Self-supervised learning for fully unsupervised re-identification in real-world applications

Gabriel C. Bertocco[1], **Fernanda A. Andaló**[1], **Anderson Rocha**[1]

[1]Institute of Computing – University of Campinas (UNICAMP)
13083-852 – Campinas – SP – Brazil

gabriel.bertocco@ic.unicamp.br, feandalo@ic.unicamp.br, arrocha@unicamp.br

Abstract. *Re-Identification (ReID) enables real-world applications such as AI-powered surveillance, criminal identification, event understanding, and smart city development. However, it remains challenging due to occlusions, viewpoint changes, and background similarities. Supervised methods perform well but rely on costly, biased annotations, limiting scalability. To address this, we propose self-supervised algorithms for Unsupervised ReID (U-ReID), extendable to modalities such as Text Authorship Verification, tackling high intra-class variation and low inter-class distinction. Our work introduces three fully unsupervised ReID methods: one using camera labels, one without side information, and one scalable to large datasets. We also present a fourth hybrid method for long-range recognition under distortions. These solutions enhance generalization and enable real-world applications in forensics and biometrics.*

1. Introduction

Re-Identification (ReID) is crucial in applications like crime investigation and surveillance, aiming to match whole-body images of the same individual, or images of the same object, across cameras under varying conditions. It faces challenges like illumination changes, occlusions, viewpoint variations, and background similarities, leading to high intra-class discrepancy and low inter-class similarity. Traditional ReID methods rely on supervised learning, which demands costly, time-consuming, and error-prone annotations that introduce bias and hinder generalization. As a result, Self-Supervised Learning (SSL) has gained attention for its ability to learn from unlabeled data. Methods like MoCo (Momentum Contrast) [He et al. 2020] and DINO (Self-Distillation with no labels) [Caron et al. 2021] use contrastive learning, while others employ feature disentangling or clustering. Although SSL matches supervised performance, it often relies on datasets like ImageNet [Deng et al. 2009], where coarse class distinctions and simple augmentations suffice. In contrast, real-world tasks like ReID require attention to fine-grained details, making SSL more challenging.

This thesis presents four Unsupervised Re-Identification (U-ReID) solutions, each enhancing performance and generalization. The first method uses unsupervised domain adaptation with triplet creation and self-ensembling. The second removes the need for camera labels, applying ensemble-based knowledge combination and clustering strategies, and extends to tasks such as Text Authorship Attribution. The third method addresses large-scale U-ReID scenarios for persons (U-PReID) and vehicles (U-VReID) by introducing ReRanking and co-training label strategies. The fourth (design during the candidate's internship at the University of Colorado Colorado Springs, USA, and sponsored by IARPA[1]) focuses on long-range recognition for face and person re-identification, contributing to counterterrorism efforts. Each method advances U-ReID and related fields, with notable publications in leading journals and conferences. In summary, the main contributions of this thesis are:

- A method that adapts to an unknown, unlabeled target domain by leveraging meta-information (e.g., camera labels) and knowledge from a different source domain, using novel strategies such as cross-camera triplet-based learning, self-ensembling without human intervention, and a simple ensembling strategy for validation and deployment.
- A fully unsupervised solution that operates without meta-information or source domain initialization, utilizing neighborhood-based ensembling, clustering fusion to mitigate human bias, and a flexible model that can be applied to both Computer Vision (Unsupervised ReID) and NLP (Text Authorship Attribution).
- A large-scale, fully-unsupervised solution that learns from unlabeled data across re-identification tasks, featuring neighborhood-based sampling and re-ranking strategies, noise-aware clustering hyperparameter scheduling, and a co-training method for knowledge sharing without relying on complex supervision or hyperparameter tuning.
- A method to enhance robustness in long-range recognition tasks under atmospheric turbulence, with novel augmentation techniques, distortion-adaptive training, and feature magnitude-based model ensembling, improving performance for distorted data.

