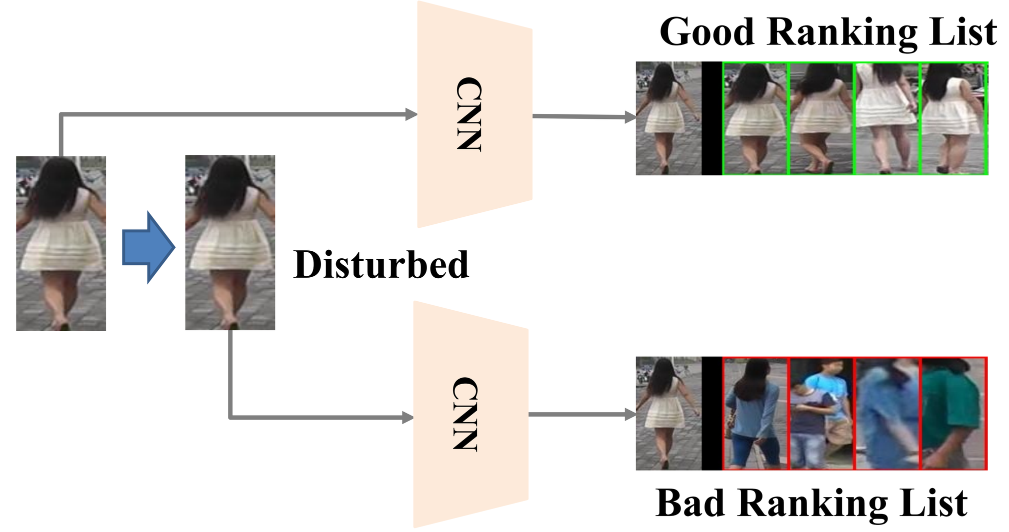
Learning to Attack Real-World Models for Person Re-Identification via Virtual-Guided Meta-Learning

Fengxiang Yang[1], Zhun Zhong[2], Hong Liu[3], Zheng Wang[3,4], Zhiming Luo[1] , Shaozi Li[1], Nicu Sebe[2,5], Shin'ichi Satoh[3]
[1]Xiamen University [2]University of Trento [3]National Institute of Informatics [4]University of Tokyo [5]Huawei Research, Ireland

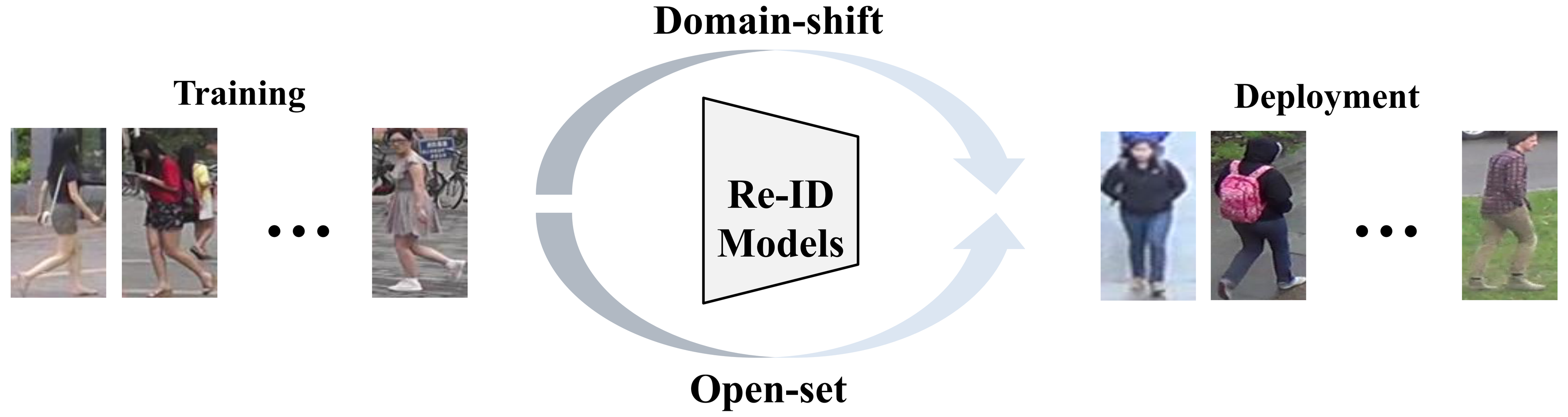
Problem Definition



Motivation: Verifying and improving the robustness of re-ID models.

