Collapse-Aware Triplet Decoupling for Adversarially Robust Image Retrieval

Qiwei Tian

School of Cybersecurity and Engineering

Xi’an JiaoTong University

July, 2024

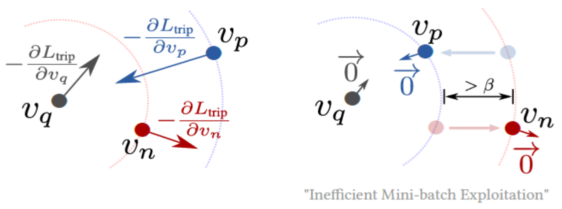
Outline

1. Research Background
2. Our Method
3. Experiment & Conclusion

Research Background

➢ Current Adversarial Defense in Deep Metric Learning

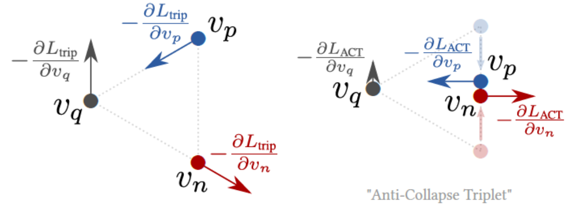
Adversarial Training(AT)



(a) Before Attack (b) After Attack

Embedding-Shifted Triplet (EST)

- **Unguided perturbation**
- **Inefficient AT**

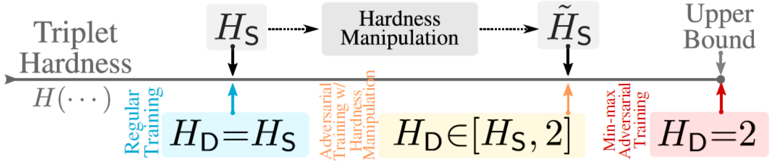


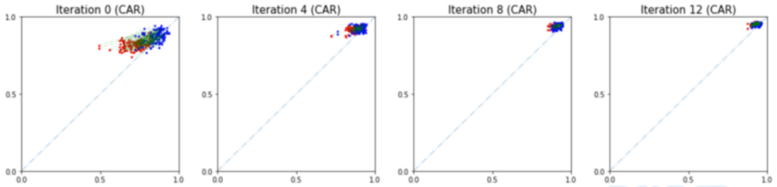
(a) Before Attack (b) After Attack

Anti-Collapse Triplet (ACT)

- **Pull pos and neg together**
- **Disastrously destroy embeddings**

Hardness Manipulation (HM)





➢ Neglect the triplet structure in DML- **Weak Adversary**
➢ DML cannot handle excessively hard triplets- **Model Collapse**

[1] Zhou, M., Niu, Z., Wang, L., Zhang, Q., & Hua, G. (2020). Adversarial ranking attack and defense. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16 (pp. 781-799).
[2] Zhou, M., Wang, L., Niu, Z., Zhang, Q., Zheng, N., and Hua, G. Adversarial attack and defense in deep ranking. CoRR, abs/2106.03614, 2021b. URL https://arxiv.org/ abs/2106.03614.
[3] Zhou, M. and Patel, V. M. Enhancing adversarial robustness for deep metric learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15325–15334, 2022.

