



G2SAM: Graph-Based Global Semantic Awareness Method for Multimodal Sarcasm Detection

Yiwei Wei^{*1,5}, Shaozu Yuan^{*†2}, Hengyang Zhou⁵, Longbiao Wang^{‡1,4},
Zhiling Yan², Ruosong Yang², Meng Chen³

¹ Tianjin Key Laboratory of Cognitive Computing and Application,
College of Intelligence and Computing, Tianjin University, Tianjin, China

² JD AI Research, Beijing, China ³ Yep AI, Melbourne, Australia

⁴ Huiyan Technology (Tianjin) Co., Ltd, Tianjin, China

⁵ China University of Petroleum(Beijing) at Karamay, Karamay, China

Code: <https://github.com/upccpu/G2SAM>

— AAAI 2024

2024. 2. 22 • ChongQing



gesis
Leibniz-Institut
für Sozialwissenschaften



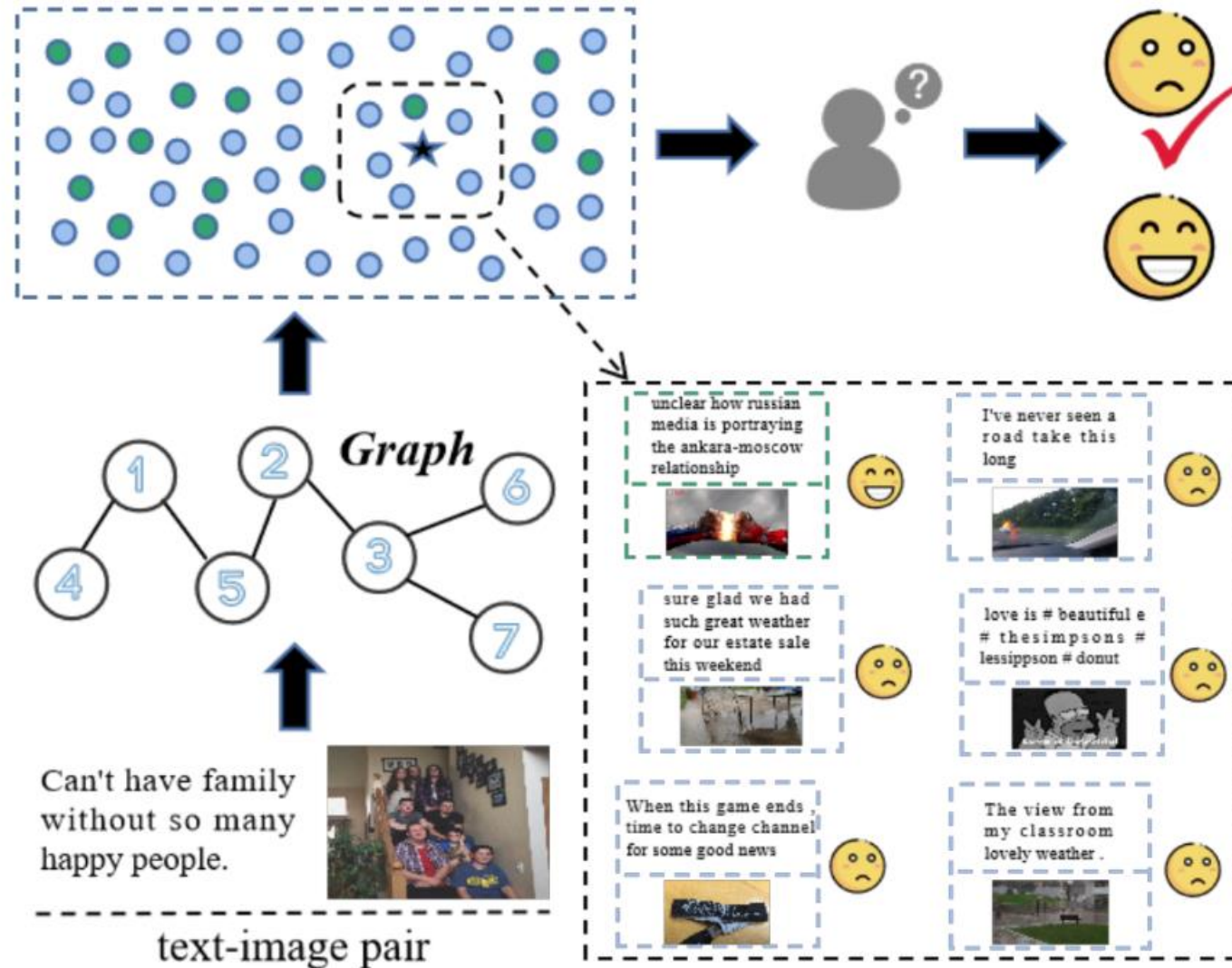
Reported by Renhui Luo



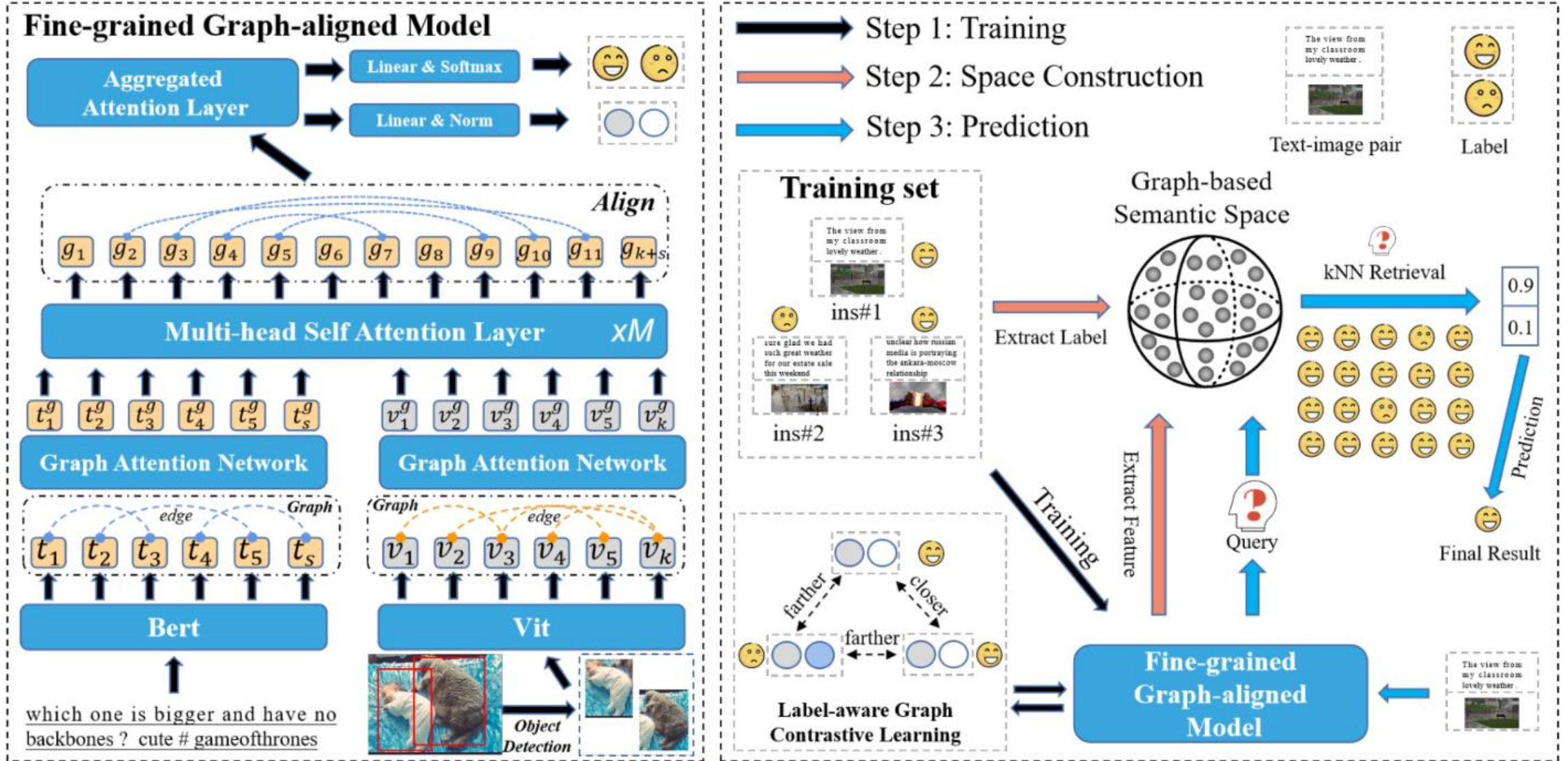
- 1.Introduction**
- 2.Overview**
- 3.Methods**
- 4.Experiments**



