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

Gabriel Bertocco[1], Fernanda Andaló, Anderson Rocha
Recod.ai Lab, University of Campinas (UNICAMP), Campinas, São Paulo 13083-852
Emails: {gabriel.bertocco, feandalo, anderson.rocha}@ic.unicamp.br

***Abstract*—Re-Identification (ReID) is vital for real-world applications such as AI-powered security, event understanding, and smart city development. It aims to retrieve all instances of a given person or object across a network of non-overlapping cameras, based solely on visual appearance. It is 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 learning algorithms for Unsupervised ReID, extendable to other modalities, like Text Authorship Verification, marked by 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 larger 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. We have released open-source code and demonstrated practical impact, including a consultancy project for the Public Prosecutor's Office of the State of São Paulo (MPSP), and House of Representatives (Brazilian Federal Chamber and Senate). This research was recognized by the Brazilian Computing Society (SBC) as the best Ph.D. thesis defended in Brazil in 2024.**

I. Introduction

Re-Identification (ReID) is crucial in forensics and biometrics applications, aiming to match whole-body images of the same individual, or 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 biased annotations that hinder generalization. As a result, Self-Supervised Learning (SSL) has gained attention for its ability to learn from unlabeled data. Methods like MoCo [1] and DINO [2] use contrastive learning, while others employ feature disentangling or clustering. Although some SSL matches supervised performance, it often relies on datasets like ImageNet [3], where coarse class distinctions and simple augmentations suffice. ReID, in contrast, demands attention to fine-grained details.

ReID aims to retrieve the same person (or object) across a network of cameras. From one camera to another, the same person may appear under different illumination (e.g., inside vs. outside a building), occlusions, viewpoints (e.g., back vs. front view), resolutions (due to varying distances and pixel density), and backgrounds. Moreover, people in the same environment (e.g., an airport) often dress similarly (carrying backpacks, bags, or luggage), making different classes semantically closer. As a result, Person ReID (PReID) faces high intra-class discrepancy and high inter-class similarity, as depicted in Figure 1. Considering a similar task, Vehicle Re-Identification (VReID), in which the "identities" are vehicles, the same challenges are faced. A vehicle can be recorded under different environmental conditions and the license plate is not always visible.

Fig. 1. Illustration of typical challenges in PReID: The first row shows high intra-class variance, with the same person varying in pose, resolution, background, and lighting. The second row highlights low inter-class disparity, where different individuals wear similar clothing and differ only in subtle features like shoes, hair, backpacks, or faces. Images reproduced from [4].

Unlike in ImageNet, PReID and VReID has test identities (query and gallery) disjoint from the training set, and their main datasets has a number of classes comparable to ImageNet [5]. Additionally, on real deployment, the person (or vehicle) images come from heterogeneous domains (surveillance, mobiles, or media cameras), requiring a system that handles multi-domain data in a fully-unsupervised way.

This research stands out for its strong real-world applicability, with fully unsupervised ReID methods already deployed in high-impact scenarios. These techniques have been used in consultancy projects with police and public prosecutor offices around the world (detailed herein), demonstrating their value in enhancing public safety and security. They can help authorities identify individuals across camera networks in events like the U.S. Capitol attack in 2021, or Brazil's January 8 riots, in 2023. Beyond investigations, the methods support biometric behavior analysis in public spaces, aiding research on crowd dynamics and preventive safety. As noted by the BBC, improved crowd monitoring could help prevent tragedies. These solutions also align with strategic needs in high-security domains such as infrastructure protection, surveillance, and counter-terrorism [Link]. This is reinforced by initiatives like IARPA's VIDEO_LINCS, which seeks fully-unsupervised ReID in open-world conditions without prior knowledge, highlighting the timely importance of fully unsupervised ReID, the core contribution of this Ph.D. research.

[1]This is a short summary of Gabriel Bertocco's Ph.D. thesis.

This thesis presents four Unsupervised Re-Identification (U-ReID) solutions. The first uses domain adaptation with triplet creation and self-ensembling. The second removes camera label dependence through ensemble-based clustering and extends to tasks like Text Authorship Attribution. The third scales to large U-PReID and U-VReID scenarios via ReRanking and co-training strategies. The fourth targets long-range recognition in face and person ReID, relevant to counterterrorism. All methods advance the state of the art, with results published in top journals and conferences.

The thesis was defended in English and evaluated by an international committee of world-renowned researchers: Prof. Dr. Patrick Flynn (Univ. Notre Dame/USA), Prof. Dr. Sébastien Marcel (IDIAP/Switzerland), Prof. Dr. Vitomir Štruc (Univ. of. Ljubljana/Slovenia) and Profa. Dra. Esther Luna Colombini (UNICAMP/Brazil).