Outline

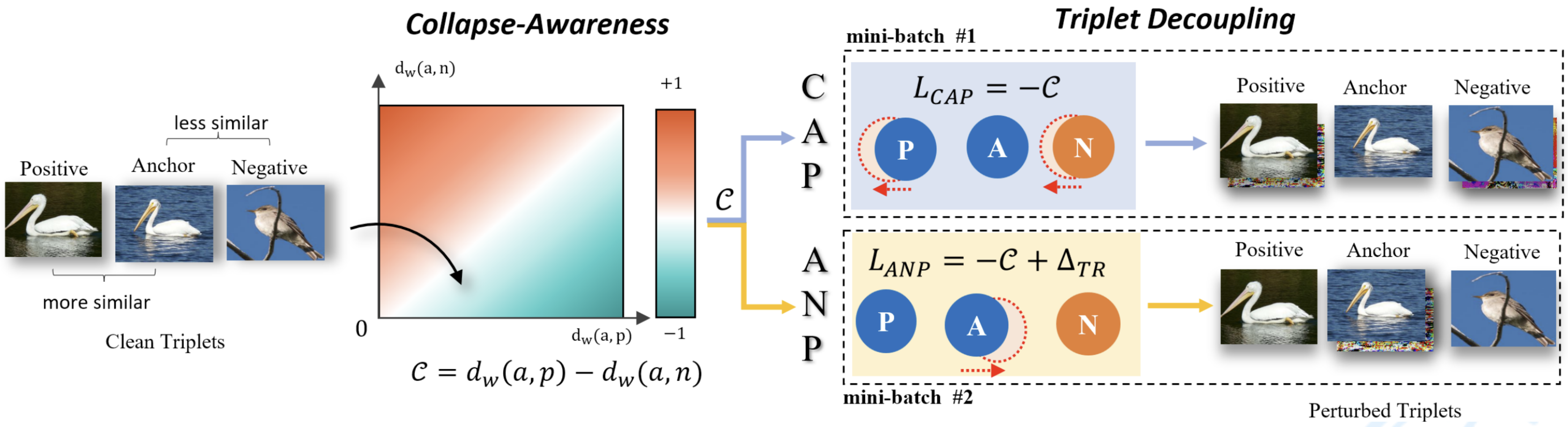
1. Research Background

2. **Our Method**

3. Experiment & Conclusion

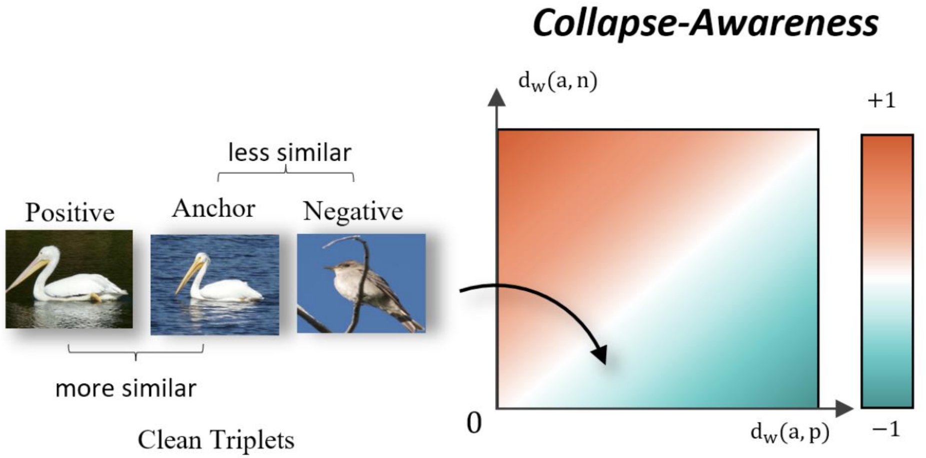
Our Method

➢ Collapse-Aware Triplet Decoupling (CA-TRIDE)



Our Method

➢ Collapse-Aware (CA) -> Model Collapse



$$\mathcal{C} = d_w(a,p) - d_w(a,n)$$

Current Method

$$H(\mathbf{A},\mathbf{P},\mathbf{N}) = d(\mathbf{A},\mathbf{P}) - d(\mathbf{A},\mathbf{N}), \quad H \in [-2, 2]$$