Approach: Learning adversarial examples to corrupt re-ID models.

Challenges



- Domain shift and open-set property in re-ID requires attackers to adapt to different envirnments, i.e., attackers should be **universal**.
- Recent works[1,2] on re-ID attack generate adversarial examples individually and are not **efficient** enough.

Our Solution & Contributions

How to achieve efficient ?: Universal Adversarial Perturbation[3].

Why UAP ?: Simplify the attack by adding UAP to queries.

How to achieve universal ?: Synthetic Data & Meta-Learning.

Why Virtual Data ?

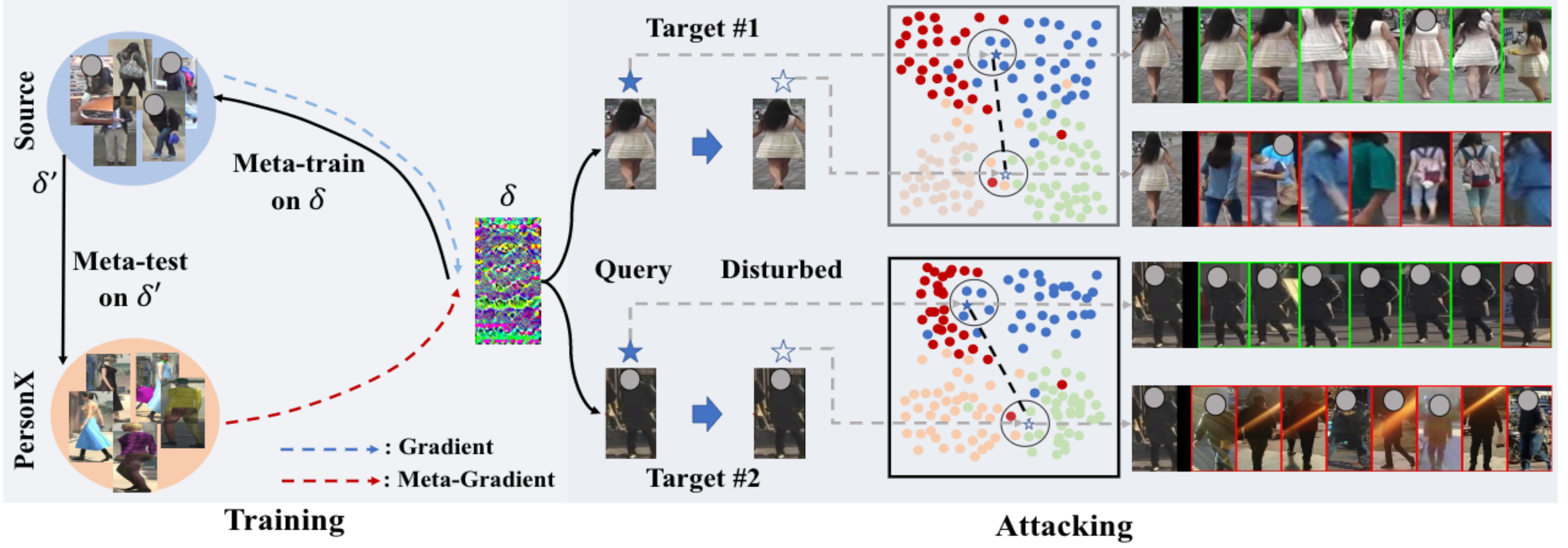
(1) Easy to collect (2) Privacy-free (3) Balanced data distribution.

