A Appendix

A.1 Additional analyses

Effect of triplet sampling strategies on ranking defenses Previous works (Roth et al., 2020; Liu et al., 2022) suggest that Soft-hard sampling ($\mathcal{S}$) (Roth et al., 2020) should achieve better Recall@1 as compared to random ($\mathcal{R}$) sampling. And although recall is very similar for EST (for most ϵ) and HM, we see a significant drop in recall for the Anti-Collapse Triplet (ACT) defense when the Soft-hard sampling is used to generate triplets. We hypothesize the Random sampling strategy outperforms the Soft-hard sampling strategy due to the conflict of how Soft-hard sampling generates triplets and the goal of the adversary in ACT. Soft-hard sampling generates triplets that "likely" have hard positives and hard negatives (see Supplement of Liu et al. (2022)), and as a result it is more likely that the positive and negative samples for a given anchor are further apart, making them harder to "collapse"[3].

A.2 Limitations of our work

Though our empirical results at the moment lack theoretical justification, we present strong empirical results which are consistent across multiple datasets and defenses and provide a evidence for benefit of initializing robust ranking defenses with robust features. Given the marked improvement of both retrieval performance and resistance to ranking attacks, we believe the empirical results are interesting in their own right and provide future avenues of research to the community.

A.3 Discussion of societal impact

Though adversarial attacks pose profound risk for various scenarios, our work aims to alleviate such risk by identifying methods to improves adversarial robustness of real work systems. By addressing vulnerabilities in AI systems, this work contributes to the responsible and beneficial deployment of AI technologies. It helps create a more secure, reliable, and equitable environment for individuals, organizations, and society as a whole.

A.4 Use of Existing Assets

These experiments use publicly available code. All of the experiments are run with default parameters found in the corresponding repositories. The GitHub repositories and the licenses can be found at:

- The robust-models-transfer (**?**) repository is released under MIT license and contains the l_∞ robust models used for robust weight initialization in our experiments. The repository can be found at
`https://github.com/Microsoft/robust-models-transfer`
- RobRank (Zhou et al., 2021) repository us released under Apache (Version 2.0) license and can be found at
`https://github.com/cdluminate/robrank`

All of the experiments are run on publicly available datasets

- The CUB-200-2011 (Welinder et al., 2010) dataset can be found at
`http://www.vision.caltech.edu/datasets/cub_200_2011/`
- The CARS (Krause et al., 2013) dataset can be found at
`http://ai.stanford.edu/~jkrause/cars/car_dataset.html`
- The Stanford Online Products (SOP) (Oh Song et al., 2016) dataset can be found at
`https://cvgl.stanford.edu/projects/lifted_struct/`

[3] The goal of ACT defense is to separate "collapsed" embeddings of the positive and negative samples Zhou & Patel (2022).

A.5 Hardware Configuration

Our code is developed on an internal cluster, where each server node is equipped with 4 NVIDIA Tesla A100 cards (each with 40 GB of VRAM), paired with a 64-core AMD EPYC cpu and 256GB of memory. All of our experiments utilize the ResNet-18 architecture.

A.6 Adversarial ranking attack: Candidate attack (CA)

Consider the case of a Candidate attack (CA) Zhou et al. (2020a), where you aim to raise or lower the rank of single candidate, c, with respect to a set of Queries, $Q = \{q_1, \ldots, q_w\}$. Note that locally in DML, a candidate, c_p, is ranked ahead of c_n if it is closer to the query image, i.e. $d(q, c_p) < d(q, c_n)$. Such rank ordering can also be formulated in the form of a triplet loss

$$L_{trip}(q, c_p, c_n) = \max(0, \beta + d(q, c_p) - d(q, c_n)). \tag{3}$$

Note that a CA that raises the rank of c with respect to every query $q \in Q$ ahead of all the candidates (C) can be formulated as a series of inequalities and subsequently a sum of triplet losses,

$$L_{CA+}(c, Q; X) = \sum_{q \in Q} \sum_{x \in X} L_{trip}(q, c, x). \tag{4}$$

Zhou et al. (2020a) point out that the attack / perturbed image is a solution to the following constrained optimization problem:

$$r^* = \arg\min_{r \in \Gamma} L_{CA+}(c + r, Q; X). \tag{5}$$

For a more detailed description of adversarial ranking attacks, see Zhou et al. (2020a).

