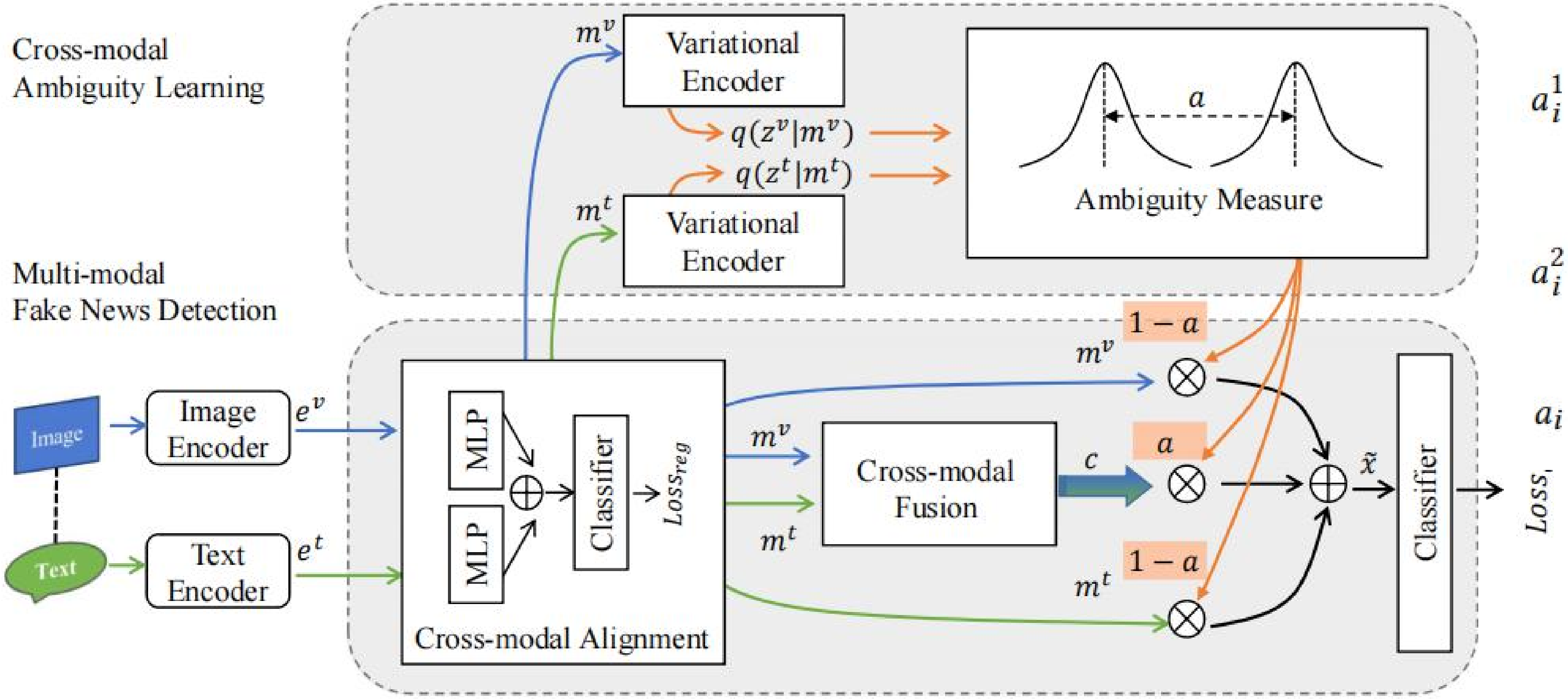


Figure 3: The architecture of the proposed CAFE method. For news with different levels of ambiguity, the proposed cross-modal ambiguity learning module can adaptively aggregate the unimodal features and cross-modal correlations to improve fake news classification. We set the weight of cross-modal correlation as a and the weight of unimodal features as $1-a$, so that the classifier will rely more on cross-modal correlation when a is large, i.e., stronger ambiguity appears.