Why Meta-learning ? Improve universality.

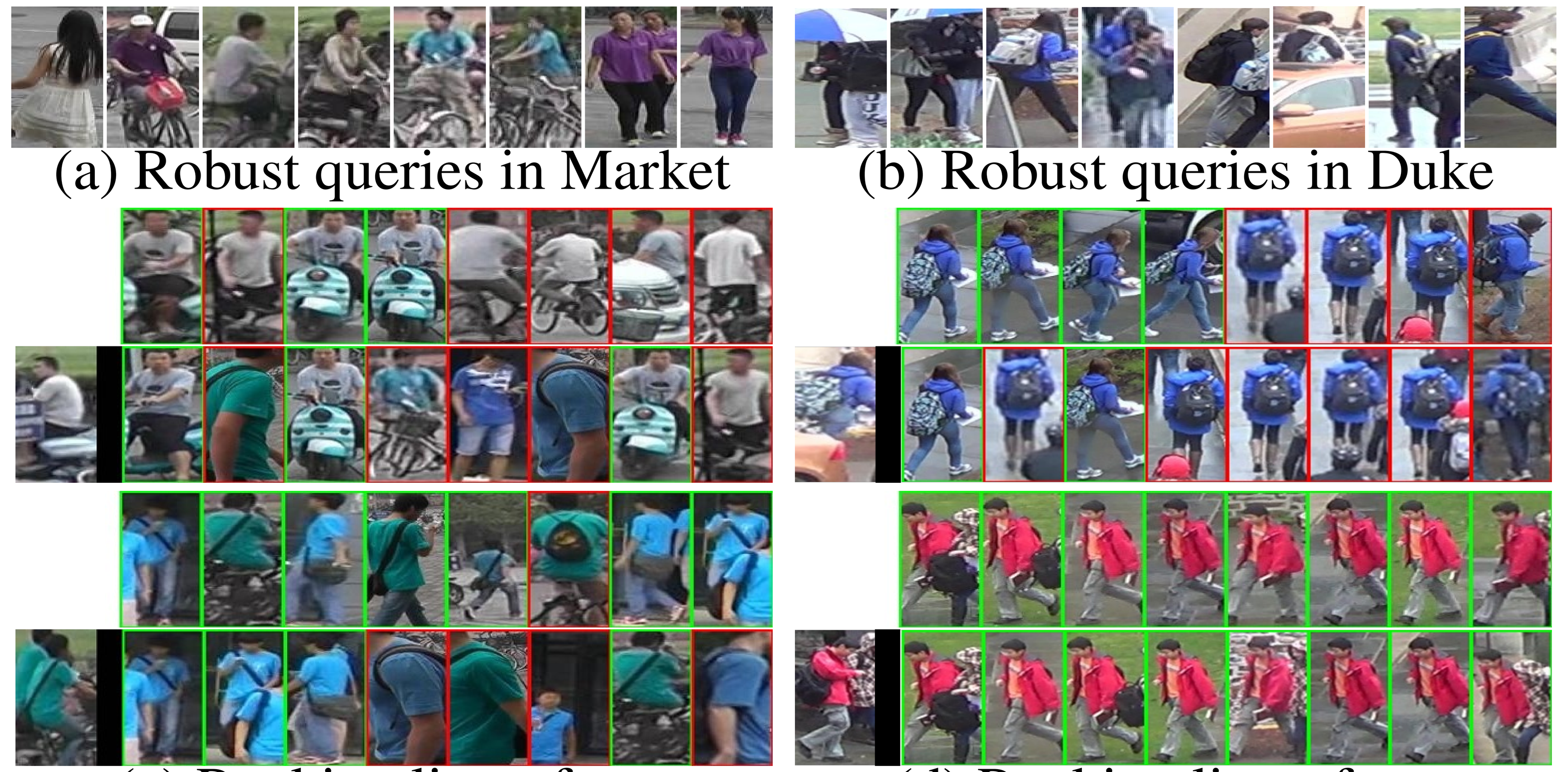
Contributions:

(1) Meta-learning strategy. (2) Virtual data for optimization. (3) Inspiration of improving robustness obtained from visualization.