A.7 Generating adversarial examples for the EST defense.

This is done by running a k-step PGD attacks against a metric loss (e.g., cosine similarity or euclidean distance). Let x_{orig} be a copy of the original image. To find the shifted adversarial examples, the image x is first perturbed in its ϵ-neighborhood, i.e.

$$x_0 \coloneqq x + U(-\epsilon, \epsilon), \tag{6}$$

where U is a uniform distribution. It then follows the generation process of the Basic Iterative Method Kurakin et al. (2016),

$$x^{k+1} \coloneqq \Pi_S \left(x_k + \alpha \cdot \text{sign} \left(\nabla_x \mathcal{L}_{\text{metric}}(x_k, x_{orig}; F_\theta) \right) \right) \tag{7}$$

where K is the number of "attacks steps", or iterations of projected gradient descent performed to find the worst case l_p bounded adversarial example x^{adv}. Π_p is an operator that will project the iterates onto an l_p ball, where in our case $p = 2$ or ∞.

A.8 Further experimental results

Tables A5 and A1 are summarized in Figure 4. Table A3 shows the detailed model performance results for Table A1.

A.9 Empirical robustness score - ERS

The ensemble of adversarial rankings attacks to quantify ranking robustness include the following attacks:

(a) Candidate attacks ($CA+$, $CA-$) (Zhou et al., 2020a) aim is to increase or lower ranks of chosen candidate c with respect to a query set Q

(b) Query attacks ($QA+$, $QA-$) (Zhou et al., 2020a) aim to raise or lower the rank of a set of candidates C by perturbing a single query q

(c) Targeted Mismatch Attack (TMA)(Tolias et al., 2019) increases the similarity between two arbitrary samples

H_S / H_D	Random [$\mathcal{R}$]		Semihard [$\mathcal{M}$]		Softhard [$\mathcal{S}$]		Distance [$\mathcal{D}$]		Hardest [$\mathcal{H}$]		g_LGA	
	R@1	ERS	R@1	ERS	R@1	ERS	R@1	ERS	R@1	ERS	R@1	ERS
Random [$\mathcal{R}$]	42.5	15.1	31.1	45.5	30.6	49.4	28.9	50.3	29.6	50.6	31.0	47.7
Semihard [$\mathcal{M}$]	30.0	12.5	28.6	9.4	39.0	42.4	32.6	49.9	32.9	50.0	34.6	40.7
Softhard [$\mathcal{S}$]	45.7	36.9	41.7	44.6	48.3	13.0	20.8	41.1	20.1	41.8	40.4	47.2
Distance [$\mathcal{D}$]	47.5	10.0	47.8	14.8	23.0	26.1	47.9	9.9	2.3	27.9	42.5	34.4
Hardest [$\mathcal{H}$]	47.4	10.0	48.3	15.0	22.9	26.4	2.3	27.7	47.7	9.5	42.5	34.4

Table A1: Clean (R@1) and robust (ERS) performance (for the CUB Dataset) of ResNet-18 models initialized with a l_∞ robust ImageNet checkpoint ($\epsilon = 4.0/255$) for different combinations of source & destination hardness sampling strategies. Models on the diagonal are regularly (instead of adversarially) trained, and the last-epoch performance is reported to stay consistent with Zhou & Patel (2022). Last columns shows the effect of robust initialization on gradual adversary as H_D in hardness manipulation.

Dataset	Defense	ϵ	Benign Example				White-Box Attacks for Robustness Evaluation										ERS↑
			R@1↑	R@2↑	mAP↑	NMI↑	CA+↑	CA-↓	QA+↑	QA-↓	TMA↓	ES:D↓	ES:R↑	LTM↑	GTM↑	GTT↑	
CUB	HM[$\mathcal{R}$, g_LGA]	4.0	31.0	40.5	13.4	44.2	27.7	25.0	31.6	15.8	0.4	0.6	19.9	19.3	20.3	8.0	47.7
	HM[$\mathcal{M}$, g_LGA]	4.0	34.6	44.9	16.7	48.7	20.3	31.8	21.5	25.4	0.4	0.8	18.5	16.8	20.9	3.3	40.7
	HM[$\mathcal{S}$, g_LGA]	4.0	40.4	51.3	18.6	51.3	21.9	23.3	23.9	15.4	0.5	0.5	28.1	28.8	26.4	9.5	47.2
	HM[$\mathcal{D}$, g_LGA]	4.0	42.5	52.8	19.9	51.3	11.8	44.3	10.9	41.0	0.4	0.9	21.1	18.0	23.4	2.8	34.4
	HM[$\mathcal{H}$, g_LGA]	4.0	42.5	52.5	19.3	50.9	12.0	43.9	11.3	41.5	0.4	0.9	21.0	18.0	23.5	2.6	34.5