The proposed methods have been applied in real-world cases, contributing to Brazil's FAPESP Déjàvu Project and the U.S. IARPA Biometric Recognition and Identification at Altitude and Range (BRIAR) program. They have been presented to the Federal Police of Brazil, the Forensics Department of the Police of Dubai, UAE, and used in investigations by the Public Prosecutor's Office of the State of São Paulo and Bahia in Brazil. The societal impact of this Ph.D. research was also featured in a UOL interview with Gabriel Bertocco.

In summary, the main contributions of this thesis are:

- A method for adapting to an unknown, unlabeled target domain by leveraging meta-information (e.g., camera labels) and knowledge from a separate source domain, using novel strategies such as cross-camera triplet-based learning, self-ensembling without human intervention, and a simple ensembling approach for deployment.
- A fully unsupervised solution that operates without meta-information or source domain initialization, employing neighborhood-based ensembling and clustering fusion to reduce human bias. This flexible model generalizes to both Computer Vision (Unsupervised ReID) and NLP (Text Authorship Attribution).
- A scalable, fully unsupervised approach for learning from unlabeled data across re-identification tasks, featuring neighborhood-based sampling, re-ranking, noise-aware clustering with hyperparameter scheduling, and a co-training strategy for knowledge sharing—without complex supervision or manual tuning.
- A method to improve robustness in long-range recognition under atmospheric turbulence, incorporating novel augmentation techniques, distortion-adaptive training, and feature magnitude-based model ensembling to enhance performance on degraded data.

Each proposed solution and its outcomes are detailed in the following sections.

II. Related Work

Unsupervised Domain Adaptation (UDA) for ReID transfers knowledge from a labeled source domain to an unlabeled target domain and is categorized into generative, attribute alignment, and label-proposing approaches. Generative methods [6], [7] synthesize data to bridge domain gaps. Attribute alignment methods [8], [9] align soft-biometric attributes across domains. Label-proposing methods [10], [11] assign pseudo-labels to target samples via clustering. Memory-based models [12], [13] iteratively store and update feature representations. Some methods leverage metadata, like camera labels [14], [15] or tracklets [9], [16]. In contrast, fully unsupervised methods [17]–[20] refine clustering without relying on metadata.

Our proposed approaches not only operate without annotations but also introduce strategies to reduce hyperparameter sensitivity and integrate multi-model knowledge, improving adaptability to unlabeled datasets, an aspect often overlooked in prior work. Moreover, our second method extends to fully unsupervised Text Authorship Attribution, enabling authorship identification without labeled data. While models such as BERT [21], BERTweet [22], and T5 [23] rely on self-supervised learning, our approach targets authorship attribution in unlabeled corpora. Compared to Ad-Hominem [24] (attention-based LSTMs), n-grams [25] or siamese networks [26], our solution enhances flexibility and scalability for real-world use. A detailed literature review is presented in the thesis.

III. Unsupervised and Self-Adaptative Techniques for Cross-Domain Person Re-Identification

The first proposed method (Chapter 2 of the thesis) addresses person re-identification in a camera system, assuming that camera labels are available, i.e., the identity of individuals in each frame is unknown, but the source camera is known. The approach begins with a small set (typically three) of Deep Convolutional Neural Network (DCNN) models pre-trained on a different ReID dataset to extract initial features. We then propose an Unsupervised Domain Adaptation algorithm (Figure 2) to adapt these models to the unlabeled target domain. This involves a novel Cross-Camera Triplet creation strategy during training, a self-ensembling mechanism, and backbone ensembling at evaluation. This work, with Gabriel as first author, was published in an IEEE Transactions journal (impact factor of 8.0) and presented at the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) workshops. All related achievements are listed in Section IX-A.

This solution can identify all occurrences of a suspect, victim, or person of interest within a monitored area, supporting investigations. It is applicable to real-world cases when camera sources are known.

IV. Leveraging Ensembles and Self-Supervised Learning for Fully-Unsupervised Person Re-Identification and Text Authorship Attribution

The second solution (Chapter 3 of the thesis) addresses a real-world deployment scenario by disregarding camera information and using the same backbones but without pre-training (Figure 3). Only the person's bounding boxes are available, without identity or camera annotations, and the backbones are initialized with ImageNet weights.