Method



$$a_i^1 = \left(\frac{\mathrm{D}_{KL}\left(q\left(z_i^t \mid m_i^t\right) \| q\left(z_i^v \mid m_i^v\right)\right)}{\mathrm{D}_{KL}\left(q\left(z^t\right) \| q\left(z^v\right)\right)} \right), \tag{5}$$

$$a_i^2 = \left(\frac{\mathrm{D}_{KL}\left(q\left(z_i^v \mid m_i^v\right) \| q\left(z_i^t \mid m_i^t\right)\right)}{\mathrm{D}_{KL}\left(q\left(z^v\right) \| q\left(z^t\right)\right)} \right), \tag{6}$$

$$a_i = \mathrm{sigmoid}\left(\frac{1}{2}\left(a_i^1 + a_i^2\right)\right). \tag{7}$$

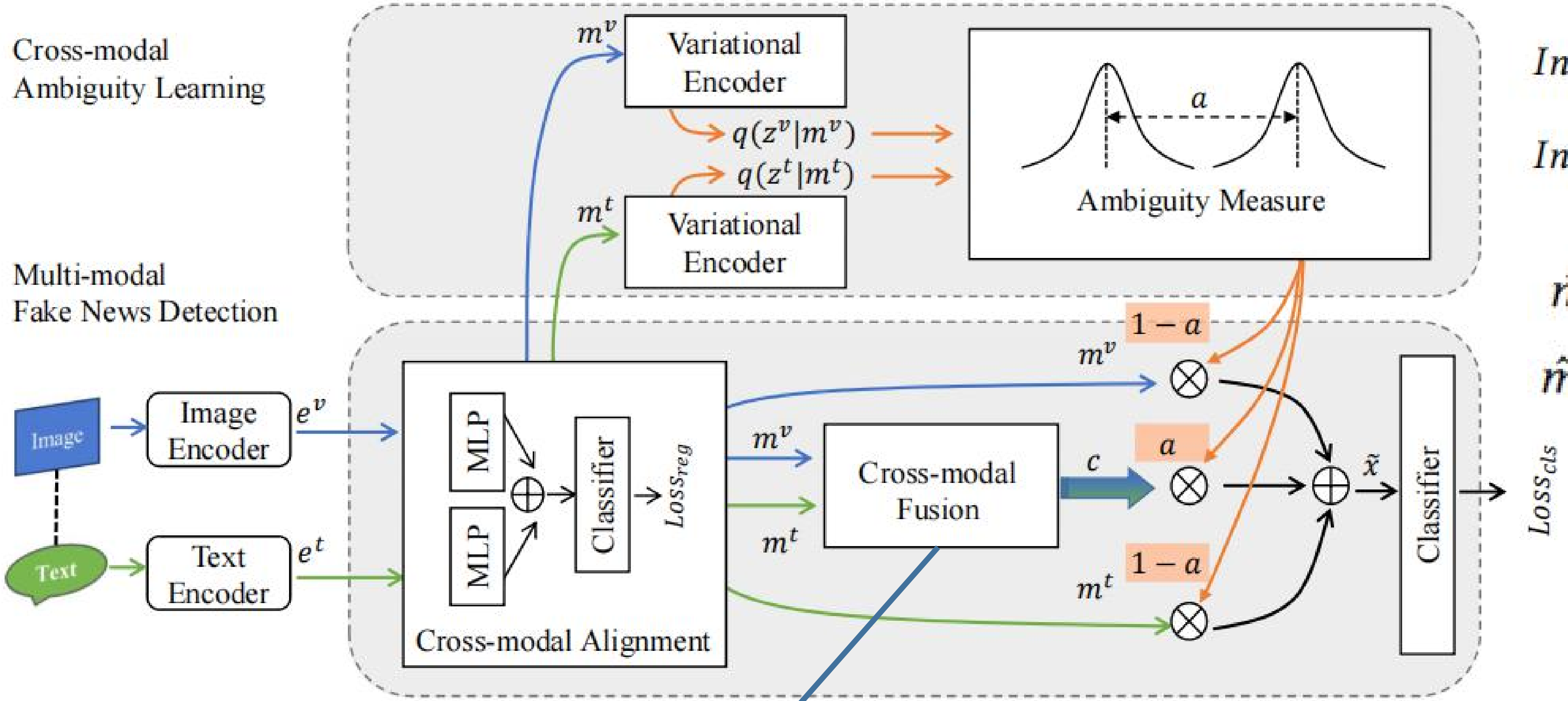
$$\mathcal{L}_{reg} = \begin{cases} 1 - \cos\left(e^t, e^v\right) & \text{if } y_2 = 1. \\ \max\left(0, \cos\left(e^t, e^v\right) - d\right) & \text{if } y_2 = 0. \end{cases} \tag{1}$$

$$q\left(z_i^t \mid m_i^t\right) = \mathcal{N}\left(z_i^t \mid \mu\left(m_i^t\right), \sigma\left(m_i^t\right)\right), \tag{2}$$