Table A2: Effect of robust initialization on gradual adversary as H_D in hardness manipulation.. The "↑" mark means "the higher the better", while "↓" means the opposite.

(d) Embedding Shift attack (ES)(Feng et al., 2020) moves the embedding of a query as far as possible from its original position

(e) Learning-To-Misrank attack (LTM) (Wang et al., 2020) aims to perturb rank ordering by minimizing / maximizing distance of matched / unmatched pairs

(f) Greedy Top-1 Misranking attack (GTM) (Zhou et al., 2021) reduces the distance between adversarial query and closest non-matching candidate, and

(g) Greedy Top-1 Translocation attack (GTT) (Zhou et al., 2021) moves the top retrieval result of out top-k ranked items.

Dataset	Defense	η	Benign Example				White-Box Attacks for Robustness Evaluation										ERS
			R@1↑	R@2↑	mAP↑	NMI↑	CA+↑	CA-↓	QA+↑	QA-↓	TMA↓	ES:D↓	ES:R↑	LTM↑	GTM↑	GTT↑	
CUB	HM[$\mathcal{R},\mathcal{R}$]	N/A	42.5	53.8	24.6	52.2	4.4	76.7	3.3	76.6	0.7	1.2	4.5	2.9	13.9	0.2	15.1
	HM[$\mathcal{R},\mathcal{M}$]	8	31.1	40.2	17.1	45.3	26.7	29.5	29.4	19.0	0.4	0.6	19.6	19.0	19.9	6.2	45.5
	HM[$\mathcal{R},\mathcal{S}$]	8	30.6	40.1	16.5	44.8	29.3	20.9	32.7	11.5	0.6	0.4	22.0	22.4	21.6	12.6	49.4
	HM[$\mathcal{R},\mathcal{D}$]	8	28.9	38.9	15.2	43.1	30.9	17.5	35.0	9.5	0.7	0.3	21.5	21.9	21.4	16.2	50.3
	HM[$\mathcal{R},\mathcal{H}$]	8	29.6	39.8	15.7	43.4	31.2	17.1	35.1	9.2	0.6	0.3	21.6	22.1	21.5	16.5	50.6
CUB	HM[$\mathcal{M},\mathcal{R}$]	8	30.0	41.1	16.8	46.1	3.9	88.1	5.0	85.1	0.7	1.4	3.3	0.8	11.5	0.0	12.5
	HM[$\mathcal{M},\mathcal{M}$]	N/A	28.6	39.5	15.8	45.5	2.9	95.2	3.4	92.2	0.7	1.5	1.2	0.3	10.8	0.0	9.4
	HM[$\mathcal{M},\mathcal{S}$]	8	39.0	49.5	22.6	49.5	19.0	28.1	20.8	22.4	0.5	0.6	25.2	21.7	24.2	6.2	42.4
	HM[$\mathcal{M},\mathcal{D}$]	8	32.6	42.6	18.2	44.9	28.2	13.9	33.2	7.8	0.8	0.3	25.8	25.3	23.2	16.6	49.9
	HM[$\mathcal{M},\mathcal{H}$]	8	32.9	42.7	18.5	44.8	28.6	14.1	33.0	7.9	0.8	0.3	26.4	25.0	22.9	16.9	50.0