Framework



Visualization of Robust Queries



(a) Robust queries in Market (b) Robust queries in Duke

(c) Ranking lists of two robust queries in Market (d) Ranking lists of two robust queries in in Duke

We visualize robust queries that survived from our attack and have 2 findings:

Finding 1: Occlusion is robust to attack. **Suggestion 1**: Erasing may improve robustness[4].

Finding 2: Camera styles are robust. **Suggestion 2**: Camera styles may improve robustness.

Contact Us

If you have any problem, please send email to us (**yangfx@stu.xmu.edu.cn**) or ask in Github.

Scan the right QR code for code and other resources.

Experimental Results

Tab 1. Results for attacking re-ID systems. We use our method to attack different backbones (IDE) and part-based PCB), then compare our method with state-of-the-arts (MisRank and UAP-Retrieval). "Before Attack": re-ID accuracies of unseen target model on target set.

Backbone	Methods	Duke → Market		Duke → MSMT		Market → Duke		Market → MSMT	
		mAP	rank-1	mAP	rank-1	mAP	rank-1	mAP	rank-1
IDE	Before Attack	78.2	88.7	42.3	69.8	66.7	80.9	42.3	69.8
	MisRank	28.2	38.6	11.7	30.3	36.7	48.8	11.1	28.5
	MisRank + PersonX	38.5	51.5	20.9	55.8	43.4	71.2	12.4	31.0
	MisRank ($\epsilon = 16$)	10.3	13.0	3.0	7.2	13.7	18.3	1.6	4.2
	UAP-Retrieval	8.2	9.7	5.5	15.4	14.8	20.4	5.3	13.9
	MetaAttack (Ours)	**4.9**	**7.0**	**3.5**	**8.3**	**11.2**	**15.2**	**3.4**	**8.3**
	MetaAttack (Ours, $\epsilon = 16$)	**0.7**	**0.9**	**0.3**	**0.7**	**1.0**	**1.3**	**0.5**	**1.1**
PCB	Before Attack	76.7	91.3	50.8	88.9	68.0	84.1	50.8	88.9
	MisRank	48.1	64.2	21.1	47.7	31.2	45.4	14.4	28.5
	MisRank + PersonX	52.4	70.6	18.8	39.6	38.0	51.4	18.8	39.6
	MisRank ($\epsilon = 16$)	11.5	13.8	5.2	9.6	12.4	17.8	8.2	17.0
	UAP-Retrieval	21.6	30.4	4.4	9.1	29.0	41.9	4.3	8.9
	MetaAttack (Ours)	**19.5**	**28.2**	**4.2**	**8.7**	**26.9**	**39.9**	**3.8**	**8.2**
	MetaAttack (Ours, $\epsilon = 16$)	**4.5**	**5.9**	**0.6**	**1.4**	**4.1**	**6.6**	**0.9**	**1.9**

Tab 2. Ablation study on the proposed virtual-guided meta-learning algorithm.

