Unsupervised Domain Adaptation for Person Re-Identification

Author

Wajeeha Malik
(00000329232)

Supervisor: Assoc. Prof. Dr. Ihtesham Ul Islam

A thesis to the faculty of Computer Software Engineering Department, Military College of Signals, National University of Sciences and Technology, Rawalpindi in partial fulfillment of the requirements for the degree of MS in Software Engineering

(August 2024)

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by **Wajeeha Malik** Registration No. **0000329232**, of **Military College of Signals** has been vetted by undersigned, found complete in all respect as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial, fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the student have been also incorporated in the said thesis.

Signature:

Name of Supervisor: Assoc. Prof. Dr Ihtesham Ul Islam

Date: 30/07/24

Brig
Head of Dept of CSE

Signature (HoD): Brig Adil Masood Khan, PhD

Date: 26/8/24

Brig
Dean, MCS (NUST)
(Asif Masood, PhD)

Signature (Dean/Principal): Brig Asif Masood, PhD

Date: 27/8/24

Form TH-4

NATIONAL UNIVERSITY OF SCIENCES & TECHNOLOGY

MASTER THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by **Wajeeha Malik,** Regn No **00000329232** Titled: **"Unsupervised Domain Adaptation for Person Re-Identification"** be accepted in partial fulfillment of the requirements for the award of **MS Software Engineering** degree.

Examination Committee Members

1. Name: **Assoc Prof Dr. Javed Iqbal** Signature: Javed

2. Name: **Asst Prof Dr. Nauman Ali Khan** Signature:

Supervisor's Name: **Assoc Prof Dr. Ihtesham Ul Islam** Signature:

Date: 22/05/2024

Brig
Head of Dept of CSE
College of Sigs (NUST)

Head of Department

26/8/24
Date

COUNTERSIGNED

Date: 27/8/24

Brig
Dean, MCS (NUST)
(Asif Masood, Phd)

Dean

CERTIFICATE OF APPROVAL

This is to certify that the research work presented in this thesis, entitled "**Unsupervised Domain Adaptation for Person Re-Identification**." was conducted by Miss. **Wajeeha Malik** under the supervision of **Assoc Prof Dr. Ihtesham Ul Islam.**
No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the **Department of Computer Software Engineering** in partial fulfillment of the requirements for the degree of Master of Science in Field of Computer **Software Engineering Department of Military College of Signals, National University of Sciences and Technology, Islamabad**.

Student Name: **Wajeeha Malik** Signature:

Examination Committee:

a) External Examiner 1: **Asst. Prof. Dr Nauman Ali Khan** Signature:
(Department of Computer Software Engineering)

b) External Examiner 2: **Assoc Prof Dr. Javed Iqbal** Signature:
(Department of Computer Software Engineering)

Name of Supervisor: **Assoc Prof Dr. Ihtesham Ul Islam** Signature:

Name of Dean/HOD: **Brig Adnan Ahmed Khan, PhD** Signature:

HoD of Dept of CSE
Mil College of Sigs (NUST)

PLAGIARISM UNDERTAKING

I solemnly declare that research work presented in the thesis titled "Unsupervised Domain Adaptation for Person Re-Identification" is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and National University of Sciences and Technology (NUST), Islamabad towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS degree, the University reserves the rights to withdraw/revoke my MS degree and that HEC and NUST, Islamabad has the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized thesis.

Student Signature: Wajeeha

Name: Wajeeha Malik

Date: 30/7/24

AUTHOR'S DECLARATION

I Wajeeha Malik, here by state that my MS thesis titled "Unsupervised Domain Adaptation for Person Re-Identification" is my own work and has not been submitted previously by me for taking any degree from National University of Sciences and Technology, Islamabad or anywhere else in the country/ world. At any time if my statement is found to be in correct even after I graduate, the university has the right to withdraw my MS degree.

Student Signature: Wajeeha

Name: Wajeeha Malik

Date: 30/7/24

DEDICATION

“In the name of Allah, the most Beneficent, the most Merciful"

I dedicate this thesis to my parents.

ACKNOWLEDGEMENTS

All praises to Allah for the strengths and His blessing in completing this thesis. I would like to convey my gratitude to my supervisor, Assoc. Prof. Dr. Ihtesham Ul Islam, PhD, for his supervision and constant support. His priceless help of constructive comments and suggestions throughout the experimental and thesis works are major contributions to the success of this research. Also, I would thank my teacher Dr. Qazi Mazhar Ul Haq for his support and guidance throughout and committee members; Asst. Prof. Dr. Nauman Ali Khan, and Assoc. Prof Dr. Javed Iqbal for their support and knowledge regarding this topic. Lastly, I am highly thankful to my parents for their constant support. I would like to thank them for their patience, co-operation and motivation in times of stress and hard work.