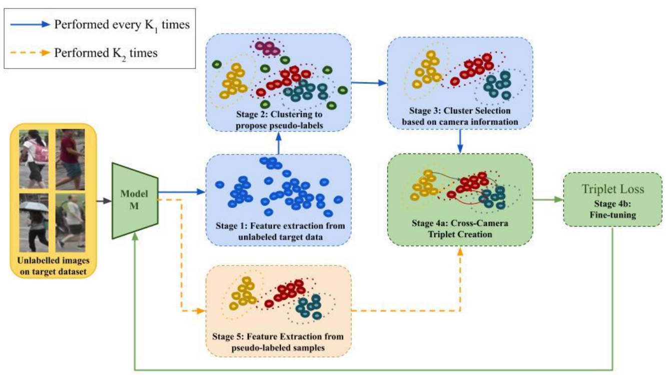

Fig. 2. The training pipeline uses camera information but no identity labels from the target domain. It alternates between two flows: the blue flow (every K_1 steps) includes feature extraction, clustering (OPTICS), cluster selection, and triplet training; the orange flow (every K_2 steps) reuses selected clusters for further training. The process iterates through blue-flow "iterations" and orange-flow "epochs" to progressively refine the model.

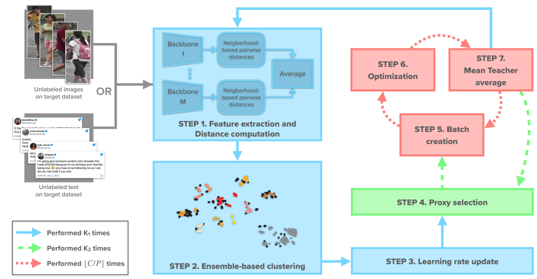

Fig. 3. The proposed approach consists of seven steps: feature extraction, distance calculation and refinement, and multi-backbone distance fusion. An ensemble-based clustering method then generates pseudo-labels and removes outliers. This is followed by warm-up learning rate scheduling, proxy selection for each cluster, and optimization using a Mean Teacher strategy.

We propose a novel ensemble-based strategy that merges neighborhood-based distances between samples from each backbone's manifold into a unified distance matrix, effectively combining diverse representations. Moreover, we introduce an ensemble-based clustering strategy that fuses clustering results across different hyperparameter values to reduce false positives. This solution requires no task-related meta-information, making it broadly applicable beyond PReID. To demonstrate this, we evaluate it on a second task in the Natural Language Processing (NLP) domain: Text Authorship Attribution (TAA) for short messages. The goal is to group tweets (short messages from Twitter, now X) by 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, we were the first to apply the same self-supervised learning pipeline across different modalities with only minor adjustments, addressing two forensic tasks with high intra-class disparity and inter-class similarity. This pipeline outperforms the state of the art in U-ReID and shows promising results in text analysis. With this work (Gabriel as first author), we published again in an IEEE Transactions journal (impact factor of 8.0), presented in the journal track of a leading biometrics conference, and achieved third place in a biometrics competition. All related achievements are listed in Section IX-B.

This solution is more flexible as it requires no meta-information. It can also detect short texts authored by the same individual on social media, supporting investigations and suspect identification.

V. Large-Scale Fully-Unsupervised Re-Identification

The third proposed method (Chapter 4 of the thesis) builds on the same constraints as the second solution but introduce new strategies to handle large-scale scenarios, including an extension to Unsupervised Vehicle Re-Identification (U-VReID). This solution incorporates several enhancements: a self-supervised model pre-initialization on the target data, a sampling technique to reduce per-iteration data size, a more efficient ReRanking technique for scalabiliy, clustering hyperparameter scheduling, and a co-training label strategy.

Our method outperforms existing state-of-the-art approaches, which often rely on unrealistic assumptions, such as access to the full dataset for re-ranking and tuning hyper-parameters based on the final query/gallery split. However, in real-world scenarios, data is fully unlabeled, making grid search for optimal hyperparameters impractical.

In summary, our third method addresses large-scale learning with local ReRanking, reduces hyperparameter sensitivity, and uses a co-training label strategy that improves clustering performance. This work (with Gabriel as first author) was published in an IEEE Transactions journal (impact factor of 5.0), featured in the "Noted in the Literature" section of the IEEE Biometrics Council Letter, and presented at InterForensics, the largest forensics conference in Latin America, promoted by the Federal Police of Brazil. We also designed a software interface (Fig. 4) made available to authorities in Brazil; our solution has already been requested by Public Prosecutor's Offices of Bahia and São Paulo. All achievements are listed in Section IX-C.

Fig. 4. System designed to help expert use of our third solution. It allows users to input surveillance videos and match a probe person (leftmost image) against candidate images. All images are from the publicly EvvE dataset [27].

This solution can detect and group occurrences of the same person or vehicle across different locations, supporting investigations and monitoring systems by uncovering potential relationships between people and vehicles in large-scale scenarios. Qualitative results are shown in Figures 5 and 6.