$$q\left(z_i^v \mid m_i^v\right) = \mathcal{N}\left(z_i^v \mid \mu\left(m_i^v\right), \sigma\left(m_i^v\right)\right). \tag{3}$$

$$\begin{aligned} q\left(z^t\right) &= \mathbb{E}_{\mathrm{Pr}_{\mathrm{data}}(m^t)}\left[q\left(z^t \mid m^t\right)\right] = \frac{1}{N}\sum_{i=1}^{N} q\left(z_i^t \mid m_i^t\right), \\ q\left(z^v\right) &= \mathbb{E}_{\mathrm{Pr}_{\mathrm{data}}(m^v)}\left[q\left(z^v \mid m^v\right)\right] = \frac{1}{N}\sum_{i=1}^{N} q\left(z_i^v \mid m_i^v\right). \end{aligned} \tag{4}$$

Method



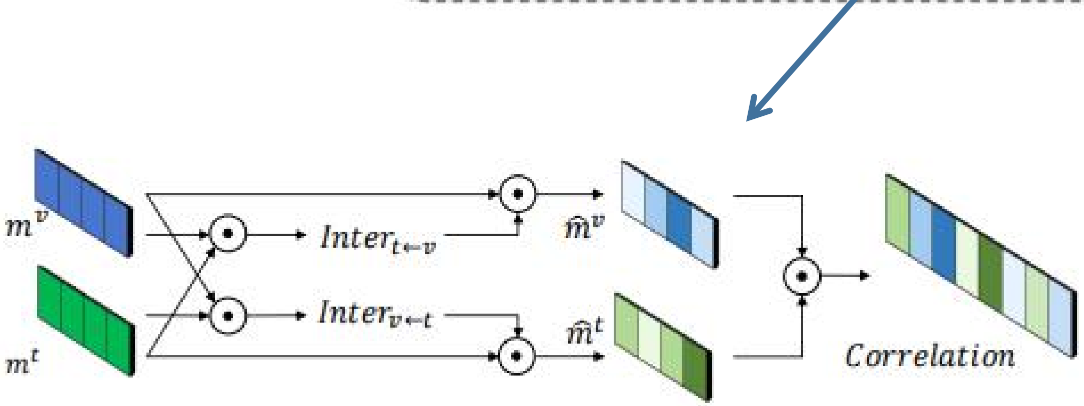
$$InterC_{t\leftarrow v} = \text{softmax}\left([m^t][m^v]^T/\sqrt{dim}\right). \quad (8)$$

$$InterC_{v\leftarrow t} = \text{softmax}\left([m^v][m^t]^T/\sqrt{dim}\right). \quad (9)$$

$$\hat{m}^t = InterC_{T\leftarrow I}\times m^t. \quad (10)$$

$$\hat{m}^v = InterC_{I\leftarrow T}\times m^v. \quad (11)$$

$$c = \hat{m}^t \otimes \hat{m}^v. \quad (12)$$



$$\tilde{\mathbf{x}} = (a_{\mathbf{x}}\times c)\oplus\left((1-a_{\mathbf{x}})\times m^t\right)\oplus\left((1-a_{\mathbf{x}})\times m^v\right) \quad (13)$$

$$\tilde{y}_1 = \text{softmax}\left(MLP\left(\tilde{\mathbf{x}}\right)\right) \quad (14)$$

$$\mathcal{L}_{cls} = y_1 log\left(\tilde{y}_1\right) + \left(1-\tilde{y}_1\right) log\left(1-y_1\right) \quad (15)$$

$$\mathcal{L} = \mathcal{L}_{cls} + \beta\mathcal{L}_{reg}. \quad (16)$$

Experiments

Table 1: Performance comparison between CAFE and the two unimodal and six multi-modal baseline meth