Introduction



Overview



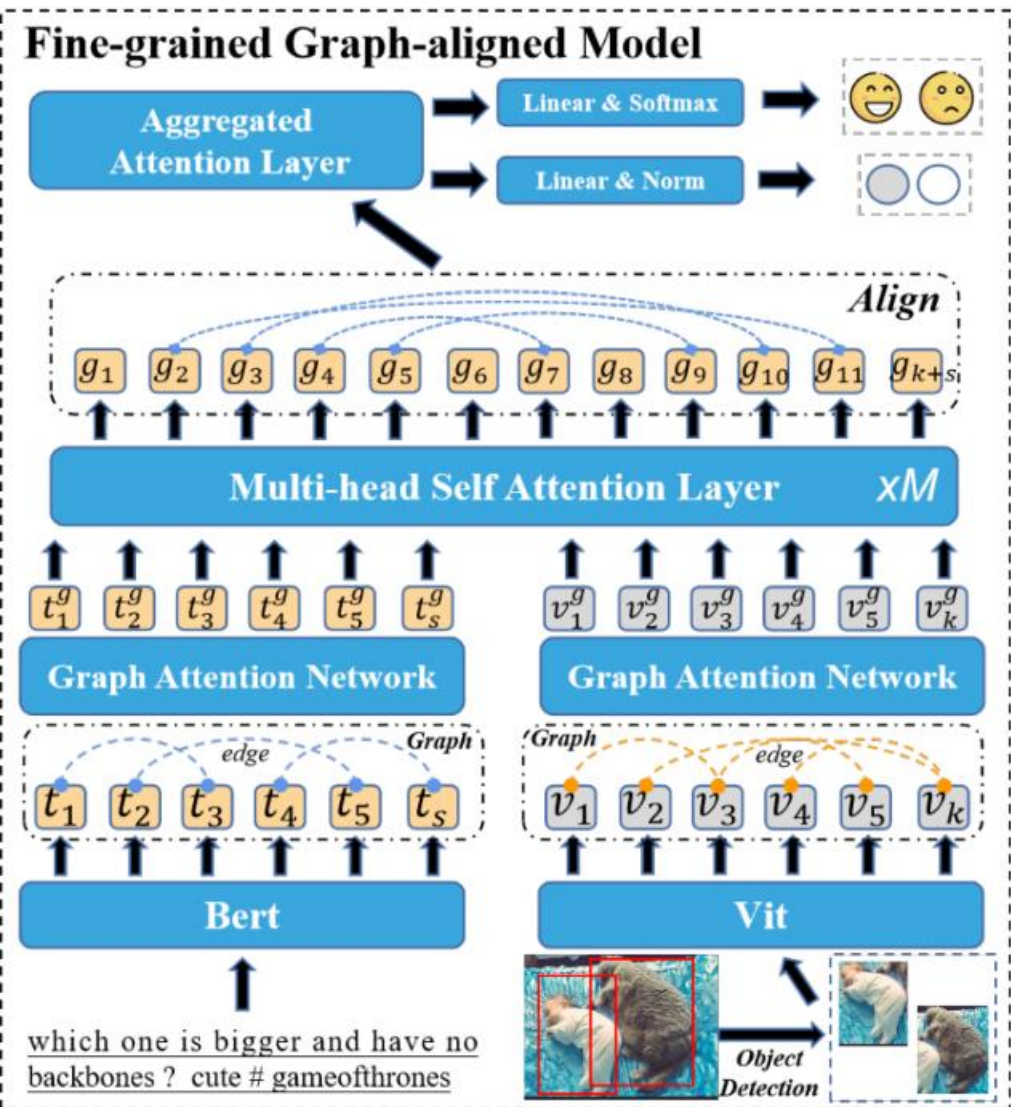
Method

$$G = \text{softmax}\left(\frac{(G^{[v,t]}W_q)^T}{\sqrt{d}}(G^{[v,t]}W_k)\right)(G^{[v,t]}W_v) \quad (1)$$