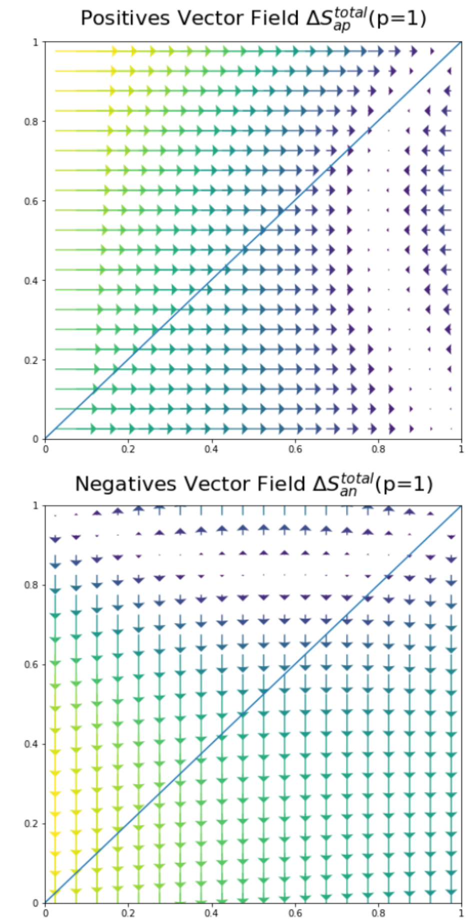
Our Method

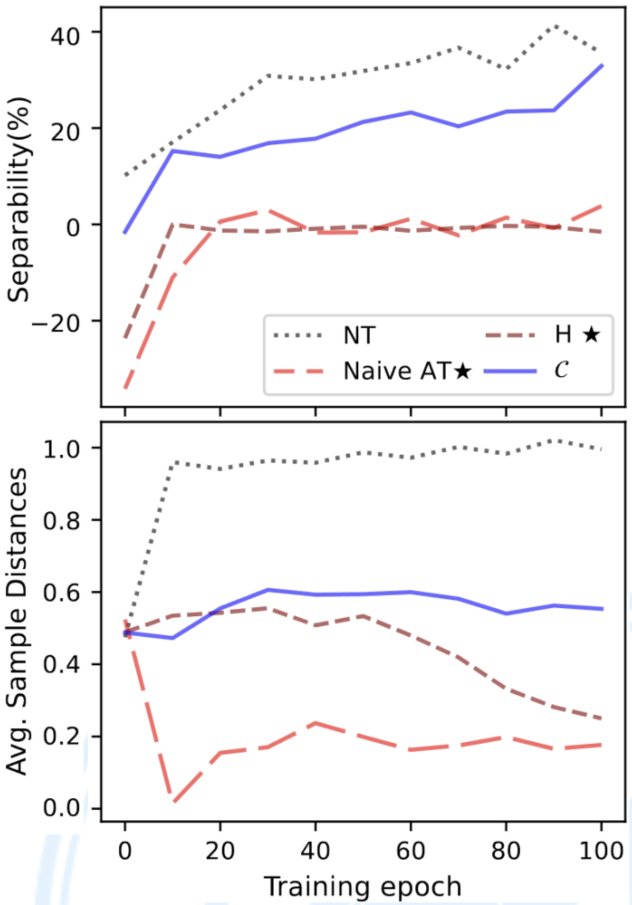
$$\mathcal{C}(\mathbf{A},\mathbf{P},\mathbf{N}) = d_\omega(\mathbf{A},\mathbf{P}) - d_\omega(\mathbf{A},\mathbf{N})$$

$$d_\omega(\mathbf{A},\mathbf{P}) = \frac{\sum_i^{\mathbf{A},\mathbf{P}} (w_{p_i} \cdot d(a_i,p_i))}{\sum_i^{\mathbf{P}} w_{p_i}}$$

$$w_{p_i} = exp\Big(- \lambda \big(d(a_i,p_i) - \min_{\forall a_i \in \mathbf{A}, p_i \in \mathbf{P}} d(a_i,p_i) \big) \Big)$$

$$\arg\max_{\delta} \mathcal{C}(\tilde{\mathbf{A}}, \tilde{\mathbf{P}}, \tilde{\mathbf{N}})$$