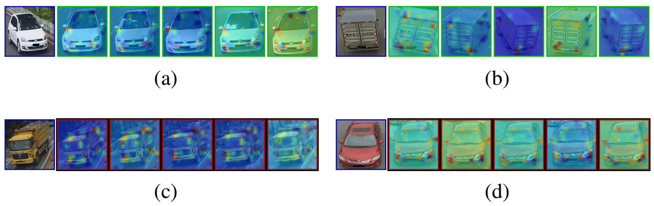

Fig. 5. Activation maps for the top-5 gallery images retrieved given a query image (blue border) from the `Veri` dataset. Images (a) and (b) show successful cases, and images (c), (d) failure cases. Visualizations were generated using the ResNet50 backbone.

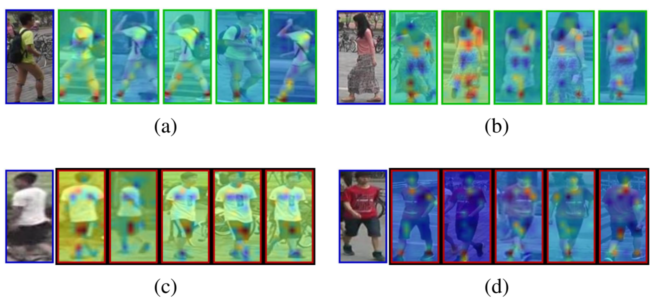

Fig. 6. Activation maps for the top-5 gallery images retrieved given a query image (blue border) from the `Market` dataset. Images (a) and (b) show successful cases, and images (c) and (d) failure cases. Visualizations were generated using the ResNet50 backbone.

VI. 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 long-range recognition. It is designed to perform Face Recognition and Person Re-Identification on images degraded by varying levels of distortion, primarily caused by atmospheric turbulence. It introduces DaliID, a distortion-adaptive learning framework. Its key contributions are: a novel distortion augmentation that combines spatial distortions and blur via an atmospheric turbulence simulator; an adaptive weighting mechanism that gradually increases the influence of heavily distorted samples during training; and a class center and proxy-based training for person re-identification. During inference, predictions from two backbones (one trained with distortions and one on clean data) are fused using a magnitude-weighted distance combination. An illustration is shown in Figure 7.

Gabriel led the entire Person Re-Identification research front and contributed to the experiments and data collection for long-range Face Recognition. He co-authored this work during his internship at the University of Colorado Colorado Springs (UCCS), USA, where he was part of the Biometric Recognition and Identification at Altitude and Range (BRIAR) program. BRIAR is a U.S. government-supported initiative sponsored by IARPA, focused on counterterrorism, critical infrastructure protection, transportation security, military force

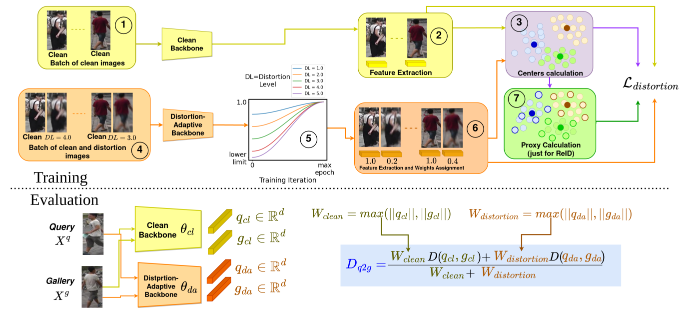

Fig. 7. The DaliID pipeline supports both face recognition (DaliFace) and person re-identification (DaliReID). Steps 1–3 cover training on clean data, while Steps 3–6 introduce distortion-adaptive learning. In Step 4, batches of clean and distorted images are built; Step 5 assigns dynamic weights based on distortion level (*DL*) to optimize the distortion loss ($\mathcal{L}_{distortion}$). Step 7, specific to DaliReID, addresses high intra-class variation. At inference time, predictions from the clean and adaptive models are fused for the final retrieval.

protection, and border security[1]. Our proposed solution can match individuals at long distances (up to 1 km), supports long-range monitoring and counterterrorism efforts.

VII. 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). Evaluated datasets include `Market1501` [4], `DukeMTMC-ReID` [28][2], `MSMT17` [29], `DeepChange` [30], `Veri776` [31], `VehicleID` [32], and `Veri-Wild` [33]. For text analysis, we use [34]. Additional details, along with full tables and analyses, are provided in the thesis.