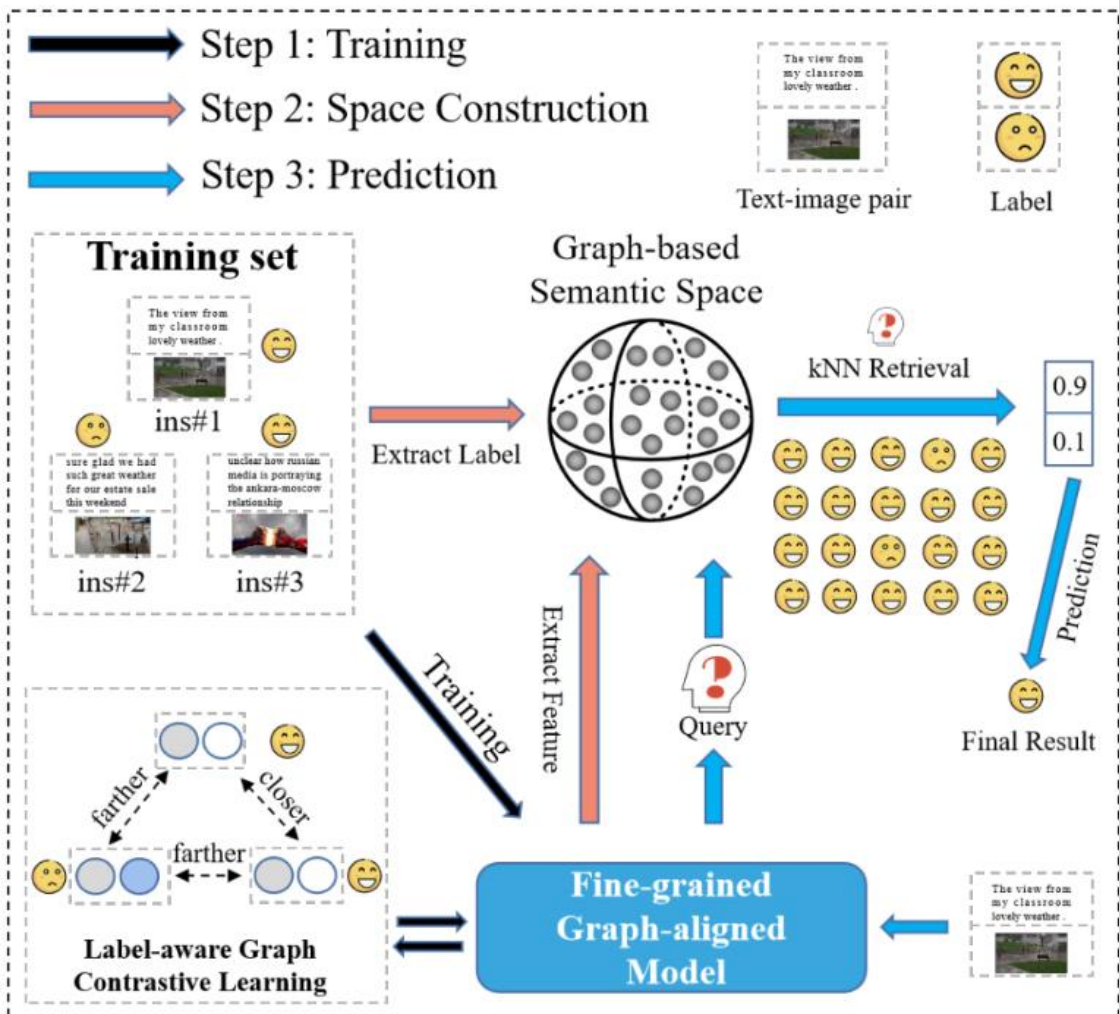
$$\tilde{r}_i = \text{GELU}(g_iW_1 + b_1)W_2 + b_2 \quad (2)$$

