南京航空航天大學
Nanjing University of Aeronautics and Astronautics

模式分析与机器智能
工业和信息化部重点实验室
MIIT Key Laboratory of
Pattern Analysis & Machine Intelligence

Multi-Label Supervised Contrastive Learning

Pingyue Zhang, Mengyue Wu*

MoE Key Lab of Artificial Intelligence, AI Institute
X-LANCE Lab, Department of Computer Science and Engineering
Shanghai Jiao Tong University, Shanghai, China
{williamzhangsjtu, mengyuewu}@sjtu.edu.cn

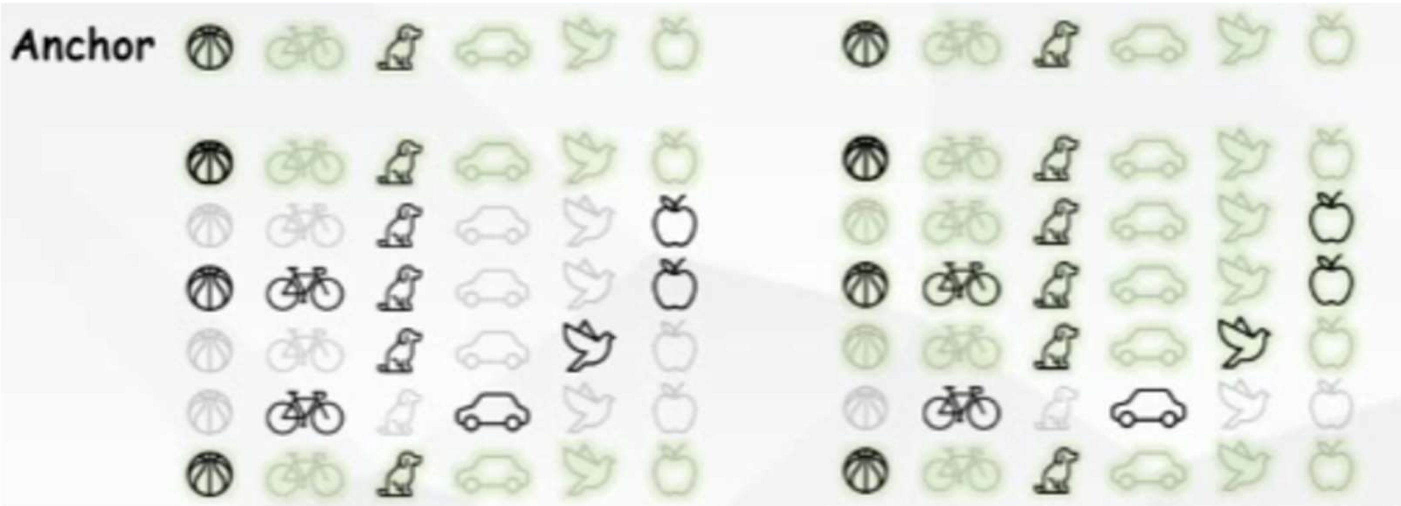
AAAI 2024

Introduction: Multi-label

- SupCon:
  - ➢ same-label as positives
  - ➢ succeed in single-label classification
- Challenge: how to find positives

Introduction: Apply SupCon – All and Any

- All : Exactly same label
- Any : Overlapped label



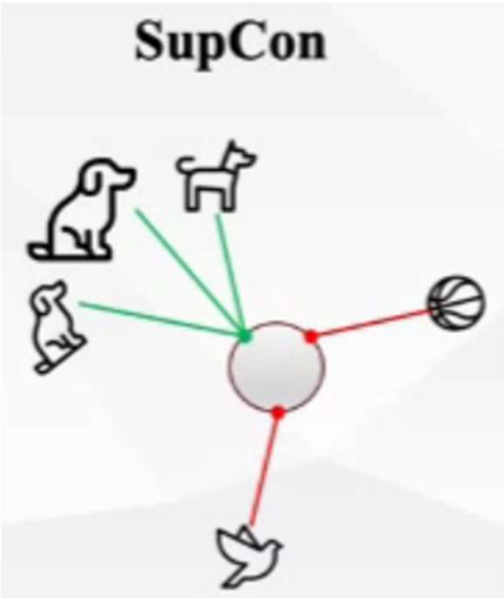
$$\nabla \approx \frac{z^{(1)} + z^{(6)}}{2}$$