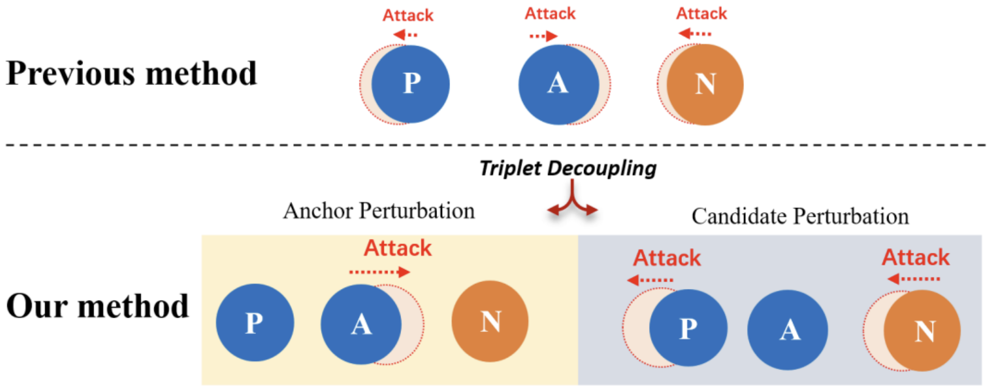


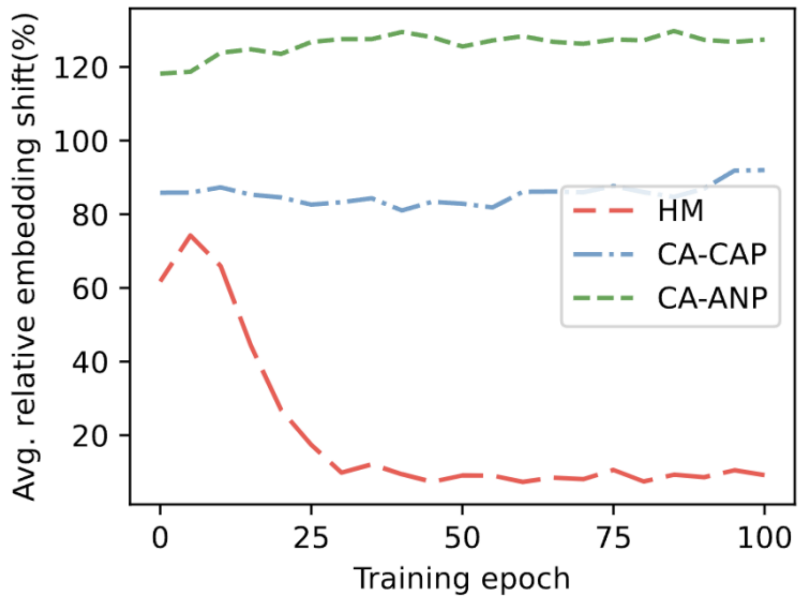
Xuan, H., Stylianou, A., Liu, X., and Pless, R. Hard negative examples are hard, but useful. In Vedaldi, A., Bischof, H., Brox, T., and Frahm, J.-M. (eds.), Computer Vision – ECCV 2020, pp. 126–142, Cham, 2020. Springer International Publishing. ISBN 978-3-030-58568-6.