$$\tilde{q} = \sum_{i=1}^{k+s} \exp\left(\frac{\tilde{r}_i}{\sum_{j=1}^{k+s} \tilde{r}_j}\right)(g_i) \quad (3)$$

$$q = \text{GELU}(\tilde{q}W_3 + b_3) \quad (4)$$



Method



$$\hat{y}_{kNN} = \sum_{i=1}^k \alpha_i y_i, \quad \alpha_i = \frac{e^{-\|q_i - q\|_2^2 / \tau}}{\sum_j e^{-\|q_j - q\|_2^2 / \tau}} \quad (5)$$

$$\begin{cases} \hat{y} = 1, & \text{if } \hat{y}_{kNN} \geq 0.5 \\ \hat{y} = 0, & \text{if } \hat{y}_{kNN} < 0.5 \end{cases} \quad (6)$$

$$L_{ce} = \text{CrossEntropy}(\text{GELU}(qW_{ce} + b_{ce})) \quad (7)$$

$$L = L_{ce} + \gamma L_{LGCL} \quad (8)$$



Experiments

Dataset	Label	Train	Val	Test
HFM	Positive	8642	959	959
	Negative	11174	1451	1450
	All	19816	2410	2409

Table 1: Statistics of the experimental data.



Experiments

MODALITY	METHOD	Acc(%)	Pre(%)	Rec(%)	F1(%)	Macro-average		
						Pre(%)	Rec(%)	F1(%)
image	Resnet	64.76	54.41	70.80	61.53	60.12	73.08	65.97
	ViT	67.83	57.93	70.07	63.43	65.68	71.35	68.40
text	Bi-LSTM	81.90	76.66	78.42	77.53	80.97	80.13	80.55
	BERT	83.85	78.72	82.27	80.22	81.31	80.87	81.09
image+text	HFM	83.44	76.57	84.15	80.18	79.40	82.45	80.90
	Res-BERT	84.80	77.80	84.15	80.85	78.87	84.46	81.57
	Att-BERT	86.05	78.63	83.31	80.90	80.87	85.08	82.92
	InCrossMGs*	86.10	81.38	84.36	82.84	85.39	85.80	85.60
	CMGCN*	86.54	-	-	82.73	-	-	-
	HKEmodel*	87.36	81.84	86.48	84.09	-	-	-
	MILNet* [†]	88.72	84.97	87.79	86.37	87.75	88.29	88.04
	DIP	89.59	87.76	86.58	87.17	88.46	89.13	89.01
Ours*	90.48	87.95	89.02	88.48	89.44	89.79	89.65	



Experiments

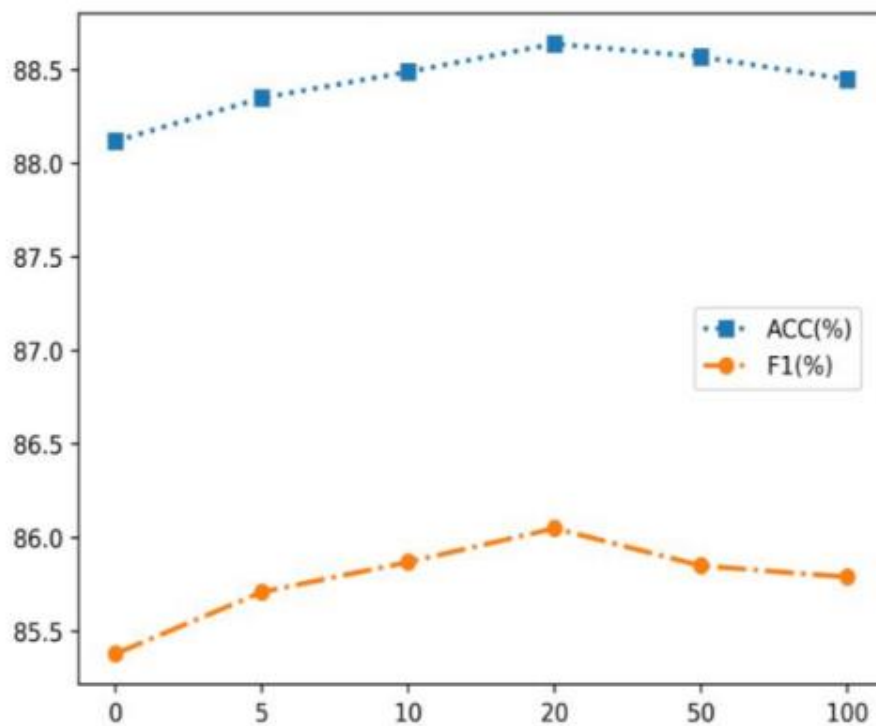
Model	ACC(%)	Pre(%)	Rec(%)	F1(%)
MILNet	88.72	84.97	87.79	86.37
FGM	89.01	85.73	87.15	86.43
FGM+kNN	89.86	86.82	87.89	87.35
FGM+LGCL	89.33	86.29	87.56	86.92
FGM+kNN+LGCL	90.48	87.95	89.02	88.48



