Fusion-Extraction Network for Multimodal Sentiment Analysis

Tao Jiang, Sun Yat-Sen University

Trend

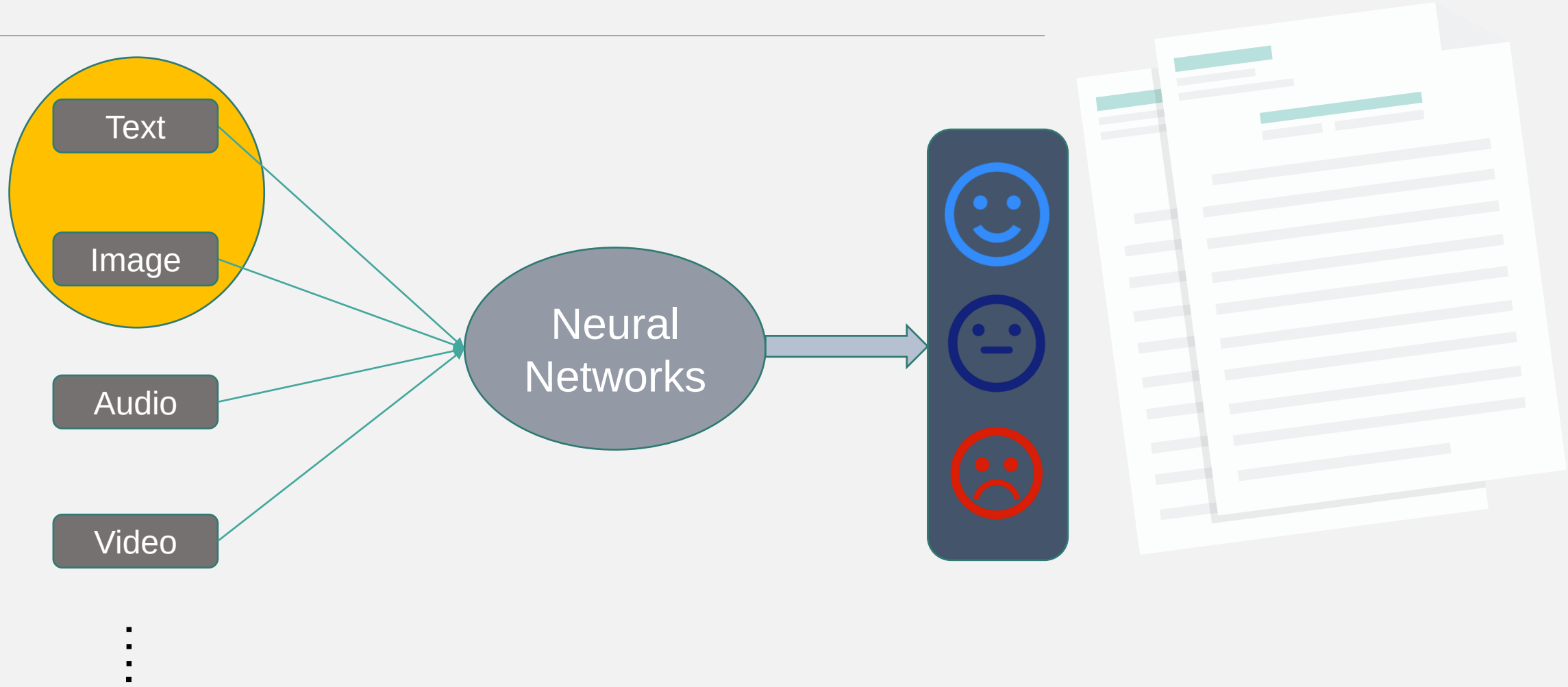

Kobe Bryant @kobebryant · 7月19日
#LegacyandtheQueen is dropping September 3rd and is finally available for pre-order! bit.ly/legacyandthequ...
翻译推文
"THE GREATEST PLAYERS PLAY WELL BECAUSE THEY'RE AFRAID OF WHAT MIGHT HAPPEN IF THEY DON'T. THEY PLAY WELL BECAUSE IF THEY FAIL, THEY'LL LOSE EVERYTHING THAT MAKES THEM THEMSELVES. BECAUSE ALL THEY ARE IS CHAMPIONS. IF THEY LOSE, THEY'RE NOTHING, NOT EVEN THEMSELVES."
- LEGACY AND THE QUEEN
Granity
51
624
3.2千