	Method	Acc	Rumor			Non Rumor		
			P	R	F_1	P	R	F_1
Twitter	CAR	0.637	0.574	0.690	0.682	0.724	0.602	0.617
	VS	0.617	0.635	0.644	0.639	0.639	0.630	0.634
	RA	0.664	0.749	0.615	0.676	0.589	0.728	0.651
	EANN	0.648	0.810	0.498	0.617	0.584	0.759	0.660
	MAVE	0.745	0.801	0.719	0.758	0.689	0.777	0.730
	MKEMN	0.715	0.814	0.756	0.708	0.634	0.774	0.660
	SAFE	0.762	**0.831**	0.724	0.774	0.695	0.811	0.748
	MCNN	0.784	0.778	0.781	0.779	0.790	0.787	0.788
	CAFE	**0.806**	0.807	**0.799**	**0.803**	**0.805**	**0.813**	**0.809**
Weibo	CAR	0.745	0.705	0.765	0.750	0.756	0.725	0.740
	VS	0.726	0.732	0.712	0.722	0.720	0.74	0.73
	RA	0.772	0.854	0.656	0.742	0.720	0.889	0.795
	EANN	0.795	0.806	0.795	0.800	0.752	0.793	0.804
	MVAE	0.824	0.854	0.769	0.809	0.802	0.875	0.837
	MKEMN	0.814	0.823	0.799	0.812	0.723	0.819	0.798
	SAFE	0.816	0.818	0.815	0.817	0.816	0.818	0.817
	MCNN	0.823	**0.858**	0.801	0.828	0.787	0.848	0.816
	CAFE	**0.840**	0.855	**0.830**	**0.842**	**0.825**	**0.851**	**0.837**

Experiments

Table 2: Ablation study on the architecture design of CAFE on two datasets.

Method	Data	*Acc*	*Pre*	*Rec*	*F1*
CAFE w/o R	Twitter	0.791	0.834	0.744	0.787
	Weibo	0.830	0.875	0.801	0.837
CAFE w/o A	Twitter	0.786	0.767	0.790	0.779
	Weibo	0.829	0.831	0.826	0.828
CAFE w/o C	Twitter	0.806	0.807	0.799	0.803
	Weibo	0.827	0.863	0.805	0.833
CAFE	Twitter	0.806	0.807	0.799	0.803
	Weibo	0.840	0.855	0.830	0.842

Table 3: Performance comparison of different distance measurement methods in ambiguity learning methods.

Method	Data	*Acc*	*Pre*	*Rec*	*F1*
CAFE-COS	Twitter	0.793	0.823	0.753	0.787
	Weibo	0.837	0.848	0.829	0.838
CAFE-DIS	Twitter	0.784	0.801	0.753	0.776
	Weibo	0.834	0.843	0.828	0.835
CAFE-KL	Twitter	0.806	0.807	0.799	0.803
	Weibo	0.840	0.855	0.830	0.842

Table 4: Performance comparison between different cross-modal fusion methods.

Method	Data	*Acc*	*Pre*	*Rec*	*F1*
CAFE-CAT	Twitter	0.789	0.801	0.756	0.778
	Weibo	0.828	0.863	0.805	0.833
CAFE-CNN	Twitter	0.794	0.801	0.763	0.782
	Weibo	0.832	0.843	0.825	0.834
CAFE	Twitter	0.806	0.807	0.799	0.803
	Weibo	0.840	0.855	0.830	0.842

Experiments

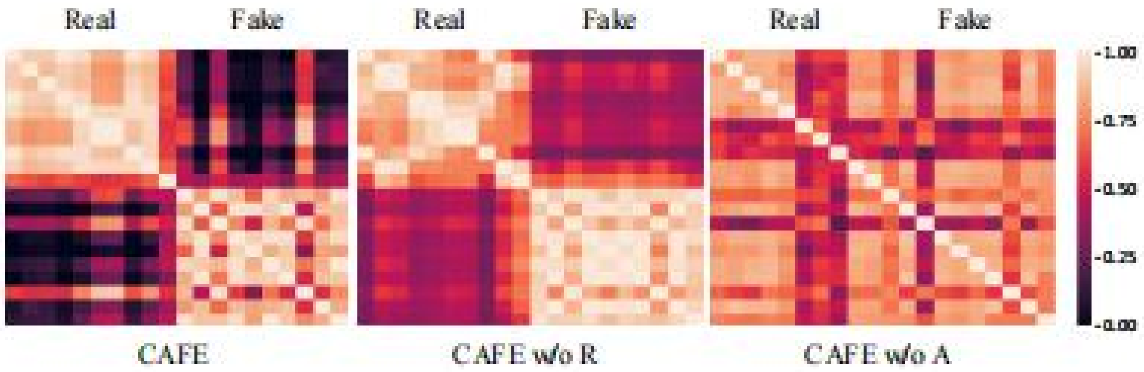


Figure 5: The result of quantitative analysis. CAFE presents clear inter-class difference and intra-class similarity, while CAFE w/o A and CAFE w/o R yield poor capability to learn inter-class difference.

Thanks