A. Results of the first method

Our first method is superior in challenging adaptation scenarios, where difficulty is defined by the number of cameras. The most complex case, `Market` → `MSMT17`, involves transferring from an environment with 6 cameras (recorded on the same day and season) to one with 15 cameras spanning indoor and outdoor scenes (recorded across different days and seasons). As shown in Table I, our method outperforms the state of the art by 1.5 and 2.1 percentage points (p.p.) in mAP and R1, respectively, for `Duke` → `MSMT17`, and by 2.2 and 4.2 p.p. for the most challenging, `Market` → `MSMT17`. The success is due to our method's ability to handle camera diversity in clustering, combined with a simpler training process that uses only one hyperparameter. In contrast, many prior works use complex loss functions with multiple hyperparameters, often tuned specifically for specific setups, which can introduce bias. Further details are in Chapter 2 of the thesis.

B. Results of the second method

Our results in Table II highlight the effectiveness of our fully unsupervised method, which operates without identity labels or

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

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

TABLE I

RESULTS ON MARKET1501 TO MSMT17 AND DUKEMTMCRE-ID TO MSMT17 ADAPTATION SCENARIOS. THE BEST RESULT IS IN BLUE, THE SECOND IN GREEN, AND THE THIRD IN ORANGE. RR MEANS RERANKING.

		Duke → MSMT17		Market → MSMT17	
Method	reference	mAP	R1	mAP	R1
SSKD [35]	NDIC'21	26.0	53.8	23.8	49.6
ABMT [17]	WACV'20	33.0	61.8	27.8	55.5
SpCL [18]	NeurIPS'20	-	-	31.0	58.1
Ours (w/o RR)	This Work	34.5	63.9	33.2	62.3
Ours (w/ RR)	This Work	46.6	69.6	45.2	68.1

meta-information. It achieves the best performance on the most challenging datasets, Duke and MSMT17, and ranks second on Market. Designed to handle complex, fully unlabeled multimodal scenarios, our method excels on difficult datasets. While many existing approaches rely on metadata such as camera labels and tracklets, our method does not. Camera information can boost performance and this is evident when comparing our results with camera-based methods. Nevertheless, our approach achieves the highest mAP on MSMT17 and delivers the best overall performance, even against methods that use **strong camera metadata**.

TABLE II

COMPARISON WITH RELEVANT FULLY-UNSUPERVISED PERSON REID METHODS. THE BEST IS IN BLUE, THE SECOND BEST IN GREEN, AND THE THIRD IN ORANGE. FULL TABLE VERSION IS IN THE THESIS.

		Market		**Duke**		**MSMT17**	
Method	Reference	mAP	R1	mAP	R1	mAP	R1
RLCC [19]	CVPR'21	77.7	90.8	69.2	83.2	27.9	56.5
CACL [36]	TIP'22	80.9	92.7	69.6	82.6	23.0	48.9
PPLR [20]	CVPR'22	81.5	92.8	-	-	31.4	61.1
ISE [37]	CVPR'22	84.7	94.0	-	-	35.0	64.7
Ours		83.4	92.9	72.7	83.9	42.6	68.2

We also address the Text Authorship Attribution (TAA) task by adapting BERT [21], BERTweet [22], and T5 [23], using a dataset with two test sets [34]. Our fully unsupervised method is compared to a prior **supervised** model and outperforms it by 7.0 and 24.5 p.p. in mAP and R1 on the first subset, and by 2.6 and 11.7 on the second. These results demonstrate our method's ability to reduce reliance on labeled data.

C. Results of the third method

Our method is compared to prior works in Table III. We highlight methods that tune clustering parameters per dataset, as they are not applicable in real-world fully unsupervised settings. AdaMG [38] performs well on Market and MSMT17, but its performacne drops when using a fixed 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 in the thesis. Additionally, our method outperforms prior works in Vehicle Re-Identification. On the Veri-Wild dataset, in the most challenging VW-Large setup, we achieve gains of 4.5, 1.7, and 0.6 p.p. in mAP, R1, and R5, respectively, using only 75% of the data. Qualitative results are shown in Figures 5 and 6. Chapter 4 provides detailed analysis and comprehensive comparisons with previous methods.

TABLE III

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	mAP	R1
CCons [39]	ACCV'22	$\mathcal{O}(N^2)$	No	83.0	92.9	33.0	62.0
ISE [37]	CVPR'22	-	Yes	84.7	94.0	35.0	64.7
HHCL [40]	NIDC'21	$\mathcal{O}(N^2)$	No	84.2	93.4	-	-
GRACL [41]	TCSVT'22	$\mathcal{O}(N^2)$	No	83.7	93.2	34.6	64.0
AdaMG [38]	TCSVT'23	$\mathcal{O}(N^2)$	Yes	84.6	93.9	38.0	66.3
Ours (50%)		$\mathcal{O}(kN)$	No	-	-	24.3	50.4
Ours (75%)		$\mathcal{O}(kN)$	No	82.9	92.6	39.3	67.3
Ours (100%)		$\mathcal{O}(kN)$	No	85.8	94.0	43.2	70.9

D. 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. To demostrate generalization, we trained DaliID on DeepChange, where subjects wear different clothes across views, and images exhibit greater distortion and lower quality. In this setting, our method outperforms recent work by 2.9 and 6.8 p.p. in mAP and R1. DaliID proves especially effective under low-quality conditions. Further details and a comprehensive comparison are provided in Chapter 5 of the thesis.