jaychou · 关注
ラフォーレ原宿
1 分钟 回复
yaojiayoudexiaok 和朋友一起喝奶茶很好啊
1 分钟 回复
xiaowen.huang.330 dun ~ dun ~ dun ~
1 分钟 回复
236,904 次赞
19 小时前
登录 即可点赞或评论。


京八婆
#斯科特拒绝和孙杨合影#这事儿闹的越来越大，孙杨也是怒了。最好的回应就是
们脸！孙杨加油


中山大学
7月8日 12:02 来自 康乐园里的iPhone客户端
#中大分享# 毕业快乐，想把中大唱给你听！
凤凰花开又一季
云山沧海赋别离
康园故事 最初的回忆是否还清晰
临行临别 也想再爱一次这片土地
这首《毕业组曲》... 展开全文

Introduce



Case Analysis

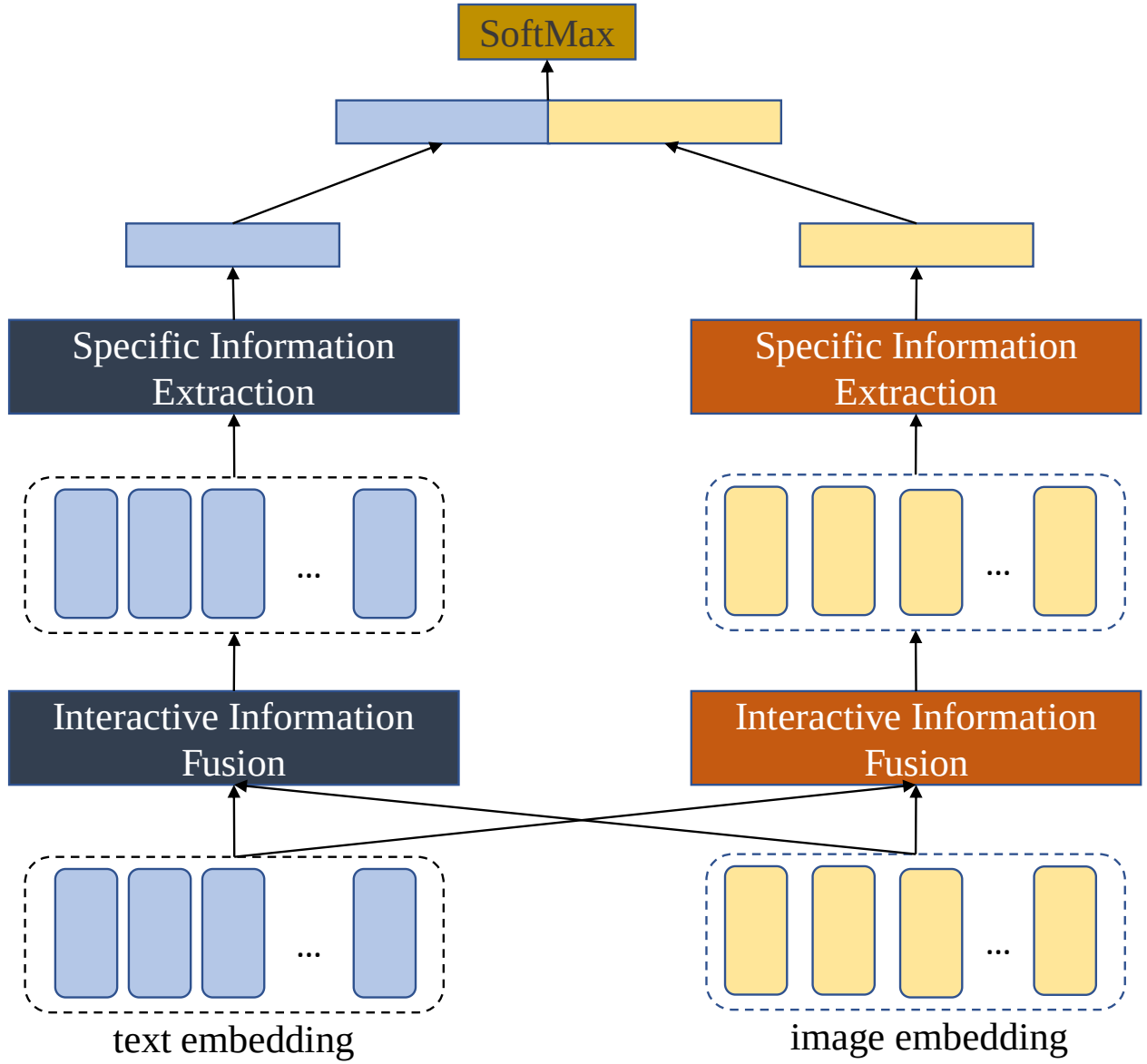
RT @OscarRomeo1268: Only 1 serious injury from #RTC on the #A64 with a few broken bones but talking. Other 3 walking wounded #incredible.