$$\nabla \approx \frac{z^{(1)} + z^{(2)} + z^{(3)} + z^{(4)} + z^{(6)}}{5}$$

Method: SupCon

- SupCon : $\mathcal{L}^{(i)}_{\text{supcon}} = \frac{-1}{|\mathcal{P}^{(i)}|} \sum_{p \in \mathcal{P}^{(i)}} \log \frac{e^{s_p^{(i)}/\tau}}{\sum_{a \in \mathcal{A}^{(i)}} e^{s_a^{(i)}/\tau}}$ $\quad s_j^{(i)} = \boldsymbol{z}_q^{(i)} \cdot \boldsymbol{z}_k^{(j)}$

- Pushing towards "class" representation

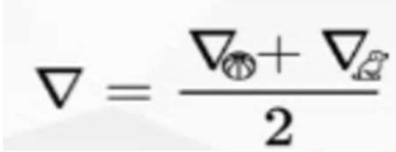


$$\nabla_{\boldsymbol{z}_q^{(i)}} \mathcal{L}^{(i)}_{\text{supcon}} = \bar{\boldsymbol{z}} + \hat{\boldsymbol{z}}$$

$$\bar{\boldsymbol{z}} = \frac{-1}{\tau} \cdot \frac{1}{|\mathcal{P}^{(i)}|} \sum_{p \in \mathcal{P}^{(i)}} \boldsymbol{z}_k^{(p)}$$

$$\hat{\boldsymbol{z}} = \sum_{a \in \mathcal{A}^{(i)}} \frac{1}{\tau} \cdot \frac{e^{s_a^{(i)}/\tau}}{C} \cdot \boldsymbol{z}_k^{(a)}$$

Method: MulSupCon



$$
\begin{aligned}
\nabla_{\boldsymbol{z}_q^{(i)}}\mathcal{L}^{(i)} &= \sum_{y_j^{(i)}\in\boldsymbol{y}^{(i)}} \frac{-1}{\tau}\frac{1}{|\mathcal{P}_j^{(i)}|}\sum_{p\in\mathcal{P}_j^{(i)}} \boldsymbol{z}_k^{(p)} \\
&\quad + |\boldsymbol{y}^{(i)}| \sum_{a\in\mathcal{A}^{(i)}} \frac{1}{\tau}\cdot\frac{e^{s_a^{(i)}/\tau}}{C}\cdot \boldsymbol{z}_k^{(a)} \\
&= \sum_{y_j^{(i)}\in\boldsymbol{y}^{(i)}} \bar{\boldsymbol{z}}_j + |\boldsymbol{y}^{(i)}|\cdot\hat{\boldsymbol{z}},
\end{aligned}
$$

$$
\bar{\boldsymbol{z}}_j = \frac{-1}{\tau}\frac{1}{|\mathcal{P}_j^{(i)}|}\sum_{p\in\mathcal{P}_j^{(i)}} \boldsymbol{z}_k^{(p)}
$$

- MulSupcon:

$$
\mathcal{L}^{(i)} = \frac{1}{|\boldsymbol{y}^{(i)}|}\sum_{y_j^{(i)}\in\boldsymbol{y}^{(i)}} \frac{-1}{|\mathcal{P}_j^{(i)}|}\sum_{p\in\mathcal{P}_j^{(i)}} \log\frac{e^{s_p^{(i)}/\tau}}{\sum_{a\in\mathcal{A}^{(i)}} e^{s_a^{(i)}/\tau}}
$$

Positive set for each label j:

$$
\mathcal{P}_j^{(i)} = \{m \mid \forall m, y_j^{(i)} \in \boldsymbol{y}^{(m)}\}
$$

- Pushing towards average “class” representation

Experimental results

Metric	example-F1					
Dataset	media	yeast	scene	bkms	deli	nus
LaMP	-	0.624	0.728	0.389	0.372	0.376
MPVAE	-	0.648	0.751	0.382	0.373	0.468
ASL	-	0.613	0.770	0.373	0.359	0.468
RBCC	-	0.605	0.758	-	-	0.466
C-GMVAE	0.623†	0.656	0.777	0.392	0.381	0.481
Ours	**0.627**	**0.659**	**0.787**	**0.394**	**0.386**	**0.484**