VIII. CONCLUSION AND FUTURE WORK

Our methods address challenges in fully unsupervised re-identification, offering adaptable solutions for forensic and biometric applications. A promising extension is the integration into broader investigative pipelines to uncover relationships between individuals, vehicles, and locations. Recent advances in Large Vision-Language Models (LVLMs) offer new opportunities for enhancing feature extraction in re-identification. Additionally, self-supervised clustering methods used in DeepFake detection suggest potential applications in analyzing 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 LVLMs, developing XAI techniques, and addressing real-world deployment constraints. Overall, this Ph.D. advances the fields of biometrics and forensic science, and pave the way for broader impact in AI and Computer Science research.

IX. PUBLICATIONS, AWARDS, SOFTWARES AND DISTINCTIONS

All Gabriel's publications during his Ph.D. are listed below. His work was recognized by the Brazilian Computing Society

(SBC) as the best Ph.D. thesis defended in Brazil in 2024. The codes for all solutions are publicly available in Github.

A. First Solution (Section III)

1) **G. C. Bertocco**, F. Andaló and A. Rocha. Unsupervised and Self-Adaptative Techniques for Cross-Domain Person Re-Identification. *IEEE Trans. on Inf. Foren. and Secur.*, vol. 16, pp. 4419-4434, 2021. (**IF: 8.0**, **H5-Index: 100**, **Qualis A1** and Citations: 22) [DOI].
2) Presented in the Journal Section of the A3 13^{th} IEEE Inter. Workshop on Inf. Foren. and Secur. (WIFS) [Link]
3) Presented at the IEEE/CVF Workshop of Long-Range Recognition during the Qualis A1 2023 Winter Conf. on Appl. of Comp. Vis. (WACV) [Link]

B. Second Solution (Section IV)

1) **Gabriel C. Bertocco**, Antonio Theophilo, Fernanda Andaló, and Anderson De Rezende Rocha. Leveraging ensembles and self-supervised learning for fully-unsuper-vised person re-identification and text authorship attribution. *IEEE Trans. on Inf. Foren. and Secur.*, 18:3876–3890, 2023. (**IF: 8.0**, **H5-Index: 100**, **Qualis A1** and Citations: 16)[DOI]
2) Presented in the Journal Track Session of the Qualis A2 2023 IEEE Inter. Joint Conf. on Biom. (IJCB) [Link]
3) Presented at InterForensics 2023, the largest forensics conference in LATAM [Link]
4) Integrated in a solution that secured **third place** in the AGReID competition promoted during IJCB 2023 (**Citations: 8**)[DOI],[Leaderboard]

C. Third Solution (Section V)

1) **G. Bertocco**, F. Andaló, T. E. Boult and A. Rocha, "Large-Scale Fully-Unsupervised Re-Identification," in *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 7, no. 2, pp. 156-169, April 2025. (**IF: 5.0**, **H5-Index: 37**, **Qualis A1** and **Citations: 1**) [DOI]
2) Featured in the "Noted in the Literature" section of the IEEE Biom. Council Letter, Vol. 51, Oct. 2024 [Link]
3) Presented at InterForensics 2023, the largest Forensics conference in LATAM [Link]
4) Designed a user-friendly software interface (Figure 4) to facilitate real-case use and expert application [Ex01] [Ex02] [Downloads].
5) Presented in a consultancy to the Forensics Dpt. of the Police of Dubai [Link]
6) Employed in a consultancy and shared the software interface with the Public Prosecutor's Office of the State of São Paulo (MPSP) and of the State of Bahia (MPBA).
7) Employed in a consultancy to the House of Representatives (Brazilian Federal Chamber and Senate) of Brazil.

