Master's Degree Thesis

Synthetic-to-Real Domain Transfer with Joint Image Translation and Discriminative Learning for Pedestrian Re-Identification

Masters Degree in Data Science and Engineering
Politecnico di Torino

Supervisors:
Prof. Barbara Caputo
Dott. Mirko Zaffaroni

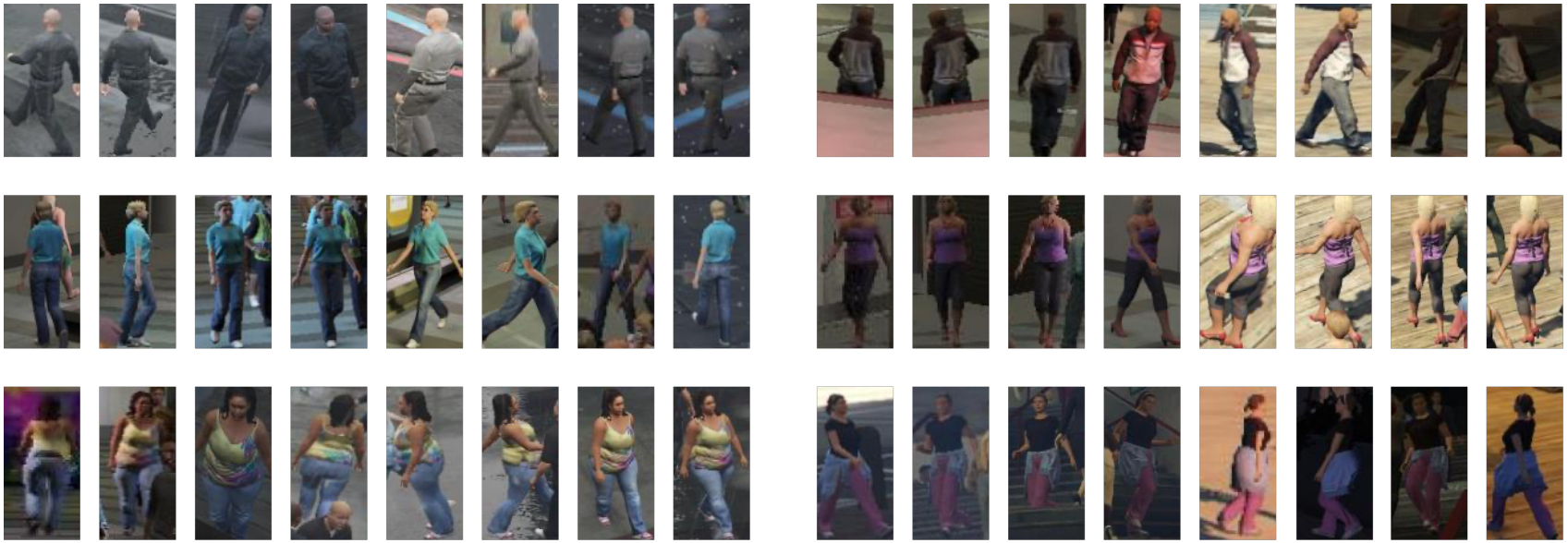
Candidate:
Antonio Dimitris Defonte

Objective: generalize from our synthetic dataset to real-world data using a generative framework applied to pedestrian re-identification.

- Recognizing the same pedestrian across several camera views.
- Person re-identification systems provide further intelligence and security in high-risk areas.
- The thesis and internship were completed at the *Links Foundation*.