CUB	HM[$\mathcal{S},\mathcal{R}$]	8	45.7	57.4	28.6	54.9	12.8	37.2	13.7	31.1	0.7	0.5	26.1	26.1	25.4	4.1	36.9
	HM[$\mathcal{S},\mathcal{M}$]	8	41.7	53.6	25.7	52.5	20.0	26.6	22.2	19.2	0.6	0.5	27.9	28.1	25.7	7.1	44.6
	HM[$\mathcal{S},\mathcal{S}$]	N/A	48.3	60.5	29.6	55.9	1.2	86.3	1.2	84.5	0.9	0.7	5.3	3.4	15.1	0.0	13.0
	HM[$\mathcal{S},\mathcal{D}$]	8	20.8	30.2	6.4	33.2	23.1	36.2	32.6	21.0	1.0	0.0	12.0	19.0	18.3	7.5	41.1
	HM[$\mathcal{S},\mathcal{H}$]	8	20.1	29.5	6.5	34.4	24.0	32.7	32.6	18.3	1.0	0.0	13.7	17.4	17.7	7.6	41.8
CUB	HM[$\mathcal{D},\mathcal{R}$]	8	47.5	59.1	28.8	55.2	1.1	91.8	0.8	92.7	0.7	1.4	3.1	1.9	14.5	0.0	10.0
	HM[$\mathcal{D},\mathcal{M}$]	8	547.8	59.7	30.1	55.8	2.4	80.8	1.6	83.5	0.7	1.1	7.0	4.6	17.0	0.0	14.8
	HM[$\mathcal{D},\mathcal{S}$]	8	23.0	32.9	8.6	35.7	6.5	75.6	16.3	46.3	1.0	0.0	10.7	9.4	14.3	2.5	26.1
	HM[$\mathcal{D},\mathcal{D}$]	N/A	47.9	58.8	29.3	54.4	1.1	91.7	0.6	93.2	0.7	1.4	3.0	1.7	14.5	0.0	9.9
	HM[$\mathcal{D},\mathcal{H}$]	8	2.3	3.1	1.2	7.7	14.8	86.8	39.9	52.2	1.0	0.0	1.5	1.6	1.6	3.8	27.9
CUB	HM[$\mathcal{H},\mathcal{R}$]	8	47.4	58.9	29.1	54.4	1.1	91.8	0.8	92.9	0.7	1.4	3.0	2.0	14.6	0.1	10.0
	HM[$\mathcal{H},\mathcal{M}$]	8	48.3	59.7	30.3	55.8	2.6	80.2	1.9	82.2	0.7	1.1	7.0	4.7	17.0	0.1	15.0
	HM[$\mathcal{H},\mathcal{S}$]	8	22.9	33.0	8.2	33.9	6.5	76.2	17.2	46.3	1.0	0.0	12.9	10.1	13.7	2.3	26.4
	HM[$\mathcal{H},\mathcal{D}$]	8	2.3	3.0	1.2	9.4	14.6	85.7	38.9	52.8	1.0	0.0	1.9	1.2	1.2	4.0	27.7
	HM[$\mathcal{H},\mathcal{H}$]	N/A	47.7	60.0	29.4	55.3	0.9	93.0	0.6	94.2	0.7	1.4	2.9	1.4	14.5	0.0	9.5

Table A3: Detailed model performance for all combinations of source and destination sampling strategies in Table 2. in the Paper. The symbols $\mathcal{R},\mathcal{M},\mathcal{S},\mathcal{D},\mathcal{H}$ denote Random, Semihard, Softhard, Distance and Hardest triplet sampling strategies, respectively.