D. Fourth solution (Section VI)

1) W. Robbins, **G. Bertocco** and T. E. Boult. DaliID: Distortion-Adaptive Learned Invariance for Identification – a Robust Technique for Face Recognition and Person Re-Identification.". *IEEE Access*, 2024. (**IF: 3.6**, **H5-Index: 266[Link]**, **Qualis A3**, and **Citations: 12**)[DOI]
2) Du, Dawei; Hill, Cole; **Bertocco, Gabriel Capiteli**; Pamplona Segundo, Mauricio ; Robbins, Wes J; Richard-Webster, Brandon; Collins, Roderic; Sarkar, Sudeep; Boult, Terrance E; McCloskey, Scott. DOERS: Distant Observation Enhancement and Recognition System. *IEEE Inter. Joint Conf. on Biom.*, 2023. (**H5-index: 32**, **Qualis A2**, and **Citations: 3**)[DOI]
3) **G. Bertocco**, F. Andaló, T. Boult and A. Rocha, "Vision through distortions: Atmospheric Turbulence-and Clothing-invariant long-range recognition," 2024 IEEE Inter. Workshop on Inf. Foren. and Secur. (WIFS), Rome, Italy, 2024.(**H5-index: 21** and **Qualis A3**)[DOI]
4) Delivered it to IARPA as part of the final solution.

E. Other publications

Gabriel also collaborated during his Ph.D. in the following publications:

1) Rafael Padilha, Caroline Mazini Rodrigues, Fernanda Alcantara Andaló, **Gabriel Bertocco**, Zanoni Dias, and Anderson Rocha. Forensic event analysis: From seemingly unrelated data to understanding. *IEEE Security and Privacy*, 18(6):23–32, 2020. (**IF: 3.0**, **H5-index: 37**, **Qualis A1**, and Citations: 3) [DOI]
2) Yang, Jing, José Nascimento, **Gabriel Bertocco**, Antonio Theophilo, Rafael Padilha, Aurea Soriano-Vargas, Fernanda A. Andaló, and Anderson Rocha. "AI Knows What You Did Last Summer: Applications in Digital Forensics." In Computer Vision, pp. 82-108. Chapman and Hall/CRC, 2024. (Book chapter [Link])
3) Rafael Padilha, Antônio Theóphilo, Fernanda A. Andaló, Didier A. Vega-Oliveros, Joao P. Cardenuto, **Gabriel Bertocco**, Jose Nascimento, Jing Yang, and Anderson Rocha. The Artificial Intelligence and the challenges on the Digital Forensics Science in the XXI century (in PT-BR). *USP Advanced Studies*, 2021. [DOI]

ACKNOWLEDGMENT

We thank the Recod.ai and UCCS VAST Labs for hardware support. This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via [2022-21102100003]. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. We thank the financial support of the São Paulo Research Foundation (FAPESP), grants Horus #2023/12865-8, Déjàvu #2019/15825-1, and #2022/02299-2.

References

[1] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Conf. Comput. Vis. Pattern Recog.*, 2020, pp. 9729–9738.

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

[3] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Conf. Comput. Vis. Pattern Recog.*, 2009, pp. 248–255.

[4] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Int. Conf. Comput. Vis.*, 2015, pp. 1116–1124.

[5] G. C. Bertocco, A. Theophilo, F. Andaló, and A. De Rezende Rocha, "Leveraging ensembles and self-supervised learning for fully-unsupervised person re-identification and text authorship attribution," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 3876–3890, 2023.

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

[7] Y. Lin, Y. Wu, C. Yan, M. Xu, and Y. Yang, "Unsupervised person re-identification via cross-camera similarity exploration," *IEEE Trans. Image Process.*, vol. 29, pp. 5481–5490, 2020.

[8] L. Qi, L. Wang, J. Huo, L. Zhou, Y. Shi, and Y. Gao, "A novel unsupervised camera-aware domain adaptation framework for person re-identification," in *Int. Conf. Comput. Vis.*, 2019, pp. 8080–8089.

[9] J. Wu, Y. Yang, H. Liu, S. Liao, Z. Lei, and S. Z. Li, "Unsupervised graph association for person re-identification," in *Int. Conf. Comput. Vis.*, 2019, pp. 8321–8330.

[10] H. Tang, Y. Zhao, and H. Lu, "Unsupervised person re-identification with iterative self-supervised domain adaptation," in *Conf. Comput. Vis. Pattern Recog. Workshops*, 2019, pp. 1536–1543.

[11] Y. Fu, Y. Wei, G. Wang, Y. Zhou, H. Shi, and T. S. Huang, "Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification," in *Int. Conf. Comput. Vis.*, 2019, pp. 6112–6121.

[12] D. Wang and S. Zhang, "Unsupervised person re-identification via multi-label classification," in *Conf. Comput. Vis. Pattern Recog.*, 2020, pp. 10 981–10 990.

[13] Y. Ge, D. Chen, and H. Li, "Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification," *arXiv preprint*, vol. arXiv:2001.01526, 2020.

[14] S. Xuan and S. Zhang, "Intra-inter camera similarity for unsupervised person re-identification," in *Conf. Comput. Vis. Pattern Recog.*, 2021, pp. 11 926–11 935.