2. Related Work

Unsupervised Domain Adaptation (UDA) for ReID transfers knowledge from a labeled source domain to an unlabeled target domain and is typically categorized into generative, attribute alignment, and label-proposing. Generative methods [Li et al. 2019, Zou et al. 2020, Lin et al. 2020] synthesize data to bridge domain gaps. Attribute alignment methods [Qi et al. 2019, Wu et al. 2019] align soft-biometric attributes across domains. Label-proposing approaches [Tang et al. 2019, Fu et al. 2019] assign pseudo-labels to target samples using clustering algorithms. Memory-based models [Wang and Zhang 2020, Ge et al. 2020a] iteratively store and update feature representations. Some methods leverage metadata, like camera labels [Xuan and Zhang 2021, Wang et al. 2020a, Chen et al. 2021], or tracklets [Wang et al. 2020b, Wu et al. 2019]. Fully unsupervised methods [Chen et al. 2020, Ge et al. 2020b, Zhang et al. 2021, Cho et al. 2022] refine clustering without additional metadata.

Our proposed approaches, besides operating without annotations, introduce strategies to reduce hyperparameter sensitivity and integrate multi-model knowledge, improving adaptability to unlabeled datasets, which is not addressed by most of prior works. Moreover, our second method extends to fully unsupervised Text Authorship Attribution, identifying authorship without labels. While models such as BERT [Devlin et al. 2018], BERTweet [Nguyen et al. 2020], and T5 [Raffel et al. 2019] rely on self-supervised learning, our approach targets authorship attribution in unlabeled corpora. Compared to AdHominem [Boenninghoff et al. 2019a], which uses attention-based LSTMs, and others based on n-grams [Potha and Stamatatos 2014] or siamese networks [Boenninghoff et al. 2019b], our solution enhances flexibility and scalability for real-world applications. A more comprehensive literature review is provided in the thesis.

[1] https://www.iarpa.gov/research-programs/briar

3. Methodologies

We briefly describe the four proposed methods. Further details about publications and impact are in the subproduct report submitted along with this document.

3.1. Unsupervised and self-adaptative techniques for cross-domain person re-identification

In the first proposed method (Chapter 2 of the thesis), we aim to re-identify people in a camera system, assuming that we know from which camera each person has been recorded, i.e., we do not know who is present in a given frame, but we know the camera label. We start with a few (usually three) Deep Convolutional Neural Network (DCNN) models previously trained on another ReID dataset to learn initial task-related features. We propose an UDA algorithm to train these networks on the target unlabeled data, leveraging a novel Cross-Camera Triplet creation strategy on training, a self-ensembling strategy, and backbone ensembling during evaluation.

3.2. Leveraging ensembles and self-supervised learning for fully-unsupervised person re-identification and text authorship attribution

In the second solution (Chapter 3 of the thesis), we address a real-world deployment scenario by disregarding camera information and using the same backbones but without pre-training on any task-specific dataset. Only the person's bounding box is available (no identity or camera annotations) and the backbones are initialized with ImageNet weights. We propose a novel ensemble-based strategy that combines neighborhood-based distances between samples from each manifold into a single distance matrix, effectively ensembling distinct knowledge captured by each backbone. Additionally, we introduce a new ensemble-based clustering strategy that combines clustering results from different hyperparameter settings to produce clusters with lower false-positive rates. This solution does not rely on any task-specific metadata, generalizing it to domains beyond ReID.

To evaluate its generalization capability, we consider a second task in the Natural Language Processing (NLP) domain: Text Authorship Attribution (TAA) for short messages. The goal is to group short messages from the same author in a fully unsupervised manner, using only raw text as input. To the best of our knowledge, at the time of the proposal, this was the first attempt to apply the same self-supervised learning pipeline to different modalities with only minor adjustments for two forensic tasks. This pipeline outperforms the state of the art in U-ReID and shows promising results in text analysis.