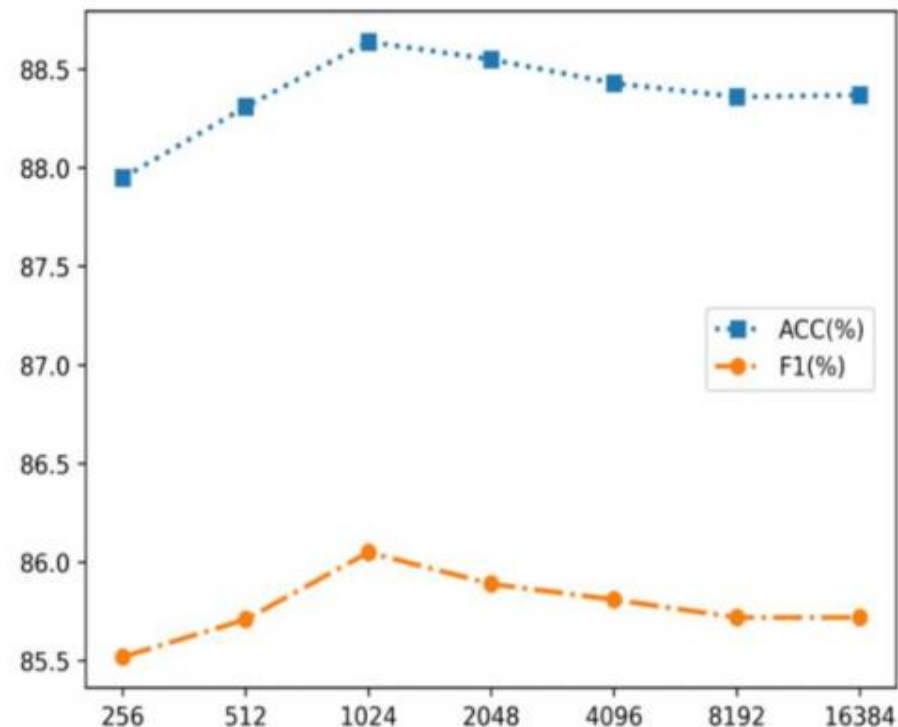
Experiments

Model	ACC(%)	Pre(%)	Rec(%)	F1(%)
ViT	67.25	57.48	69.93	63.10
G ² SAM(ViT)	68.75	58.20	70.59	63.79
BERT	84.74	79.27	83.52	81.34
G ² SAM(BERT)	85.65	80.09	84.31	82.15
DIP	89.59	87.76	86.58	87.17
G ² SAM(DIP)	90.21	88.01	87.45	87.73