[15] H. Chen, B. Lagadec, and F. Bremond, "ICE: Inter-instance contrastive encoding for unsupervised person re-identification," in *Int. Conf. Comput. Vis.*, 2021, pp. 14 960–14 969.

[16] Z. Wang, J. Zhang, L. Zheng, Y. Liu, Y. Sun, Y. Li, and S. Wang, "CycAs: Self-supervised cycle association for learning re-identifiable descriptions," in *Eur. Conf. Comput. Vis.*, 2020, pp. 72–88.

[17] H. Chen, B. Lagadec, and F. Bremond, "Enhancing diversity in teacher-student networks via asymmetric branches for unsupervised person re-identification," in *Winter Conf. Appl. Comput. Vis.*, 2020, pp. 1–10.

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

[19] X. Zhang, Y. Ge, Y. Qiao, and H. Li, "Refining pseudo labels with clustering consensus over generations for unsupervised object re-identification," in *Conf. Comput. Vis. Pattern Recog.*, 2021, pp. 3436–3445.

[20] Y. Cho, W. J. Kim, S. Hong, and S.-E. Yoon, "Part-based pseudo label refinement for unsupervised person re-identification," in *Conf. Comput. Vis. Pattern Recog.*, 2022, pp. 7308–7318.

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

[22] D. Q. Nguyen, T. Vu, and A. T. Nguyen, "BERTweet: A pre-trained language model for english tweets," *arXiv preprint*, vol. arXiv:2005.10200, 2020.

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

[24] B. Boenninghoff, S. Hessler, D. Kolossa, and R. M. Nickel, "Explainable authorship verification in social media via attention-based similarity learning," in *Int. Conf. Big Data*, 2019, pp. 36–45.

[25] N. Potha and E. Stamatatos, "A profile-based method for authorship verification," in *Hellenic Conf. AI*, 2014, pp. 313–326.

[26] B. Boenninghoff, R. M. Nickel, S. Zeiler, and D. Kolossa, "Similarity learning for authorship verification in social media," in *IEEE Int. Conf. on Acoust., Speech Signal Process.*, 2019, pp. 2457–2461.

[27] J. Revaud, M. Douze, C. Schmid, and H. Jégou, "Event retrieval in large video collections with circulant temporal encoding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 2459–2466.

[28] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Eur. Conf. Comput. Vis.*, 2016, pp. 17–35.

[29] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person transfer GAN to bridge domain gap for person re-identification," in *Conf. Comput. Vis. Pattern Recog.*, 2018, pp. 79–88.

[30] P. Xu and X. Zhu, "Deepchange: A long-term person re-identification benchmark with clothes change," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 11 196–11 205.

[31] X. Liu, W. Liu, H. Ma, and H. Fu, "Large-scale vehicle re-identification in urban surveillance videos," in *IEEE Int. Conf. Multimedia Expo*, 2016, pp. 1–6.

[32] H. Liu, Y. Tian, Y. Yang, L. Pang, and T. Huang, "Deep relative distance learning: Tell the difference between similar vehicles," in *Conf. Comput. Vis. Pattern Recog.*, 2016, pp. 2167–2175.

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

[34] A. Theophilo, R. Giot, and A. Rocha, "Authorship attribution of social media messages," *IEEE Trans. Comput. Social Syst.*, 2021.

[35] W. Liu, S. Nie, J. Yin, R. Wang, D. Gao, and L. Jin, "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)*. IEEE, 2021, pp. 81–85.

[36] M. Li, C.-G. Li, and J. Guo, "Cluster-guided asymmetric contrastive learning for unsupervised person re-identification," *IEEE Trans. Image Process.*, vol. 31, pp. 3606–3617, 2022.

[37] X. Zhang, D. Li, Z. Wang, J. Wang, E. Ding, J. Q. Shi, Z. Zhang, and J. Wang, "Implicit sample extension for unsupervised person re-identification," in *Conf. Comput. Vis. Pattern Recog.*, 2022, pp. 7369–7378.

[38] J. Peng, G. Jiang, and H. Wang, "Adaptive memorization with group labels for unsupervised person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 1–1, 2023.

[39] Z. Dai, G. Wang, W. Yuan, S. Zhu, and P. Tan, "Cluster contrast for unsupervised person re-identification," in *Asian Conf. Comput. Vis.*, 2022, pp. 1142–1160.

[40] Z. Hu, C. Zhu, and G. He, "Hard-sample guided hybrid contrast learning for unsupervised person re-identification," in *IEEE Int. Conf. Netw. Intell. Digit. Content*, 2021, pp. 91–95.

[41] H. Zhang, G. Zhang, Y. Chen, and Y. Zheng, "Global relation-aware contrast learning for unsupervised person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 12, pp. 8599–8610, 2022.