GTASynthReid

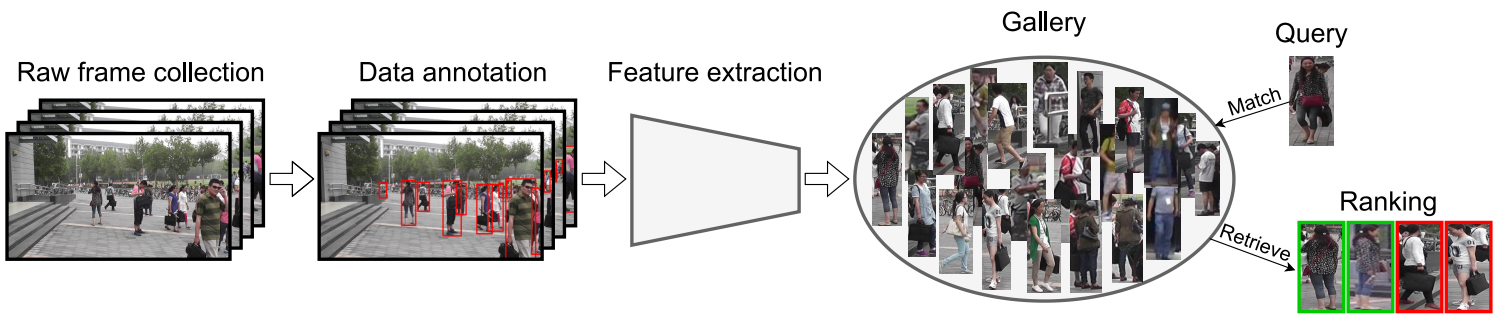
- Exploited the graphic engine of *Grand Theft Auto V*.
- A total of 94312 images with 538 synthetic identities.
- There are 19 camera views, from 2 to 5 for each pedestrian.



Person re-identification

Recognizing the same pedestrian across several camera views.

- Pose, illumination and viewpoint variations.
- Occlusions and person-to-person interactions.
- Training and test sets do not share the same identities.
- Retrieval between query and gallery images.

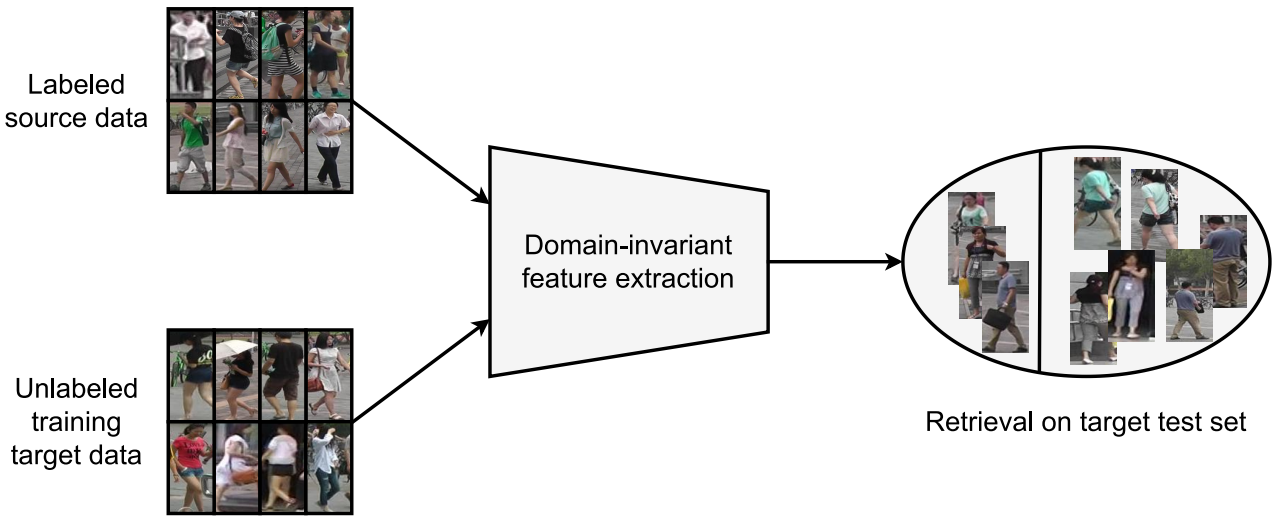


Bounding boxes are from the PRW dataset [1].

Cross-domain techniques

Source-to-target cross-domain techniques.

- Drop in performance when evaluating on other data.
- Evaluate on the test target data (unseen identities).
- Most works embrace generative approaches or *iterative pseudo labeling* techniques.

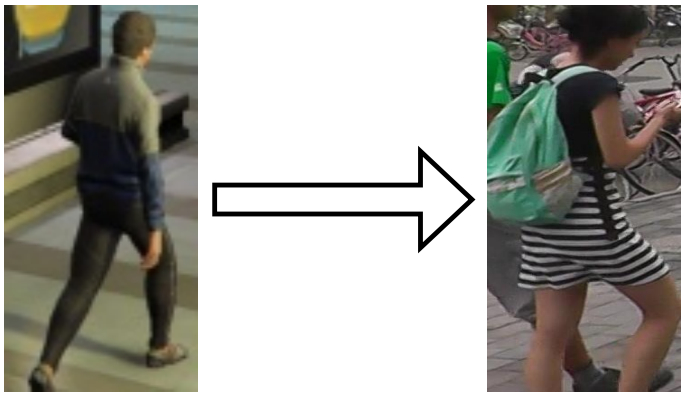


Bounding boxes are from the PRW [1] and Market [2] datasets.