Our Method

➢ **Triplet Decoupling (TRIDE)** -> Weak Adversary

Decouple current methods → Anchor Perturbation / Candidate Perturbation

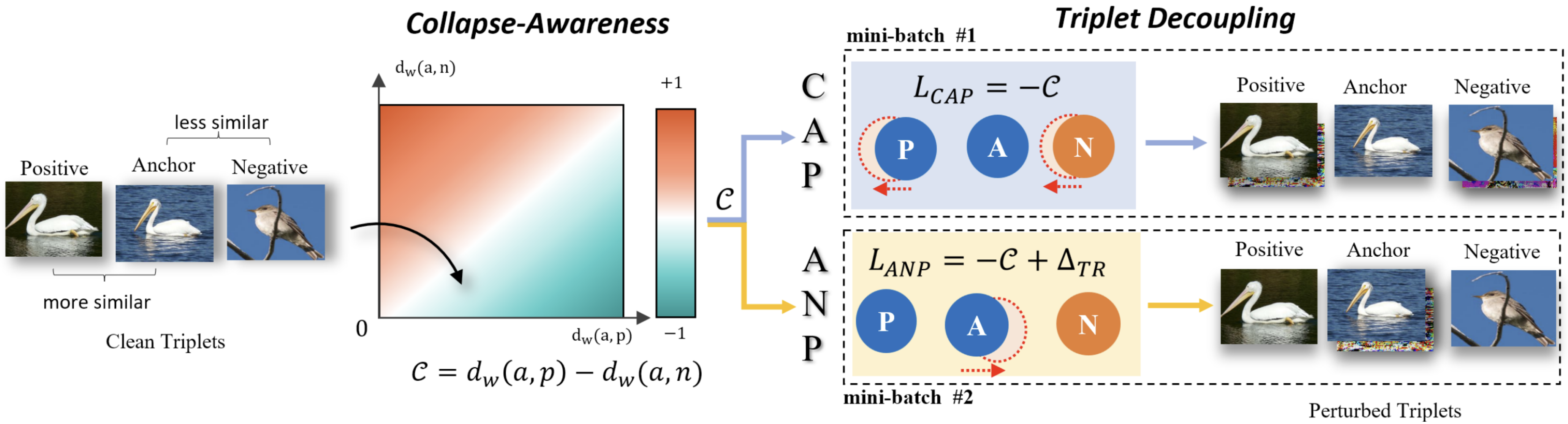


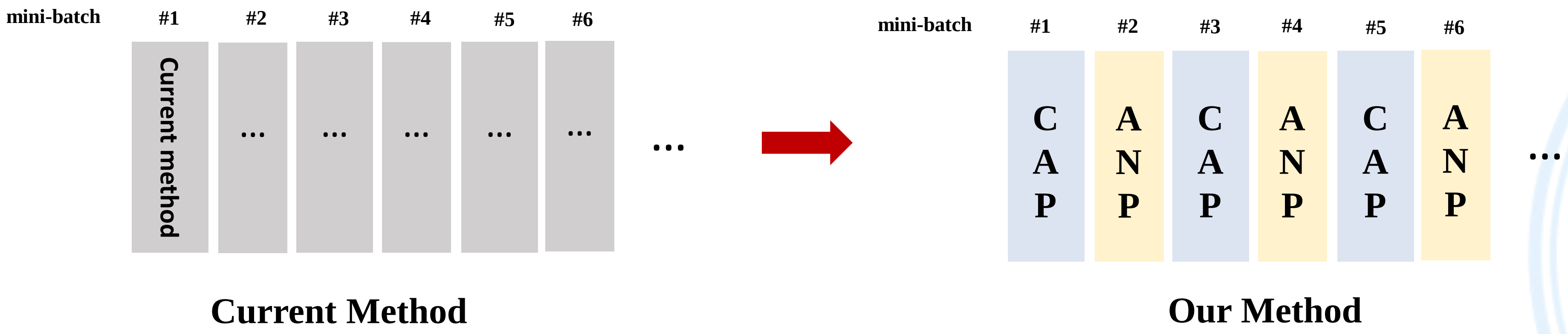


$$\arg\min_{\theta} L_{\mathcal{T}}(\tilde{\mathbf{A}}, \tilde{\mathbf{P}}, \tilde{\mathbf{N}}; \Theta) \Rightarrow \arg\min_{\Theta} \begin{cases} L_{\mathcal{T}}(\tilde{\mathbf{A}}, \mathbf{P}, \mathbf{N}; \Theta) + L_{TR}, & ANP \\ L_{\mathcal{T}}(\mathbf{A}, \tilde{\mathbf{P}}, \tilde{\mathbf{N}}; \Theta), & CAP \end{cases}$$

Our Method

➢ Collapse-Aware Triplet Decoupling (CA-TRIDE)