Table 2: Comparison with existing approaches, The example-F1 is shown, † means we run the official C-GMVAE codes with our data splits

Experimental results

Method	mAP	Hamming Accuracy	example-F1	Macro-F1	Micro-F1	precision@1
ALL	0.636	0.979	0.676	0.607	0.664	0.897
ANY	0.564	0.977	0.639	0.547	0.623	0.883
Ours	**0.672**	**0.980**	**0.700**	**0.636**	**0.688**	**0.916**

Table 5: Linear probe results of ALL, ANY, and MulSupCon (ours) method on MS-COCO datase

Experimental results

Dataset	Method	mAP	HA	example-F1	Macro-F1	Micro-F1	precision@1
MS-COCO	BCE	0.694	0.981	0.724	0.654	0.706	0.921
	linear	0.672	0.980	0.700	0.636	0.688	0.916
	finetune	**0.708**	**0.982**	**0.736**	**0.662**	**0.714**	**0.930**
NUS-WIDE	BCE	0.524	0.984	0.707	0.505	0.722	0.821
	linear	0.552	0.984	0.701	0.553	0.721	0.822
	finetune	**0.566**	0.984	**0.714**	**0.557**	**0.728**	**0.825**
Objects365	BCE	0.180	0.983	0.361	0.133	0.432	0.716
	linear	0.297	0.984	0.441	0.288	0.479	0.784
	finetune	**0.333**	**0.984**	**0.467**	**0.311**	**0.509**	**0.802**
MIRFLICKR	BCE	0.696	0.921	0.710	0.641	0.731	0.889
	linear	0.694	0.920	0.705	0.637	0.727	0.888
	finetune	**0.711**	**0.926**	**0.729**	**0.658**	**0.754**	**0.895**
PASCAL	BCE	0.673	0.962	0.684	0.628	0.692	0.769
	linear	0.694	0.962	0.680	0.644	0.697	0.778
	finetune	**0.726**	**0.965**	**0.716**	**0.670**	**0.726**	**0.802**

Table 6: Comparison of linear probing and finetuning results with BCE on 5 image datasets

Experimental results

Dataset	Method	mAP	HA	example-F1	Macro-F1	Micro-F1	precision@1
(a) Model pretrained on MS-COCO							
PASCAL	BCE	**0.865**	**0.979**	**0.838**	**0.809**	**0.842**	**0.916**
	Ours	0.853	0.977	0.818	0.792	0.824	0.907
MIRFLICKR	BCE	0.656	0.911	0.673	0.613	0.700	0.873
	Ours	**0.722**	**0.923**	**0.716**	**0.668**	**0.741**	**0.899**
(b) Model pretrained on NUS-WIDE							
PASCAL	BCE	0.661	0.964	0.713	0.624	0.713	0.801
	Ours	**0.750**	**0.969**	**0.749**	**0.693**	**0.760**	**0.853**
MIRFLICKR	BCE	0.716	0.922	0.716	0.662	0.740	0.897
	Ours	**0.750**	**0.927**	**0.736**	**0.691**	**0.758**	**0.906**

Table 7: Transfer to PASCAL and MIRFLICKR datasets.

Experimental results

Linear probing results with and without the weight in MulSupCon							
Dataset	Weight	mAP	HA	example-F1	Macro-F1	Micro-F1	precision@1
MS-COCO	Yes	0.663	0.980	0.697	0.630	0.683	0.914
	No	**0.672**	0.980	**0.700**	**0.636**	**0.688**	**0.916**
PASCAL	Yes	0.693	0.962	0.676	0.640	0.695	0.772
	No	**0.694**	0.962	**0.680**	**0.644**	**0.697**	**0.778**
MIRFLICKR	Yes	0.689	0.919	0.703	0.634	0.725	0.884
	No	**0.694**	**0.920**	**0.705**	**0.637**	**0.727**	**0.888**

Table 8: Ablation study results

Conclusion

- MulSupCon achieves a competitive results when compared to SOTA.
- MulSupCon outperforms BCE to a large extent.
- MulSupCon can generalize well.
- Ablation studies to demonstrate the effectiveness of the weights.

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

模式分析与机器智能
工业和信息化部重点实验室
MIIT Key Laboratory of
Pattern Analysis & Machine Intelligence

THANKS