Synthetic data

Adapt from computer-generated characters to real pedestrians.

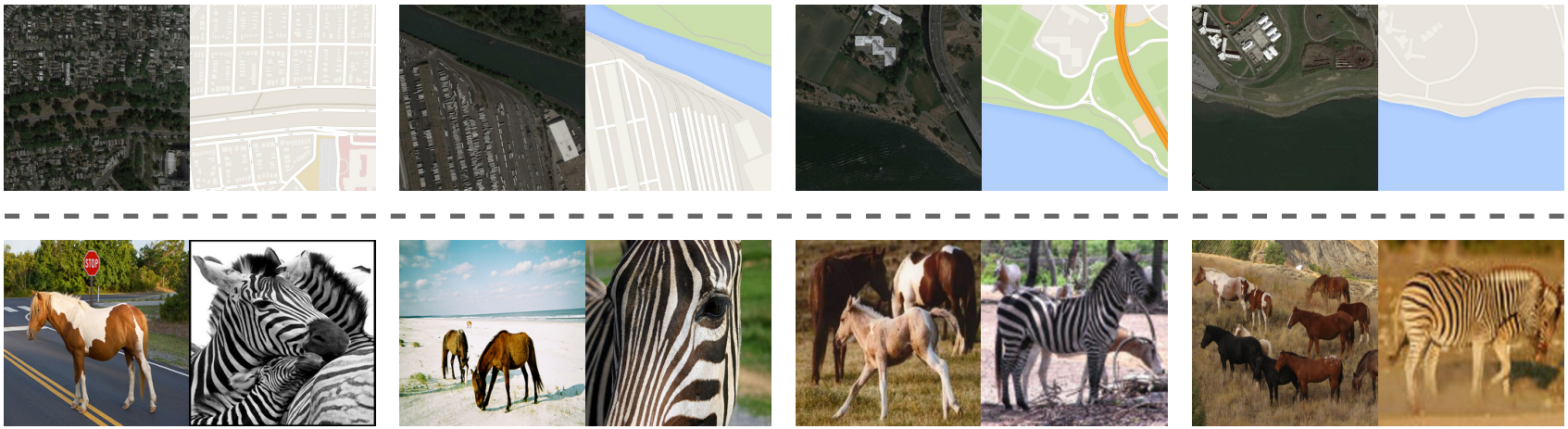
- Collecting data is error-prone and time-consuming.
- Synthetic worlds can model far more environment variations.
- Growing ethical concerns of real-world data.
- Additional complexity for synth-to-real adaptation.



Right bounding box is from the PRW [1] dataset.

Image translation

Matching the style of the target data, it can be paired or unpaired.