No.	Duke → MSMT		Market → MSMT		Extra Data		Meta
	mAP	rank-1	mAP	rank-1	Real	PersonX	Learning
1	5.6	14.3	5.8	14.9	×	×	×
2	5.1	14.5	5.7	14.3	✓	×	×
3	4.8	10.4	5.0	12.6	✓	×	✓
4	4.6	9.9	5.5	14.2	×	✓	×
5	**3.5**	**8.3**	**3.4**	**8.3**	×	✓	✓

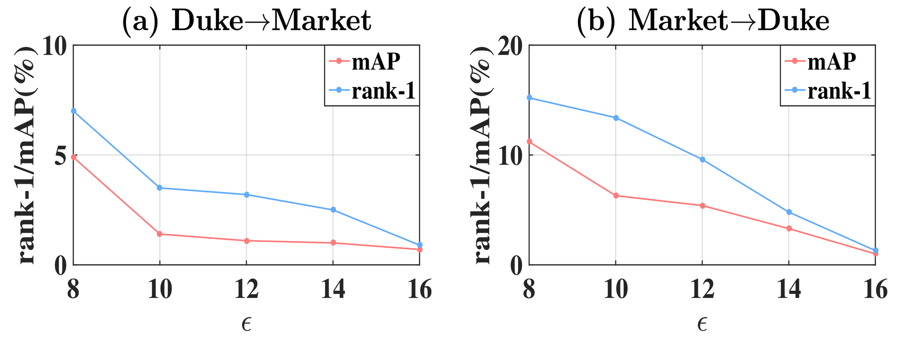


Fig 1. Sensitivity analysis of ϵ.

Tab 3. Results on source domain.

Backbone	Method	Duke		Market	
		mAP	rank-1	mAP	rank-1
IDE	Before Attack	66.7	80.9	78.2	88.7
	UAP-Retrieval	4.2	9.9	3.6	4.5
	Ours	**3.6**	**6.4**	**3.1**	**3.4**
PCB	Before Attack	68.0	84.1	76.7	91.3
	UAP-Retrieval	14.3	20.3	**10.7**	**15.1**
	Ours	**11.2**	**16.5**	10.9	15.4

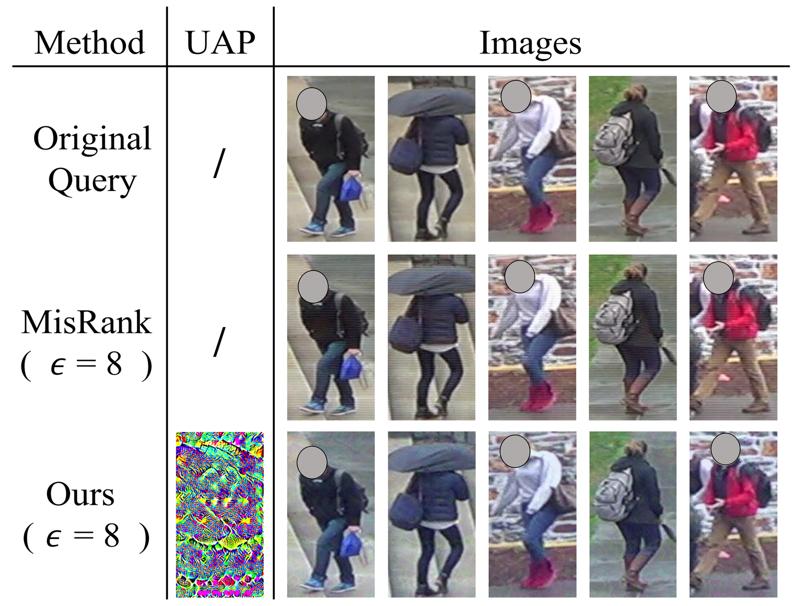


Fig 2. Visualizations of corrupted queries and obtained δ.

Experimental Settings

Train:

Optimize UAP with source and virtual data.

Test:

Directly test UAP on target datasets that have not been used in training phase.

References

[1] Wang et al. Transferable, controllable, and inconspicuous adversarial attacks on person re-identification with deep mis-ranking. CVPR'20.

[2] Bai et al. Metric attack and defense for person re-identification. TPAMI'20.

[3] Moosavi-Dezfooli et al. Universal adversarial perturbations. CVPR'17.

[4] Carmon et al. Unlabeled data improves adversarial robustness. NeurIPS'19.