Difficulties

- Information fusion between multimodal data
- Information extraction of each modal data

The architecture of FENet

FENet is composed of interactive information fusion(IIF) layer and specific information extraction(SIE) layer.



Interactive information fusion layer

$$G_i = tanh(W_g[S_i : F_i] + b_g)$$

$$G = IIF(S, A)$$

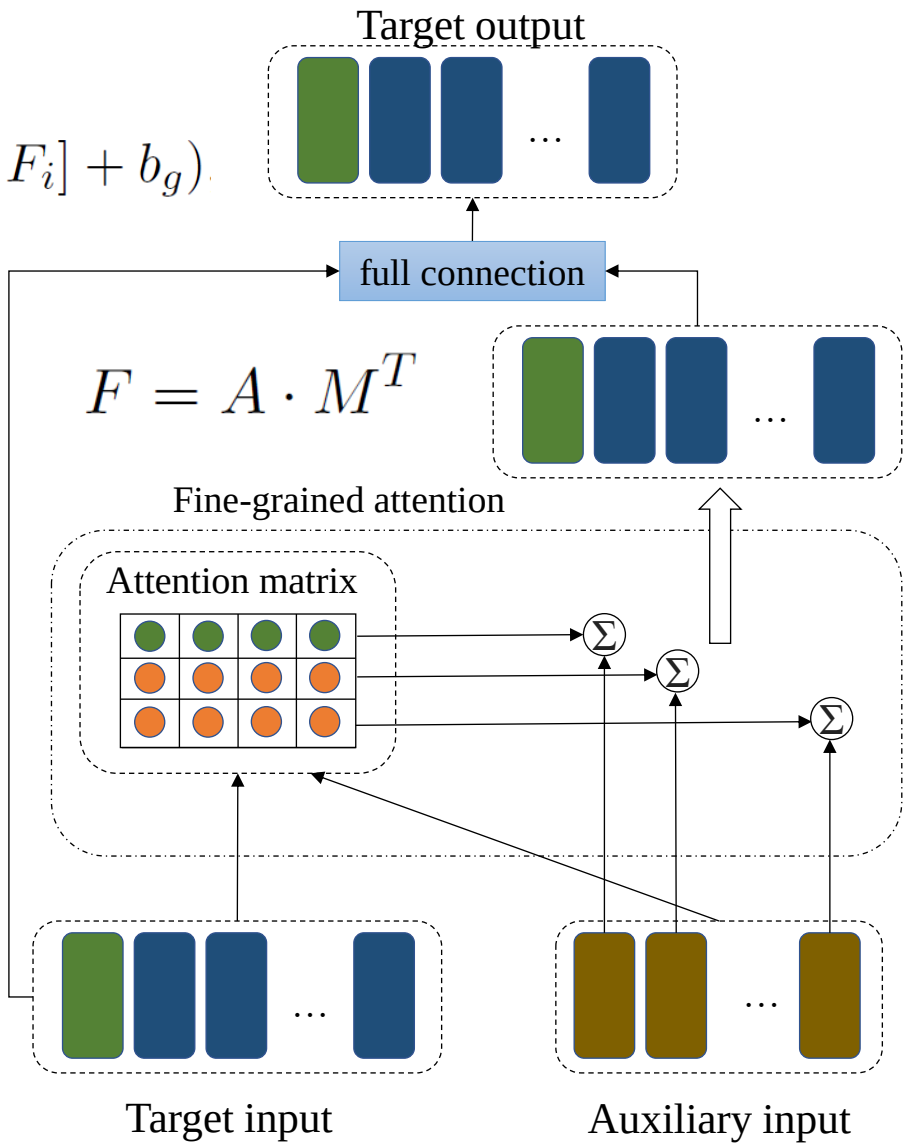
$$F = A \cdot M^T$$

$$M_{ij} = {S_{emb_i}}^T A_{emb_j}$$

$$M_{ij} = \frac{exp(M_{ij})}{\sum_{j=1}^{l} exp(M_{ij})}$$

$$S_{emb_i} = tanh(W_{S_{emb}} S_i + b_{S_{emb}})$$

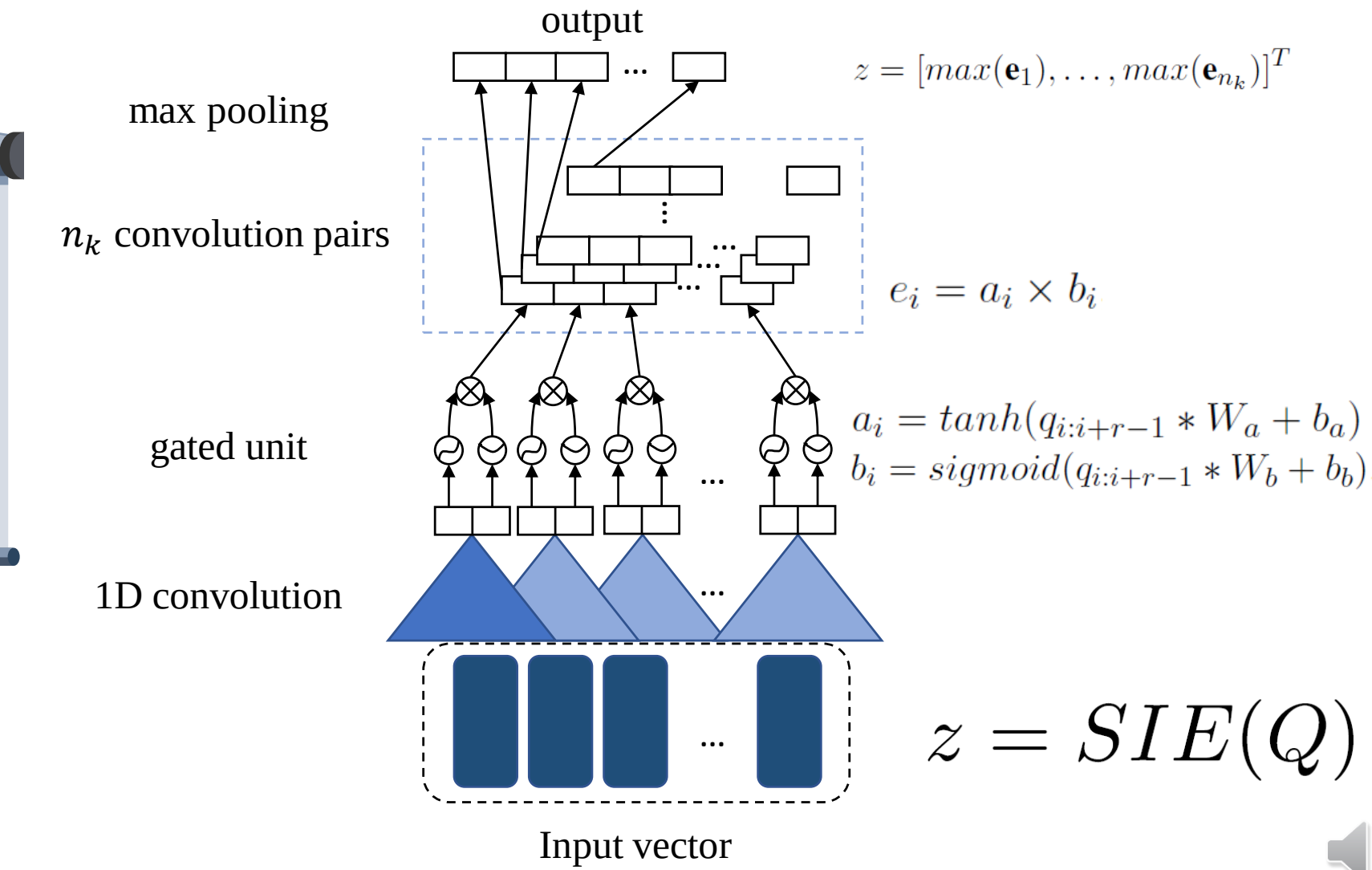
$$A_{emb_i} = tanh(W_{A_{emb}} A_i + b_{A_{emb}})$$