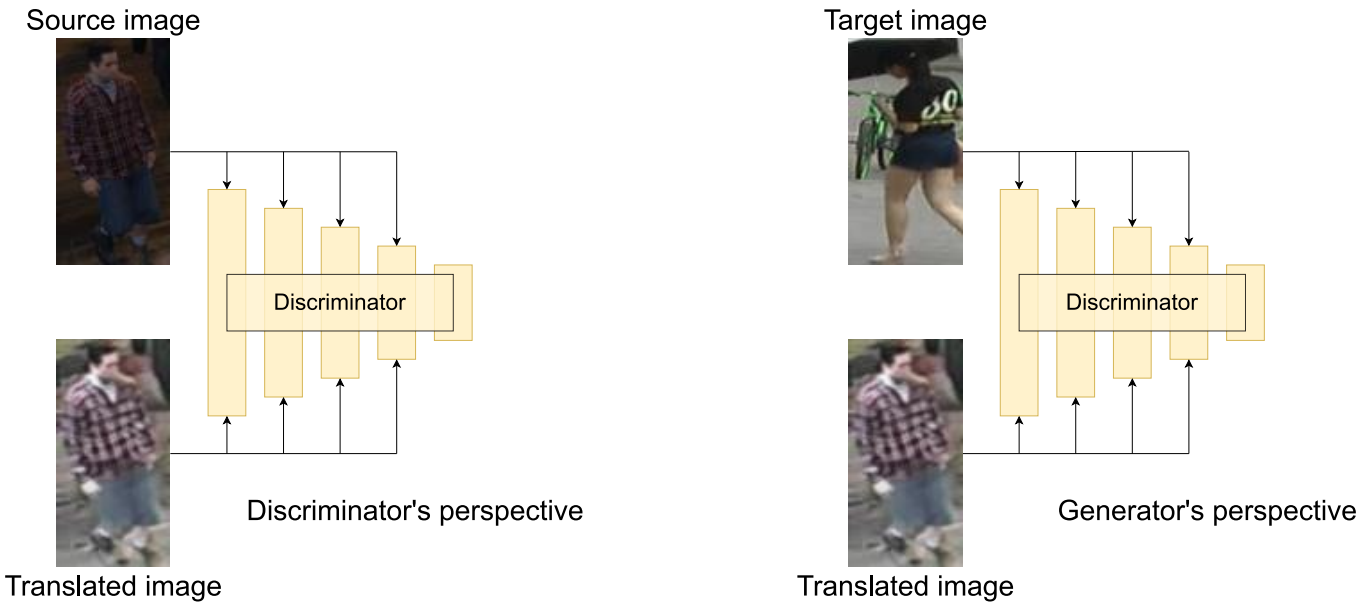


First row: paired image translation [3], second row: unpaired image translation [4].

Domain mapping

Contrastive unpaired translation framework [4] for translation.

- Generator-discriminator structure.
- *PatchGan* discriminator [3].
- Designed a feature matching loss for the discriminator.



Target image is from Market [2].

Relationship preservation

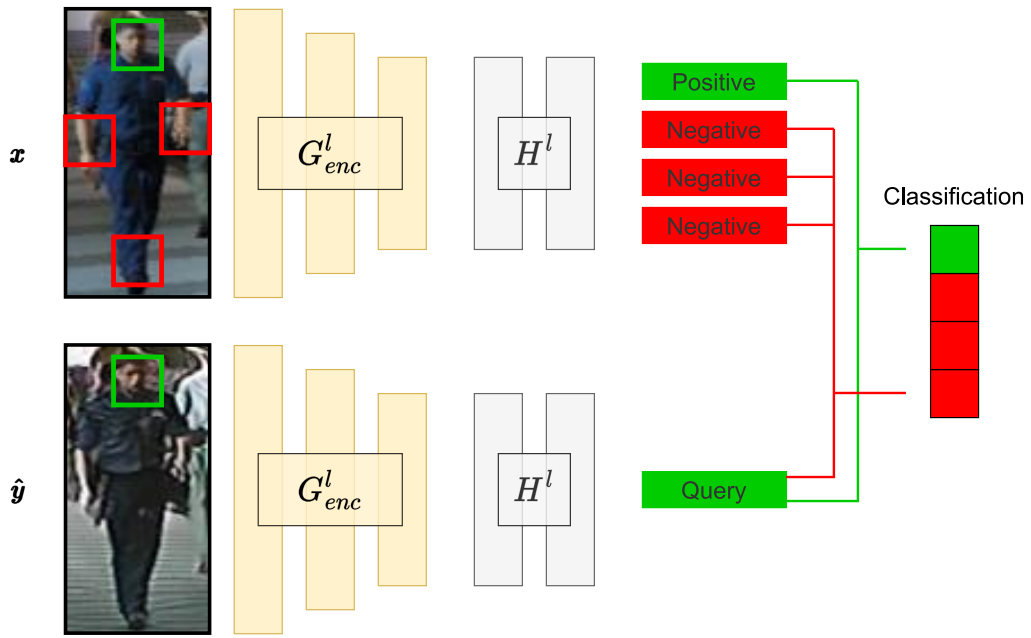
Solely relying on a generator-discriminator architecture does not preserve the original content.

Figure 1: Pairs of original and translated pictures with the basic domain mapping.

Relationship preservation

Constraining the image generation process.