Defense	η	ϵ	Benign Example				White-Box Attacks for Robustness Evaluation										ERS↑
			R@1↑	R@2↑	mAP↑	NMI↑	CA+↑	CA-↓	QA+↑	QA-↓	TMA↓	ES:D↓	ES:R↑	LTM↑	GTM↑	GTT↑	
N/A[$\mathcal{R}$]	N/A	N/A	53.9	66.4	26.1	59.5	0.0	100.0	0.0	99.9	0.883	1.762	0.0	0.0	14.1	0.0	3.8
EST[$\mathcal{R}$]	8	0.0	35.0	46.1	17.6	45.0	0.4	97.8	0.4	93.3	0.9	1.2	5.1	0.3	10.5	0.0	7.3
	8	1.0	44.3	56.0	26.0	53.2	8.2	58.6	7.1	51.1	0.6	0.9	18.8	11.7	16.9	0.4	26.1
	8	4.0	37.9	49.6	22.5	50.4	16.1	36.7	17.0	25.8	0.5	0.6	25.2	22.3	20.9	3.6	39.3
	32	0.0	7.0	11.2	2.0	25.5	1.5	98.7	0.8	98.0	0.8	1.6	1.0	0.0	3.9	0.0	4.9
	32	1.0	42.2	52.7	24.6	50.7	15.5	38.9	15.3	28.7	0.5	0.7	26.0	20.5	19.9	2.1	37.5
	32	4.0	37.0	49.0	21.0	49.6	19.8	36.6	22.3	18.2	0.5	0.6	26.9	25.3	21.6	4.4	42.7
ACT[$\mathcal{R}$]	8	0.0	31.0	41.4	16.8	45.4	13.3	47.3	12.3	41.8	0.6	0.9	11.3	7.6	14.3	0.4	29.4
	8	1.0	41.9	53.3	25.3	51.1	17.7	36.6	16.4	29.3	0.4	0.8	21.4	18.6	22.0	2.2	38.2
	8	4.0	36.2	47.6	22.6	49.7	20.6	30.0	21.3	22.1	0.4	0.7	19.4	20.6	21.5	4.2	42.1
	32	0.0	26.4	36.1	13.6	42.6	17.0	38.2	17.2	29.7	0.5	0.8	13.4	8.3	14.0	0.8	34.5
	32	1.0	40.7	50.7	24.2	50.7	19.1	33.0	19.0	25.6	0.4	0.7	23.1	19.7	21.5	2.7	40.5
	32	4.0	36.6	46.5	21.6	48.9	21.1	28.4	21.6	21.3	0.4	0.7	19.8	19.3	21.3	3.9	42.5
HM[$\mathcal{S}, g_{\text{LGA}}$]	8	0.0	39.5	49.8	23.8	50.2	10.5	51.0	10.5	48.7	0.6	0.8	12.6	11.0	17.3	0.9	28.1
	8	1.0	43.8	54.8	26.6	53.9	18.4	31.0	19.2	23.6	0.5	0.6	26.6	26.0	25.9	4.1	42.0
	8	4.0	42.2	53.1	25.1	51.6	20.8	25.9	21.8	18.3	0.5	0.5	27.6	28.2	25.5	8.0	45.2
	32	0.0	37.6	48.5	22.2	48.5	12.4	47.1	12.3	43.1	0.6	0.8	14.7	11.8	18.3	1.0	30.7
	32	1.0	43.3	53.5	25.9	52.4	19.7	27.7	20.5	20.5	0.5	0.6	25.8	26.9	26.0	5.5	43.9
	32	4.0	39.6	50.6	23.9	51.5	22.1	23.0	23.7	15.6	0.5	0.5	27.0	28.7	26.8	9.2	47.3
HM[$\mathcal{S}, g_{\text{LGA}}$]&ICS	8	0.0	38.1	48.9	22.7	49.8	11.3	47.8	11.8	43.2	0.8	0.5	14.7	12.3	18.2	0.9	29.1
	8	1.0	44.1	55.8	26.3	51.9	19.3	27.1	22.2	19.5	0.8	0.4	28.0	27.6	25.5	5.9	42.6
	8	4.0	41.1	51.9	25.1	51.8	21.8	21.3	25.5	14.9	0.8	0.3	29.0	30.0	26.1	10.4	46.0
	32	0.0	36.8	47.8	21.3	48.7	13.4	44.9	13.9	38.8	0.8	0.6	15.1	13.0	18.7	1.3	31.3
	32	1.0	42.7	54.2	25.7	51.7	20.2	25.0	21.9	17.3	0.8	0.4	27.1	29.1	27.1	6.8	43.8
	32	4.0	39.6	50.9	24.0	50.9	23.1	19.9	26.1	12.5	0.8	0.3	28.8	29.9	27.1	12.5	47.3

Table A4: Effect of η and ϵ on state-of-the-art defenses: a) Embedding-Shifted Triplet (EST) (Zhou et al., 2020a), (b) Anti-Collapse Triplet (ACT) (Zhou et al., 2021), and (c) Hardness Manipulations (HM) (Zhou & Patel, 2022).