Experiments






(a) Performance of the number(k) of neighbors

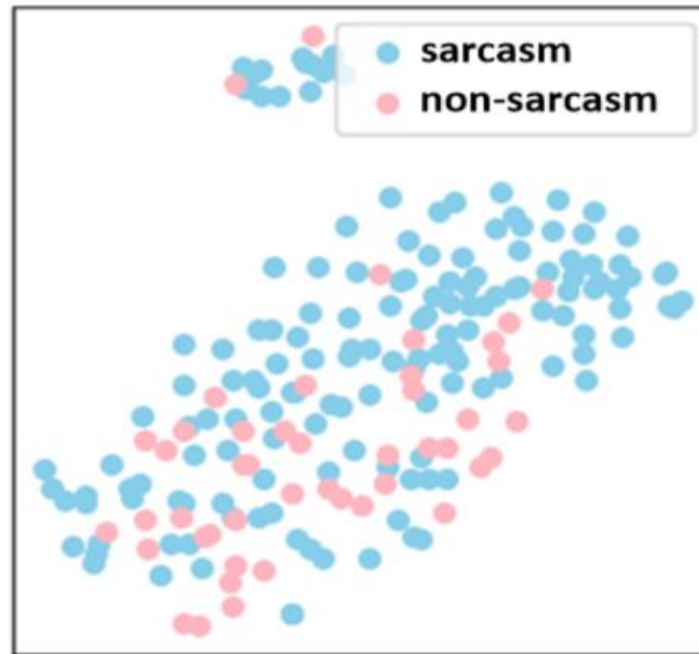


(b) Performance of the size(N) of semantic space

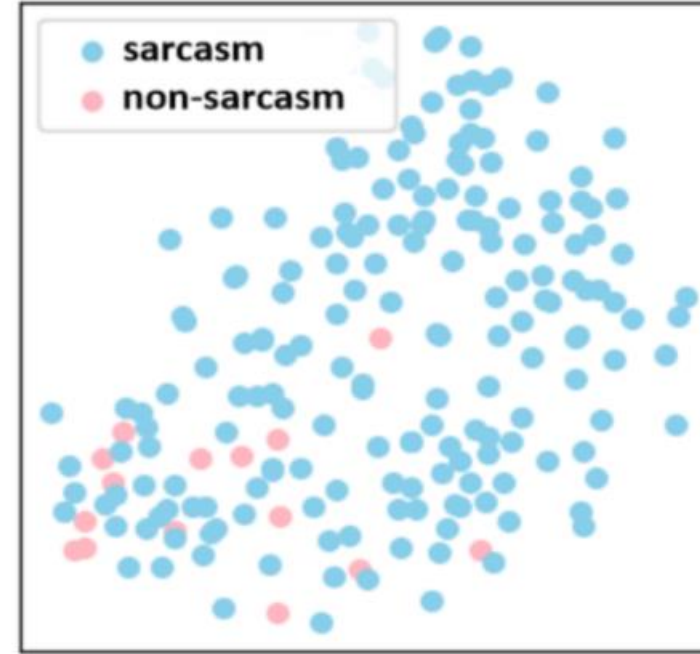
Experiments

Text-Image Pairs	kNN Prediction						GT
	Instance	label	α	Instance	label	α	
 Can't have family without so many happy people.	ins#1	1	0.205	ins#6	1	0.069	Sarcasm 1
	ins#2	1	0.102	ins#7	1	0.065	
	ins#3	1	0.087	ins#8	1	0.065	
	ins#4	1	0.082	ins#9	0	0.063	
	ins#5	1	0.072	ins#10	1	0.057	
 Happy fourth birthday to one of the cutest and happiest kids.	ins#1	0	0.163	ins#6	1	0.068	Non-sarcasm 0
	ins#2	0	0.125	ins#7	0	0.063	
	ins#3	0	0.107	ins#8	0	0.059	
	ins#4	0	0.075	ins#9	0	0.051	
	ins#5	0	0.073	ins#10	0	0.051	
 This is a man who obviously knows how to make good decisions.	ins#1	1	0.092	ins#6	1	0.071	Sarcasm 1
	ins#2	1	0.087	ins#7	0	0.071	
	ins#3	0	0.085	ins#8	1	0.068	
	ins#4	1	0.079	ins#9	1	0.065	
	ins#5	1	0.075	ins#10	0	0.064	

Experiments



(a) without LGCL



(b) with LGCL



Thanks!