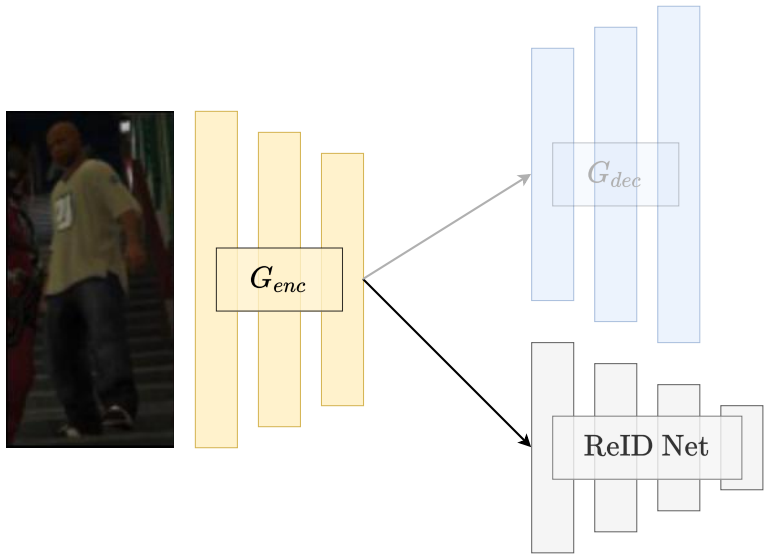
- One-way translation without cycle consistency.
- Corresponding patches of input-output images should be similar.



Discriminative learning

Joint discriminative feature learning with image translation.

- The generator encoder learns features responsible for the translation but they cannot recognize pedestrian identities.
- The last activations of the encoder are passed to a network for standard classification with the source identities.



Existing works

Evaluation on Market, Duke and CUHK03.

- Perform better than earlier domain-transfer methods without having to rely on cycle consistency.
- Evaluated each GTASynthReid-to-target adaptation also on the remaining real-world dataset.

Model	**Source**	Market	
		MaP	R1
SPGAN[5]	Duke	26.9	58.1
Cyclegan[6]	Duke	24.5	52.0
CR-GAN[7]	Duke	33.2	64.5
DG-Net++[8]	MSMT17	64.6	83.1
GLC[9]	Duke	75.4	90.5
Resnet50[10][11]	RandPerson	28.8	55.6
JVTC[12][13]	UnrealPerson	80.2	93.0
Ours	Market	**42.6**	**61.2**

Model	**Target**	Market	
		MaP	R1
Baseline	-	23.3	39.4
Ours	Market	**42.6**	**61.2**
Ours	Duke	40.7	59.5
Ours	CUHK03	41.6	60.0