1. Calculate the attention matrix of the target input to the auxiliary input.
2. Multiply the auxiliary input with the obtained attention matrix to obtain the auxiliary output.
3. Auxiliary output and target input are fused through full connection.

Specific information extraction layer

1. Feed the fused representation into n convolutions and corresponding gated units.
2. Use max pooling to get the most expressive n-dimensional representation.



$$z = SIE(Q)$$

Experiment Results

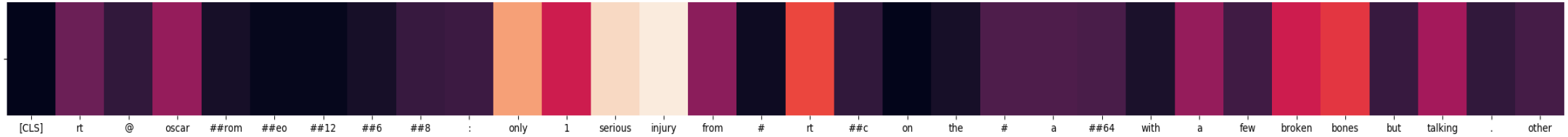
	Model	MVSA-Single		MVSA-Multiple	
		ACC	F1	ACC	F1
Baselines	SentiBank & SentiStrength	0.5205	0.5008	0.6562	0.5536
	CNN-Multi	0.6120	0.5837	0.6630	0.6419
	DNN-LR	0.6142	0.6103	0.6786	0.6633
	MultiSentiNet	0.6984	0.6963	0.6886	0.6811
	CoMN(6)	0.7051	0.7001	0.6892	0.6883
Ablated FENet	FENet w/o IIF	0.6920	0.6882	0.6837	0.6795
	FENet w/o SIE	0.7120	0.7102	0.6989	0.6964
FENet variants	FENet-Glove	**0.7254**	**0.7232**	**0.7057**	**0.7038**
	FENet-BERT	**0.7421**	**0.7406**	**0.7146**	**0.7121**

visual and textual attention visualizations.

Visual attention

RT @OscarRomeo1268: Only 1 serious injury from #RTC on the #A64 with a few broken bones but talking. Other 3 walking wounded #incredible

Textual attention



Main contributions

1. We introduce an Interactive Information Fusion (IIF) mechanism to learn fine-grained fusion features. IIF is based on cross-modality attention mechanisms, aiming to generate the visual-specific textual representation and the textual-specific visual representation for both two modality contents.

2. We propose a specific Information Extraction (SIE) mechanism to extract the informative features for textual and visual information, and leverage the extracted visual and textual information for sentiment prediction. To the best of our knowledge, no CNN-gated extraction mechanism for both textual and visual information has been proposed in the field of multimodal sentiment analysis so far.

thanks