3.3. Large-scale Fully-Unsupervised Re-Identification

In the third proposed method (Chapter 4 of the thesis), we maintain the same constraints as in the second solution but introduce novel strategies to handle large-scale scenarios, along with an extension to Unsupervised Vehicle Re-Identification (U-VReID). This solution includes several enhanced components: a self-supervised model pre-initialization on the target data, a sampling technique to reduce data size per iteration, a more efficient ReRanking technique suited for large-scale learning, clustering hyperparameter scheduling, and a co-training label strategy. It outperforms state-of-the-art methods that rely on full datasets for ReRanking and select hyperparameters based on the final query/gallery split. We argue that such prior work practices are unrealistic, as the standard assumption is that the data is fully unlabeled, making grid search for optimal hyperparameters infeasible in real-world deployments. In summary, this third pipeline targets large-scale learning through localized ReRanking, with reduced sensitivity to hyperparameter choices, and a co-training label strategy that improves clustering performance.

3.4. DaliID: Distortion-Adaptive Learned Invariance for Identification–a Robust Technique for Face Recognition and Person Re-Identification.

The fourth solution (Chapter 5 of the thesis) addresses a growing area in biometrics: long-range recognition. It was designed to perform Face Recognition and Person Re-Identification on images affected by varying levels of distortion, primarily caused by atmospheric turbulence. This collaborative work was developed during Gabriel's internship at the University of Colorado Colorado Springs (UCCS), USA, during which he was a member of the Biometric Recognition and Identification at Altitude and Range (BRIAR) program[2], a U.S. government-supported project focused on counterterrorism, critical infrastructure protection, transportation security, military force protection, and border security (sponsored by IARPA). Gabriel was one of the solution's designers and conducted all experiments, analyses, and conclusions related to the Person Re-Identification task, as well as supporting experiments and data collection for long-range Face Recognition. While not as unsupervised as the previously proposed solutions, this method is part of a broader framework that incorporates unsupervised techniques during evaluation to improve whole-body person matching performance.

4. Results

Following prior work, our results are reported using mean Average Precision (mAP) and rank-based metrics, including Rank-1 (R1), Rank-5 (R5), and Rank-10 (R10). The evaluated datasets include `Market1501` [Zheng et al. 2015], `DukeMTMC-ReID` [Ristani et al. 2016][3], `MSMT17` [Wei et al. 2018], `DeepChange` [Xu and Zhu 2023], `Veri776` [Liu et al. 2016b], `VehicleID` [Liu et al. 2016a], and `Veri-Wild` [Lou et al. 2019]. For text analysis, we employ [Theophilo et al. 2021]. Additional details are provided in Appendix B of the thesis. Due to space constraints, we present condensed result tables, focusing on the strongest competitors, while full tables and analyses are included in the thesis.

4.1. Results of the first method

Our first method demonstrates superior performance in the more challenging adaptation scenarios, where difficulty is determined by the number of cameras in the dataset. The most complex adaptation, `Market` → `MSMT17`, involves transitioning from a controlled environment (6 cameras, same day period and season) to a highly diverse setting (15 cameras, both indoors and outdoors, recorded across different times of the day and seasons). As shown in Table 1, our method outperforms the state of the art by 1.5 and 2.1 percentage points (p.p.) in mAP and Rank-1, respectively, for `Duke` → `MSMT17`, and by 2.2 and 4.2 p.p. for the most challenging case, `Market` → `MSMT17`.

[2] https://www.iarpa.gov/research-programs/briar

[3] Due to redaction, this dataset is **not** used for evaluation in the third and fourth solutions. More details are available at *https://www.dukechronicle.com/article/2019/06/duke-university-facial-recognition-data-set-study-surveillance-video-students-china-uyghur.*

This success is attributed to our explicit design for handling camera diversity by constructing triplets based on different camera views within a cluster. Additionally, our approach uses a simpler training process with only one hyperparameter (triplet loss margin), whereas many prior works rely on complex loss functions with multiple hyperparameters, often tuned specifically for Duke → Market and Market → Duke adaptations. This reliance on predefined hyperparameters may introduce bias, while our method remains robust across diverse adaptation setups. Further details and more comprehensive comparison to prior works are provided in Chapter 2 of the thesis.