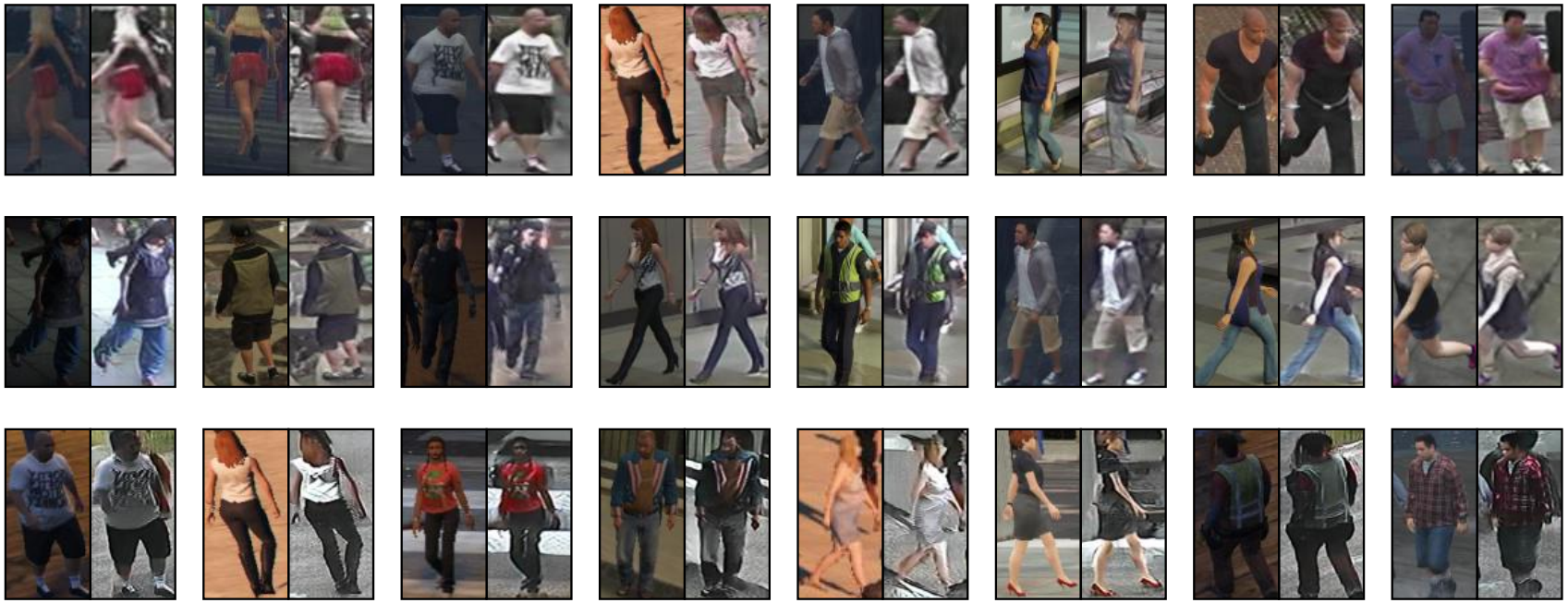
Qualitative results

The resulting pictures are closer to the real-world data.

- Computed Fréchet Inception Distance before translation.
- Computed Fréchet Inception Distance after translation.

Translation	Target	FID
GTASynthReid	Market	45.14
GTASynthReid-to-Market	Market	**24.29**
GTASynthReid	Duke	50.44
GTASynthReid-to-Duke	Duke	**19.53**
GTASynthReid	CUHK03	63.23
GTASynthReid-to-CUHK03	CUHK03	**31.54**

Qualitative results



Original (left) and translated (right) images from our GTASynthReid to Market, Duke and CUHK03 (first, second and third rows).

Remarks

- Extended a Pytorch framework [14] for person re-identification to allow learning via translation from synthetic data.
- Employed the contrastive unpaired translation framework [4] for pedestrian re-identification.
- Designed a feature matching loss that increased performance.
- Achieved better results than some earlier methods with a simpler model.
- Translated images capture the style of the target data.

Future directions

- Design more pedestrians to embed in the video game.
- Develop a method for pose and appearance disentanglement without cycle consistency or external information.
- The adopted framework could be inconsistent with this kind of objective.

Thank You

[1] L. Zheng, H. Zhang, S. Sun, M. Chandraker, and Q. Tian, "Person re-identification in the wild," *arXiv preprint arXiv:1604.02531*, 2016.

[2] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Computer Vision, IEEE International Conference on*, 2015.

[3] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[4] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive learning for unpaired image-to-image translation," in *European Conference on Computer Vision*, 2020.

[5] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang, and J. Jiao, "Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification," in *CVPR*, 2018.

[6] H. Tang, Y. Zhao, and H. Lu, "Unsupervised person re-identification with iterative self-supervised domain adaptation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1536–1543, 2019.

[7] Y. Chen, X. Zhu, and S. Gong, "Instance-guided context rendering for cross-domain person re-identification," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 232–242, 2019.

[8] Y. Zou, X. Yang, Z. Yu, B. Vijayakumar, and J. Kautz, "Joint disentangling and adaptation for cross-domain person re-identification," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

[9] H. Chen, Y. Wang, B. Lagadec, A. Dantcheva, and F. Bremond, "Joint generative and contrastive learning for unsupervised person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2004–2013, June 2021.

[10] Y. Wang, S. Liao, and L. Shao, "Surpassing Real-World Source Training Data: Random 3D Characters for Generalizable Person Re-Identification," in *28th ACM International Conference on Multimedia (ACMMM)*, 2020.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[12] T. Zhang, L. Xie, L. Wei, Z. Zhuang, Y. Zhang, B. Li, and Q. Tian, "Unrealperson: An adaptive pipeline towards costless person re-identification," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[13] J. Li and S. Zhang, "Joint visual and temporal consistency for unsupervised domain adaptive person re-identification," in *ECCV*, 2020.

[14] K. Zhou, Y. Yang, A. Cavallaro, and T. Xiang, "Learning generalisable omni-scale representations for person re-identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021.