Dataset	Defense	ϵ	Benign Example				White-Box Attacks for Robustness Evaluation										ERS↑
			R@1↑	R@2↑	mAP↑	NMI↑	CA+↑	CA-↓	QA+↑	QA-↓	TMA↓	ES:D↓	ES:R↑	LTM↑	GTM↑	GTT↑	
CUB	N/A[$\mathcal{R}$]	N/A	53.9	66.4	26.1	59.5	0.0	100.0	0.0	99.9	0.883	1.762	0.0	0.0	14.1	0.0	3.8
CUB	EST[$\mathcal{R}$]	0.0	35.0	46.1	17.6	45.0	0.4	97.8	0.4	93.3	0.9	1.2	5.1	0.3	10.5	0.0	7.3
	EST[$\mathcal{R}$]	0.5	45.4	57.6	27.6	53.8	7.3	60.7	5.7	59.9	0.6	1.0	15.1	9.3	16.5	0.3	23.5
	EST[$\mathcal{R}$]	1.0	44.3	56.0	26.0	53.2	8.2	58.6	7.1	51.1	0.6	0.9	18.8	11.7	16.9	0.4	26.1
	EST[$\mathcal{R}$]	2.0	40.9	53.1	23.9	51.2	12.0	46.4	12.2	35.2	0.6	0.7	24.2	18.2	19.6	1.3	33.6
	EST[$\mathcal{R}$]	4.0	37.9	49.6	22.5	50.4	16.1	36.7	17.0	25.8	0.5	0.6	25.2	22.3	20.9	3.6	39.3
	EST[$\mathcal{R}$]	8.0	33.7	44.1	18.4	46.9	16.6	38.0	17.4	27.0	0.5	0.6	22.7	19.9	19.4	3.6	38.5
CUB	EST[$\mathcal{S}$]	0.0	43.0	54.7	24.6	50.7	0.0	99.8	0.1	97.9	1.0	0.5	7.8	0.2	11.0	0.0	9.8
	EST[$\mathcal{S}$]	0.5	49.5	61.0	30.8	55.5	3.0	70.6	3.7	66.0	0.9	0.5	22.6	9.5	18.6	0.2	21.3
	EST[$\mathcal{S}$]	1.0	48.7	60.3	29.8	56.4	3.7	66.6	4.5	59.5	0.9	0.4	22.6	15.8	20.4	0.4	23.7
	EST[$\mathcal{S}$]	2.0	45.6	57.8	27.9	54.6	7.6	55.6	9.0	43.1	0.9	0.4	27.4	24.5	21.9	1.1	30.0
	EST[$\mathcal{S}$]	4.0	41.3	53.1	24.6	52.5	9.0	46.9	11.1	35.2	0.9	0.4	27.9	21.7	22.1	2.3	32.7
	EST[$\mathcal{S}$]	8.0	39.3	50.3	22.1	50.2	12.8	39.0	15.9	27.3	0.9	0.3	27.3	25.1	21.9	3.6	36.6
CUB	ACT[$\mathcal{R}$]	0.0	31.0	41.4	16.8	45.4	13.3	47.3	12.3	41.8	0.6	0.9	11.3	7.6	14.3	0.4	29.4
	ACT[$\mathcal{R}$]	0.5	42.0	52.7	25.7	51.7	16.4	40.0	15.2	32.7	0.5	0.8	17.6	16.1	20.5	1.7	35.6
	ACT[$\mathcal{R}$]	1.0	41.9	53.3	25.3	51.1	17.7	36.6	16.4	29.3	0.4	0.8	21.4	18.6	22.0	2.2	38.2
	ACT[$\mathcal{R}$]	2.0	40.1	50.5	24.1	51.2	18.4	34.0	17.9	26.0	0.4	0.8	20.1	19.6	21.0	3.1	39.4
	ACT[$\mathcal{R}$]	4.0	36.2	47.6	22.6	49.7	20.6	30.0	21.3	22.1	0.4	0.7	19.4	20.6	21.5	4.2	42.1
	ACT[$\mathcal{R}$]	8.0	32.8	43.1	18.5	46.8	20.6	30.1	21.2	22.2	0.4	0.7	18.1	17.8	20.6	3.9	41.5
CUB	ACT[$\mathcal{S}$]	0.0	45.3	56.6	28.6	54.6	2.7	80.8	1.8	83.2	0.7	1.3	4.8	3.6	14.4	0.0	13.7