Table 1. Results on Market1501 to MSMT17 and DukeMTMCRe-ID to MSMT17 adaptation scenarios. The best result is shown in blue, the second in green, and the third in orange. RR means ReRanking.

		Duke → MSMT17				Market → MSMT17			
Method	reference	mAP	R1	R5	R10	mAP	R1	R5	R10
SSKD [Liu et al. 2021]	NDIC'21	26.0	53.8	**66.6**	**72.0**	23.8	49.6	63.1	68.8
ABMT [Chen et al. 2020]	WACV'20	**33.0**	**61.8**	-	-	27.8	55.5	-	-
SpCL [Ge et al. 2020b]	NeurIPS'20	-	-	-	-	**31.0**	**58.1**	**69.6**	**74.1**
Ours (w/o RR)	This Work	**34.5**	**63.9**	**75.3**	**79.6**	**33.2**	**62.3**	**74.1**	**78.5**
Ours (w/ RR)	This Work	**46.6**	**69.6**	**77.1**	**80.4**	**45.2**	**68.1**	**76.0**	**79.2**

4.2. Results of the second method

Our results in Table 2 highlight the effectiveness of our fully unsupervised method, which operates without identity labels or meta-information. We achieve state-of-the-art performance on the most challenging datasets, Duke and MSMT17, and obtain the second-best result on Market in terms of mAP and Rank-1. Designed to handle complex, fully unlabeled multi-modal scenarios, our method excels on difficult datasets, even though some existing Person ReID methods perform better in simpler cases like Market.

Many existing works leverage metadata such as camera labels and tracklets. As explained in the thesis, camera information significantly boosts performance, as it naturally involves cross-camera retrieval. This is particularly evident when comparing our results with the camera-based method PPLR [Cho et al. 2022], which surpasses ours by 5.1 p.p. in Rank-1 on MSMT17 (result presented in the thesis). However, our approach still achieves the highest mAP on this dataset, demonstrating its ability to retrieve more true positive samples closer to the query. Notably, our method delivers the best overall performance, even when compared to methods that utilize **strong camera metadata**.

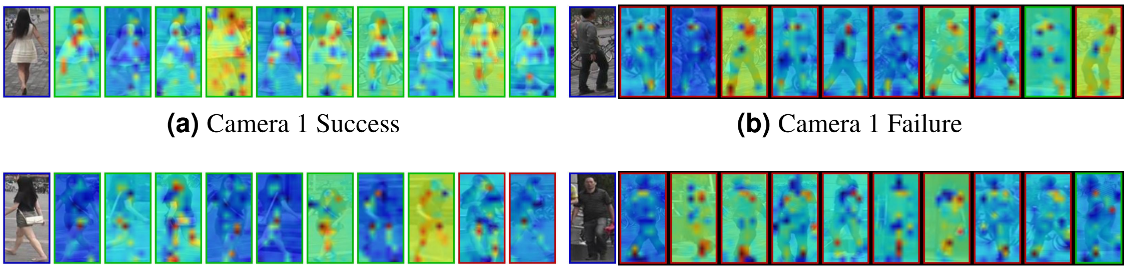
We also address the Text Authorship Attribution (TAA) task by adapting BERT [Devlin et al. 2018], BERTweet [Nguyen et al. 2020], and T5 [Raffel et al. 2019]. The employed dataset has two test sets for evaluation [Theophilo et al. 2021]. Our method is compared against a prior work employing a **supervised** Siamese network model. Despite being fully unsupervised, our approach surpasses the supervised method by 7.0 and 24.5 p.p. in mAP and R1 on the first subset, and by 2.6 and 11.7 p.p. on the second. This demonstrates that our method effectively reduces the reliance on labeled data.

Figure 1 shows one success and one failure case for two cameras from the Market1501 dataset. We observe that the model can mine fine-grained details across all images, regardless of the camera, and retrieve true positive samples within the top 10 results. Failure cases are mostly due to visual similarity between identities (e.g., similar

Table 2. Comparison with relevant fully-unsupervised Person ReID methods. The best one is in blue, the second best in green, and the third in orange. Full table version is provided in the thesis.