Outline

1. Research Background
2. Our Method
3. Experiment & Conclusion

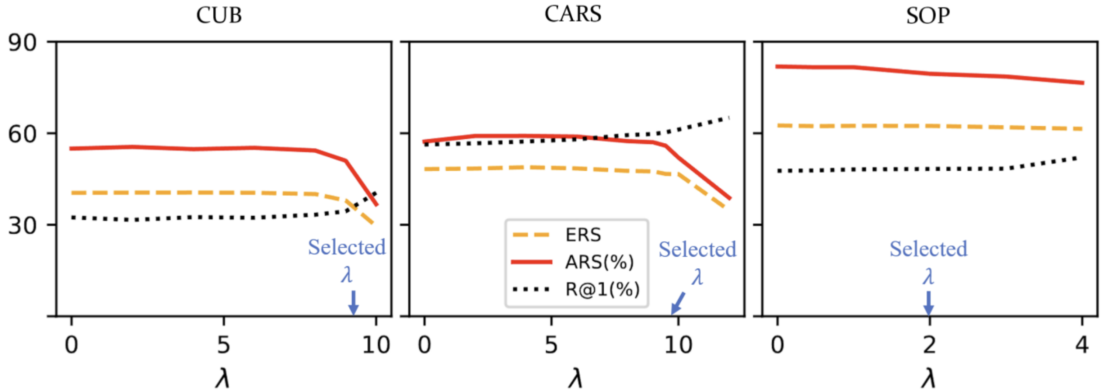
Experiments & Conclusions

Dataset	Defense Method	PGD steps	Benign Example Evaluation R@1 ↑	R@2 ↑	mAP ↑	NMI ↑	Adversarial Example Evaluation (%) CA+ ↑	CA- ↑	QA+ ↑	QA- ↑	ES:R ↑	LTM ↑	GTM ↑	GTT ↑	Overall ERS ↑	Overall ARS ↑
CUB	N/A	N/A	58.9	66.4	26.1	59.5	3.3	0.0	0.0	0.0	0.0	0.0	23.9	0.0	3.8	3.5
	ACT	32	27.5	38.2	12.2	43.0	31.0	62.9	30.2	68.5	40.3	34.2	54.2	1.0	33.9	40.3
	HM	32	34.9	45.0	19.8	47.1	31.0	62.9	33.2	69.8	51.3	47.9	**78.2**	2.9	36.0	47.2
	Ours	**16**	34.9	45.1	19.6	45.6	**32.6**	**68.5**	**41.8**	**79.2**	**61.9**	**59.0**	64.8	**5.3**	**38.6**	**51.6**
CARS	N/A	N/A	63.2	75.3	36.6	55.6	0.4	0.0	0.0	3.6	0.0	0.0	21.2	0.0	3.6	2.8
	ACT	32	43.4	54.6	11.8	42.9	36	68.4	35	70.2	37.6	35.3	47.7	1.6	38.6	41.4
	HM	32	60.2	71.6	33.9	51.2	**38.6**	74.8	39.2	75.1	50.3	61.0	**76.4**	8.8	46.1	52.9
	Ours	**16**	60.7	71.2	34.6	49.4	36	**81.0**	**47.0**	**87.5**	**64.4**	**66.9**	60.8	**13.7**	**47.7**	**57.2**
SOP	N/A	N/A	62.9	68.5	39.2	87.4	0.2	0.6	0.3	0.9	0.0	0.0	10.0	0.0	4.0	1.5
	ACT	32	47.5	52.6	25.5	84.9	48.2	90.4	45.4	91.5	44.6	45.5	58.5	15.3	50.8	54.9
	HM	32	46.8	51.7	24.5	84.7	64.0	96.8	67.4	**98.0**	83.5	85.0	81.0	45.6	61.6	77.7
	Ours	**16**	48.3	53.3	25.9	84.9	**65.8**	**97.1**	**71.4**	97.9	**89.4**	**93.4**	**82.4**	**53.1**	**62.4**	**81.3**

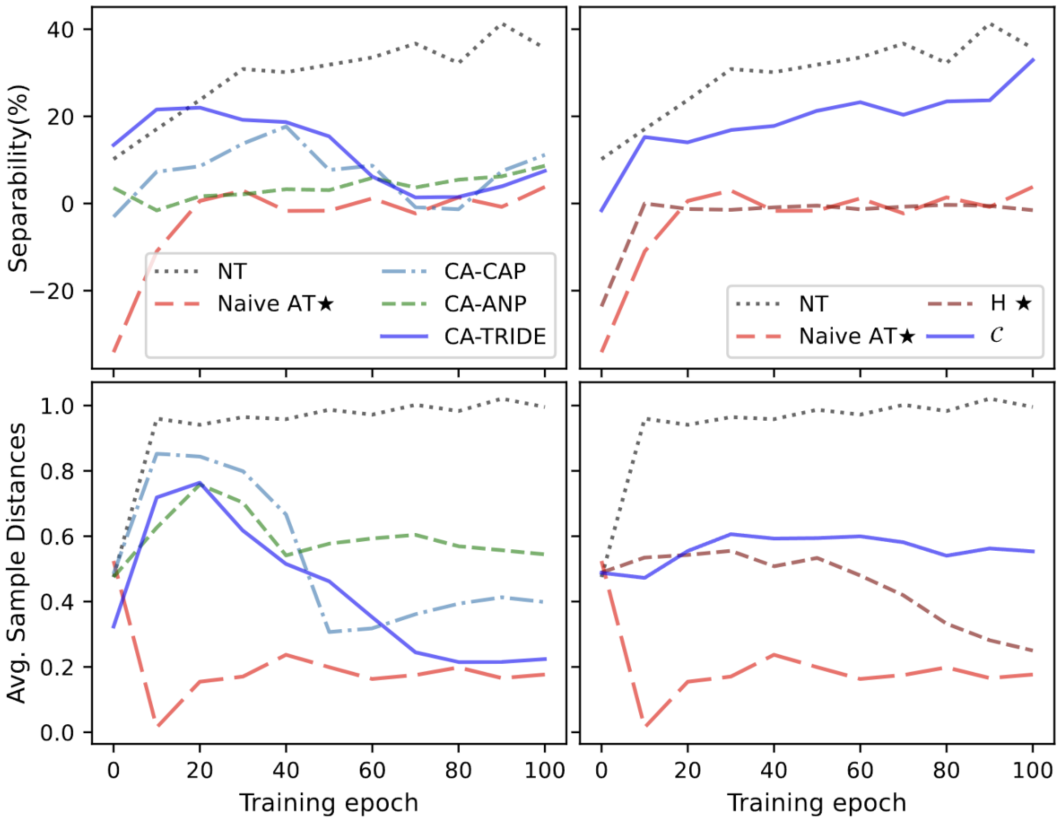
Defense Method	R@1↑	Adversarial Example Evaluation (ARS) (%) CA+↑	CA-↑	QA+↑	QA-↑	ES:R ↑	LTM↑	GTM ↑	GTT ↑	Overall ERS↑	Overall ARS↑
CA-ANP	34.2	27.4	56.7	35.6	73.3	57.3	61.1	65.8	5.1	34.0	47.8
CA-CAP	33.8	34.2	68.0	52.2	70.8	51.2	47.6	60.7	3.1	37.9	48.5
CA-TRIDE	34.9	32.6	68.5	41.8	79.2	61.9	59.0	64.8	5.3	**38.6**	**51.6**