	ACT[$\mathcal{S}$]	0.5	52.8	63.6	34.2	59.2	6.7	59.9	5.3	59.9	0.6	1.0	12.5	12.1	19.5	0.3	23.7
	ACT[$\mathcal{S}$]	1.0	50.8	62.1	32.9	59.2	8.3	56.8	6.3	54.4	0.6	0.9	17.3	15.7	20.8	0.5	26.7
	ACT[$\mathcal{S}$]	2.0	48.5	60.3	31.1	57.7	9.5	49.6	7.9	47.9	0.6	0.9	19.2	17.3	22.1	0.9	29.6
	ACT[$\mathcal{S}$]	4.0	45.0	57.1	28.7	55.3	10.4	46.4	9.4	43.8	0.6	0.8	19.0	18.2	22.1	1.3	31.2
	ACT[$\mathcal{S}$]	8.0	41.6	53.8	25.4	52.3	11.5	43.9	10.9	39.7	0.6	0.7	19.0	17.4	21.2	1.9	32.4
CUB	HM[$\mathcal{R}, g_{\mathsf{LGA}}$]	0.0	27.3	36.8	13.9	43.2	19.2	47.3	21.6	34.4	0.5	0.9	10.3	8.2	14.5	1.2	33.7
	HM[$\mathcal{R}, g_{\mathsf{LGA}}$]	0.5	33.9	43.7	18.9	46.1	22.9	36.9	25.3	25.6	0.5	0.8	16.5	14.6	20.1	3.4	40.4
	HM[$\mathcal{R}, g_{\mathsf{LGA}}$]	1.0	33.9	43.6	18.6	46.2	24.3	33.7	26.9	22.6	0.4	0.7	19.8	17.5	20.2	4.6	43.0
	HM[$\mathcal{R}, g_{\mathsf{LGA}}$]	2.0	31.6	41.8	17.5	45.3	25.5	32.1	27.7	20.6	0.4	0.6	20.3	18.1	20.2	5.6	44.2
	HM[$\mathcal{R}, g_{\mathsf{LGA}}$]	4.0	31.0	40.7	16.7	45.6	27.7	26.6	30.1	16.6	0.4	0.6	19.6	18.7	20.0	6.9	46.7
	HM[$\mathcal{R}, g_{\mathsf{LGA}}$]	8.0	26.9	35.6	13.7	42.9	28.4	23.9	32.4	15.3	0.4	0.6	18.2	17.3	19.0	9.7	47.8
CUB	HM[$\mathcal{S}, g_{\mathsf{LGA}}$]	0.0	39.5	49.8	23.8	50.2	10.5	51.0	10.5	48.7	0.6	0.8	12.6	11.0	17.3	0.9	28.1
	HM[$\mathcal{S}, g_{\mathsf{LGA}}$]	0.5	44.6	55.5	27.7	53.2	17.2	33.4	17.8	26.4	0.5	0.6	25.0	23.6	24.3	2.9	40.0
	HM[$\mathcal{S}, g_{\mathsf{LGA}}$]	1.0	43.8	54.8	26.6	53.9	18.4	31.0	19.2	23.6	0.5	0.6	26.6	26.0	25.9	4.1	42.0
	HM[$\mathcal{S}, g_{\mathsf{LGA}}$]	2.0	43.9	54.9	26.5	52.9	19.3	28.6	20.7	20.3	0.5	0.6	26.5	27.0	26.2	5.3	43.5
	HM[$\mathcal{S}, g_{\mathsf{LGA}}$]	4.0	42.2	53.1	25.1	51.6	20.8	25.9	21.8	18.3	0.5	0.5	27.6	28.2	25.5	8.0	45.2
	HM[$\mathcal{S}, g_{\mathsf{LGA}}$]	8.0	37.6	48.9	22.3	49.8	21.3	23.9	22.5	17.0	0.5	0.5	25.0	26.8	24.6	8.7	45.4

Table A5: Effect of robust initialization on ranking defenses for deep representations learning: Embedding-Shifted Triplet (EST), Anti-collapse triplet (ACT) and Hardness Manipulations (HM). The "↑" mark means "the higher the better", while "↓" means the opposite.