		Market				Duke				MSMT17			
Method	Reference	mAP	R1	R5	R10	mAP	R1	R5	R10	mAP	R1	R5	R10
Fully Unsupervised													
		Market				Duke				MSMT17			
Method	Reference	mAP	R1	R5	R10	mAP	R1	R5	R10	mAP	R1	R5	R10
HCT [Zeng et al. 2020]	CVPR'20	56.4	80.0	91.6	95.2	50.7	69.6	83.4	**87.4**	-	-	-	-
RLCC [Zhang et al. 2021]	CVPR'21	77.7	90.8	96.3	97.5	**69.2**	**83.2**	**91.6**	**93.8**	27.9	56.5	68.4	73.1
ICE [Chen et al. 2021]	ICCV'21	79.5	92.0	97.0	**98.1**	67.2	81.3	90.1	**93.0**	29.8	59.0	71.7	77.0
CACL [Li et al. 2022]	TIP'22	80.9	92.7	**97.4**	**98.5**	**69.6**	**82.6**	**91.2**	**93.8**	23.0	48.9	61.2	66.4
PPLR [Cho et al. 2022]	CVPR'22	**81.5**	**92.8**	**97.1**	**98.1**	-	-	-	-	**31.4**	**61.1**	**73.4**	**77.8**
ISE [Zhang et al. 2022b]	CVPR'22	**84.7**	**94.0**	**97.8**	**98.8**	-	-	-	-	**35.0**	**64.7**	**75.5**	**79.4**
Ours		**83.4**	**92.9**	**97.1**	97.8	**72.7**	**83.9**	**91.0**	**93.0**	**42.6**	**68.2**	**77.9**	**81.4**

T-shirts and hair). Further details and more comprehensive comparison to prior works are provided in Chapter 3 of the thesis.



(a) Camera 1 Success

(b) Camera 1 Failure

(c) Camera 3 Success

(d) Camera 3 Failure

Figure 1. Success and failure cases considering one query (leftmost image) from two cameras on `Market`. Green border means true positive and red border means false positive samples.

4.3. Results of the third method

Our method is compared to prior works in Table 3. We highlight methods that tune clustering parameters per dataset, as they are not applicable in real-world fully unsupervised settings. AdaMG [Peng et al. 2023] achieves good results on `Market` and `MSMT17` but suffers performance drops when using the same parameter across datasets. Our ε scheduling scheme outperforms AdaMG by 1.2 and 0.1 p.p. in mAP and R1 on `Market`, and by 5.2 and 4.6 p.p. on `MSMT17`. We also compare with ensemble-based methods discussed in Chapter 4 of the thesis.

In addition, our method outperforms prior works in Vehicle Re-Identification. On the `Veri-Wild` dataset, in the most challenging setup (`VW-Large`), we achieve improvements of 4.5, 1.7, and 0.6 p.p. in mAP, R1, and R5, respectively, using only 75% of the data. Figure 2 presents qualitative results on the `Veri` dataset. In successful matches, our model learns fine-grained, point-of-view-invariant features, focusing on specific discriminant regions while remaining robust to background variations. Activation maps highlight key regions, with minimal activation in the background. In failure cases, the model retrieves visually similar images that are difficult to distinguish, even for humans. Further details and more comprehensive comparison to prior works are provided in Chapter 4 of the thesis.

Table 3. Comparison of Person ReID methods, with the best result in blue, the second best in green, and the third in orange. RRMC denotes Re-Ranking Memory Complexity, while CPD indicates whether the method requires dataset-specific clustering parameters. (p%) represents the proportion of data points sampled in Local Neighborhood Sampling per epoch.