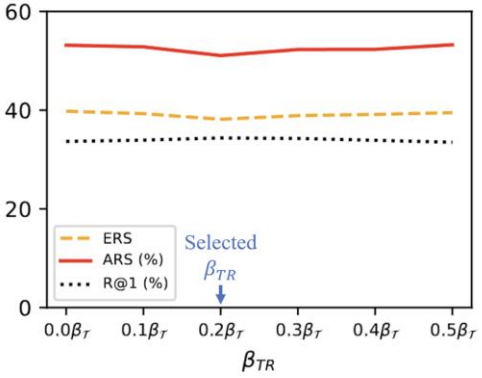
- ✓ CA-TRIDE achieves SOTA performance on both benign and adversarial examples on CUB, CARS and SOP.
- ✓ Through TRIDE, our CA-TRIDE uses less time (~15%) and half PGD steps to achieve better robustness and accuracy.

Experiments



Dataset	Inter-Class Distances	Intra-Class Distances	Entanglement	λ
CUB	0.287	0.226	0.79	10.0
CARS	0.325	0.256	0.79	9.5
SOP	0.664	0.438	0.66	2.0





- ✓ Ablation studies validate the effectiveness of CA to stop model collapse and TRIDE to address the weak adversary.
- ✓ Proven insensibility of our methods towards hyperparameters.
- ✓ Interesting correlation between attention factor λ and entanglement level.

Thanks for watching!

Qiwei Tian
michaeltqw@stu.xjtu.edu.cn

西安交通大學
XI'AN JIAOTONG UNIVERSITY

ICML
International Conference On Machine Learning

arXiv > cs > arXiv:2312.07364

Search...
Help | Adva

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 12 Dec 2023 (v1), last revised 6 Jun 2024 (this version, v4)]

Collapse-Aware Triplet Decoupling for Adversarially Robust Image Retrieval

Qiwei Tian, Chenhao Lin, Zhengyu Zhao, Qian Li, Chao Shen

Adversarial training has achieved substantial performance in defending image retrieval against adversarial examples. However, existing studies in deep metric learning (DML) still suffer from two major limitations: weak adversary and model collapse. In this paper, we address these two limitations by proposing Collapse-Aware TRIplet DEcoupling (CA-TRIDE). Specifically, TRIDE yields a stronger adversary by spatially decoupling the perturbation targets into the anchor and the other candidates. Furthermore, CA prevents the consequential model collapse, based on a novel metric, collapseness, which is incorporated into the optimization of perturbation. We also identify two drawbacks of the existing robustness metric in image retrieval and propose a new metric for a more reasonable robustness evaluation. Extensive experiments on three datasets demonstrate that CA-TRIDE outperforms existing defense methods in both conventional and new metrics. Codes are available at this https URL.

Comments: Accepted by ICML2024
Subjects: **Computer Vision and Pattern Recognition (cs.CV)**
Cite as: arXiv:2312.07364 **[cs.CV]**
(or arXiv:2312.07364v4 **[cs.CV]** for this version)
https://doi.org/10.48550/arXiv.2312.07364

Paper released on Arxiv: https://arxiv.org/pdf/2312.07364

Repository on Github (Mid July): https://github.com/michaeltian108/CA-TRIDE