				Market				MSMT17			
Method	Reference	RRMC	CPD	mAP	R1	R5	R10	mAP	R1	R5	R10
CCons [Dai et al. 2022]	ACCV'22	$\mathcal{O}(N^2)$	No	83.0	92.9	97.2	98.0	33.0	62.0	71.8	76.7
ISE [Zhang et al. 2022b]	CVPR'22	-	Yes	84.7	94.0	97.8	98.8	35.0	64.7	75.5	79.4
HHCL [Hu et al. 2021]	NIDC'21	$\mathcal{O}(N^2)$	No	84.2	93.4	97.7	98.5	-	-	-	-
GRACL [Zhang et al. 2022a]	TCSVT'22	$\mathcal{O}(N^2)$	No	83.7	93.2	97.6	98.6	34.6	64.0	75.0	79.3
AdaMG [Peng et al. 2023]	TCSVT'23	$\mathcal{O}(N^2)$	Yes	84.6	93.9	97.9	98.9	38.0	66.3	76.9	80.6
Ours (50%)		$\mathcal{O}(kN)$	No	-	-	-	-	24.3	50.4	60.6	65.4
Ours (75%)		$\mathcal{O}(kN)$	No	82.9	92.6	97.0	97.8	39.3	67.3	77.3	80.8
Ours (100%)		$\mathcal{O}(kN)$	No	85.8	94.0	97.7	98.5	43.2	70.9	80.8	84.2

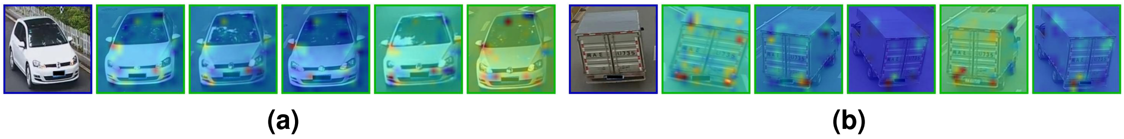


Figure 2. Activation maps for the top-5 images retrieved from the gallery, given a query image (blue border) in the `Veri` dataset.

4.4. Results of the fourth method

Our method DaliID achieves the best performance on the `Market1501` dataset, outperforming prior work by 0.8 p.p. in mAP, and tying for second place (along with FIDI) with a R1 of 94.5%. On `MSMT17`, the most challenging Person ReID benchmark, we achieve the best performance, surpassing prior works by 5.9 and 3.2 p.p. in mAP and R1, respectively. With OSNet, we obtain the best performance on both datasets in both metrics.

To show the generalization ability, we trained DaliID on `DeepChange`, where subjects wear different clothes across views. We outperform recent prior work by 2.9 and 6.8 p.p. in mAP and R1, respectively. In addition to clothing changes, `DeepChange` presents more distortions and lower-quality data than `Market` and `MSMT17`. Our method achieves the highest R1 gain and the second-highest mAP gain (after `MSMT17`), indicating that it is especially effective in low-quality conditions. Further details and more comprehensive comparison to prior works are provided in Chapter 5 of the thesis.

5. Conclusion and Future Work

Our methods address key challenges in fully unsupervised re-identification, offering adaptable solutions for forensic and biometric applications. A potential extension involves integrating our approaches into broader investigative pipelines to map relationships between individuals, vehicles, and locations. Advances in Large Language Models and Large Vision Models based on Transformers offer a promising direction for improving feature extraction in re-identification tasks. Additionally, recent self-supervised clustering techniques in DeepFake detection suggest applications in the analysis of synthetic media and other forms of synthetic reality.

Our contributions span event investigation, smart security, and biometrics, introducing three self-supervised learning algorithms and a hybrid supervised-unsupervised method. Future research directions include integrating Large Vision Models, developing explainable AI techniques, and addressing other real-world deployment constraints. Overall, these findings contribute to fields of biometrics and forensic science, while also enabling extensions into broader AI and Computer Science research. Please, check the six-page subproduct report to see direct uses and further applications of this research.

References

Boenninghoff, B., Hessler, S., Kolossa, D., and Nickel, R. M. (2019a). Explainable authorship verification in social media via attention-based similarity learning. In *Int. Conf. Big Data*, pages 36–45.

Boenninghoff, B., Nickel, R. M., Zeiler, S., and Kolossa, D. (2019b). Similarity learning for authorship verification in social media. In *IEEE Int. Conf. on Acoust., Speech Signal Process.*, pages 2457–2461.

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., and Joulin, A. (2021). Emerging properties in self-supervised vision transformers. *arXiv preprint*, arXiv:2104.14294.

Chen, H., Lagadec, B., and Bremond, F. (2020). Enhancing diversity in teacher-student networks via asymmetric branches for unsupervised person re-identification. In *Winter Conf. Appl. Comput. Vis.*, pages 1–10.

Chen, H., Lagadec, B., and Bremond, F. (2021). ICE: Inter-instance contrastive encoding for unsupervised person re-identification. In *Int. Conf. Comput. Vis.*, pages 14960–14969.

Cho, Y., Kim, W. J., Hong, S., and Yoon, S.-E. (2022). Part-based pseudo label refinement for unsupervised person re-identification. In *Conf. Comput. Vis. Pattern Recog.*, pages 7308–7318.

Dai, Z., Wang, G., Yuan, W., Zhu, S., and Tan, P. (2022). Cluster contrast for unsupervised person re-identification. In *Asian Conf. Comput. Vis.*, pages 1142–1160.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *Conf. Comput. Vis. Pattern Recog.*, pages 248–255.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint*, arXiv:1810.04805.

Fu, Y., Wei, Y., Wang, G., Zhou, Y., Shi, H., and Huang, T. S. (2019). Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. In *Int. Conf. Comput. Vis.*, pages 6112–6121.

Ge, Y., Chen, D., and Li, H. (2020a). Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. *arXiv preprint*, arXiv:2001.01526.

Ge, Y., Chen, D., Zhu, F., Zhao, R., and Li, H. (2020b). Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. *arXiv preprint*, arXiv:2006.02713.

He, K., Fan, H., Wu, Y., Xie, S., and Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. In *Conf. Comput. Vis. Pattern Recog.*, pages 9729–9738.

Hu, Z., Zhu, C., and He, G. (2021). Hard-sample guided hybrid contrast learning for unsupervised person re-identification. In *IEEE Int. Conf. Netw. Intell. Digit. Content*, pages 91–95.

Li, M., Li, C.-G., and Guo, J. (2022). Cluster-guided asymmetric contrastive learning for unsupervised person re-identification. *IEEE Trans. Image Process.*, 31:3606–3617.

Li, Y.-J., Lin, C.-S., Lin, Y.-B., and Wang, Y.-C. F. (2019). Cross-dataset person re-identification via unsupervised pose disentanglement and adaptation. In *Int. Conf. Comput. Vis.*, pages 7919–7929.

Lin, Y., Wu, Y., Yan, C., Xu, M., and Yang, Y. (2020). Unsupervised person re-identification via cross-camera similarity exploration. *IEEE Trans. Image Process.*, 29:5481–5490.

Liu, H., Tian, Y., Yang, Y., Pang, L., and Huang, T. (2016a). Deep relative distance learning: Tell the difference between similar vehicles. In *Conf. Comput. Vis. Pattern Recog.*, pages 2167–2175.

Liu, W., Nie, S., Yin, J., Wang, R., Gao, D., and Jin, L. (2021). Sskd: Self-supervised knowledge distillation for cross domain adaptive person re-identification. In *2021 7th IEEE International Conference on Network Intelligence and Digital Content (IC-NIDC)*, pages 81–85. IEEE.

Liu, X., Liu, W., Ma, H., and Fu, H. (2016b). Large-scale vehicle re-identification in urban surveillance videos. In *IEEE Int. Conf. Multimedia Expo*, pages 1–6.

Lou, Y., Bai, Y., Liu, J., Wang, S., and Duan, L. (2019). Veri-wild: A large dataset and a new method for vehicle re-identification in the wild. In *Conf. Comput. Vis. Pattern Recog.*, pages 3235–3243.

Nguyen, D. Q., Vu, T., and Nguyen, A. T. (2020). BERTweet: A pre-trained language model for english tweets. *arXiv preprint*, arXiv:2005.10200.

Peng, J., Jiang, G., and Wang, H. (2023). Adaptive memorization with group labels for unsupervised person re-identification. *IEEE Trans. Circuits Syst. Video Technol.*, pages 1–1.

Potha, N. and Stamatatos, E. (2014). A profile-based method for authorship verification. In *Hellenic Conf. AI*, pages 313–326.

Qi, L., Wang, L., Huo, J., Zhou, L., Shi, Y., and Gao, Y. (2019). A novel unsupervised camera-aware domain adaptation framework for person re-identification. In *Int. Conf. Comput. Vis.*, pages 8080–8089.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. (2019). Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint*, arXiv:1910.10683.

Ristani, E., Solera, F., Zou, R., Cucchiara, R., and Tomasi, C. (2016). Performance measures and a data set for multi-target, multi-camera tracking. In *Eur. Conf. Comput. Vis.*, pages 17–35.

Tang, H., Zhao, Y., and Lu, H. (2019). Unsupervised person re-identification with iterative self-supervised domain adaptation. In *Conf. Comput. Vis. Pattern Recog. Workshops*, pages 1536–1543.

Theophilo, A., Giot, R., and Rocha, A. (2021). Authorship attribution of social media messages. *IEEE Trans. Comput. Social Syst.*

Wang, D. and Zhang, S. (2020). Unsupervised person re-identification via multi-label classification. In *Conf. Comput. Vis. Pattern Recog.*, pages 10981–10990.

Wang, M., Lai, B., Huang, J., Gong, X., and Hua, X.-S. (2020a). Camera-aware proxies for unsupervised person re-identification. *arXiv preprint*, arXiv:2012.10674.

Wang, Z., Zhang, J., Zheng, L., Liu, Y., Sun, Y., Li, Y., and Wang, S. (2020b). CycAs: Self-supervised cycle association for learning re-identifiable descriptions. In *Eur. Conf. Comput. Vis.*, pages 72–88.

Wei, L., Zhang, S., Gao, W., and Tian, Q. (2018). Person transfer GAN to bridge domain gap for person re-identification. In *Conf. Comput. Vis. Pattern Recog.*, pages 79–88.

Wu, J., Yang, Y., Liu, H., Liao, S., Lei, Z., and Li, S. Z. (2019). Unsupervised graph association for person re-identification. In *Int. Conf. Comput. Vis.*, pages 8321–8330.

Xu, P. and Zhu, X. (2023). Deepchange: A long-term person re-identification benchmark with clothes change. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11196–11205.

Xuan, S. and Zhang, S. (2021). Intra-inter camera similarity for unsupervised person re-identification. In *Conf. Comput. Vis. Pattern Recog.*, pages 11926–11935.

Zeng, K., Ning, M., Wang, Y., and Guo, Y. (2020). Hierarchical clustering with hard-batch triplet loss for person re-identification. In *Conf. Comput. Vis. Pattern Recog.*, pages 13657–13665.

Zhang, H., Zhang, G., Chen, Y., and Zheng, Y. (2022a). Global relation-aware contrast learning for unsupervised person re-identification. *IEEE Trans. Circuits Syst. Video Technol.*, 32(12):8599–8610.

Zhang, X., Ge, Y., Qiao, Y., and Li, H. (2021). Refining pseudo labels with clustering consensus over generations for unsupervised object re-identification. In *Conf. Comput. Vis. Pattern Recog.*, pages 3436–3445.

Zhang, X., Li, D., Wang, Z., Wang, J., Ding, E., Shi, J. Q., Zhang, Z., and Wang, J. (2022b). Implicit sample extension for unsupervised person re-identification. In *Conf. Comput. Vis. Pattern Recog.*, pages 7369–7378.

Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., and Tian, Q. (2015). Scalable person re-identification: A benchmark. In *Int. Conf. Comput. Vis.*, pages 1116–1124.

Zou, Y., Yang, X., Yu, Z., Kumar, B., and Kautz, J. (2020). Joint disentangling and adaptation for cross-domain person re-identification. *arXiv preprint*, arXiv:2007.10315.