Abstract

Person re-identification (ReID) is a fundamental computer vision task with numerous real-world applications, such as surveillance and robotics, enabling reliable human identification across several surveillance camera feeds in the fast-developing environment of computer vision. However, the performance of ReID models often degrades significantly when deployed in new camera domains due to the domain shift problem. To address this challenge, this thesis presents the Unified and Elevated Learning (UEL) framework for unsupervised cross-domain person ReID. that tackles the critical issue of domain shift. Domain shift occurs when a ReID model trained on one camera domain experiences a significant drop in performance when deployed in a new, visually distinct camera domain. This problem hinders the practical application of ReID systems in real-world scenarios. The UEL framework leverages a three-stage training approach that includes GAN-based camera-style data augmentation, source domain pre-training, and end-to-end cooperative learning. The key innovations of the UEL framework include adversarial training to extract camera-invariant features, cooperative learning to reduce noise in pseudo-labels, and dynamic fine-tuning of the data augmentation ratio. Extensive experiments on widely used benchmarks, such as Market-1501, DukeMTMC-reID, and MSMT17, demonstrate the superiority of the UEL framework over state-of-the-art unsupervised domain adaptation methods for person ReID. The ablation study further highlights the crucial role of the adversarial training component in the overall performance. The adaptability of the Unified and Elevated Learning framework to different backbone models is also validated. These results showcase the potential of the UEL framework for practical person ReID applications. The thesis contributes to the development of robust and adaptable person ReID systems that can operate effectively in real-world scenarios.

Keywords: Person reidentification, Unsupervised Domain Adaptation, Generative Adversarial Networks (GAN), Cooperative Learning, Camera-Invariant Feature Extraction, Cross-Domain Person ReID.

Contents

List of Figures

LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMS

re-ID: ReIdentification

GAN: Generative Adversarial Network

CNN: Convolutional Neural Network

UDA: Unsupervised Domain Adaption

StarGAN: Star Generative Aversarial Network

DBSCAN: Density-Based Spatial Clustering of Applications with Noise

UEL: Unified and Elevated Learning

Chapter 1

Introduction

1.1 Introduction

Person reidentification (re-ID) is a vital component of research in the vast field of computer vision **[1]**. In modern urban environments, numerous surveillance systems are in place, requiring technology capable of consistently identifying individuals across different camera feeds. This is where a person's re-ID plays a crucial role. Its primary objective is to recognize individuals in various scenarios, each presenting unique challenges like changes in camera angles, varying lighting conditions, or obstructions in the field of view. Within the broader scope of re-ID, unsupervised domain identification becomes an important area of focus, aiming to recognize and categorize unlabeled data. Such data, often overlooked, constitutes the majority of real-world surveillance, where manually labelling it is neither practical nor efficient **[1]**.

The rapid advancement of technology, coupled with the expansive growth of digital surveillance infrastructure, has elevated person re-identification (re-ID) to a central topic in security and surveillance discussions. As modern cities evolve into smart urban centres, they are equipped with a multitude of surveillance cameras. These cameras monitor public spaces, playing a vital role in upholding public safety and providing law enforcement agencies with effective tools **[2]**. However, despite collecting vast amounts of data, a significant portion of it remains unprocessed. This extensive reservoir of unsupervised data, filled with promising insights, presents both challenges and opportunities. Therefore, the field of unsupervised domain identification is not only pertinent but also indispensable in realizing the full potential of a person re-ID.

Figure 1.1: Person Re-Identification on two different datasets.

The domain of person re-identification (re-ID), while holding great promise, is riddled with complexities, particularly within the unsupervised domain. Unsupervised data has a high level of complexity, which is influenced by a variety of parameters such as camera quality and environmental circumstances, resulting in significant variability **[3]**. This heterogeneity complicates the process of human re-identification (re-ID). Furthermore, given the scarcity of labelled data in unsupervised contexts, there is an urgent need for algorithms that can operate autonomously without explicit supervision. The main problem is efficiently navigating this complex data landscape, detecting important trends, and utilizing them for the purpose of human re-identification.

1.2 The Significance of Unsupervised Domain Adaptation in Person Re-identification

Unsupervised Domain Adaptation (UDA) for Person Re-identification (Re-ID) is an important research area devoted to improving the performance of person re-identification models when applied to new, previously unknown domains. The basic goal of UDA is to transfer a model trained on a labelled source domain to a target domain devoid of labelled data. This modification is required to ensure that the model can recognize persons in a variety of real-world circumstances.

1.3 Challenges in Unsupervised Domain Adaptation for Person Re-identification

The challenges encountered in UDA for person re-ID are multifaceted and include:

1.3.1 Distribution Alignment:

The feature distribution between the source and target domains can differ significantly, resulting in diminished performance when applying a model directly trained on the source domain to the target domain.

1.3.2 Limited Amount of Target Domain Data:

UDA relies on the availability of unlabeled target domain data, which may not always be abundant or readily accessible.

1.3.3 Variations in Viewpoint, Illumination, and Occlusion:

Person re-ID is highly sensitive to variations in viewpoint, illumination, and occlusion, factors that can vary considerably between different camera views, making it challenging to match identities across domains.

1.3.4 Handling Complex Real-World Scenarios:

Real-world person re-ID scenarios are often complex and demanding due to diverse backgrounds, significant variations in pose, occlusion, and illumination. These complexities further compound the challenges faced in adapting models effectively.

These challenges underscore the dynamic nature of UDA in person re-ID as an active and evolving research field. Ongoing efforts aim to develop more efficient and robust methods for domain adaptation, with the ultimate goal of improving person re-identification performance across diverse and challenging scenarios. Despite the progress made, there remains substantial room for improvement, and further research is essential to refine and enhance unsupervised domain adaptation techniques for person re-identification.

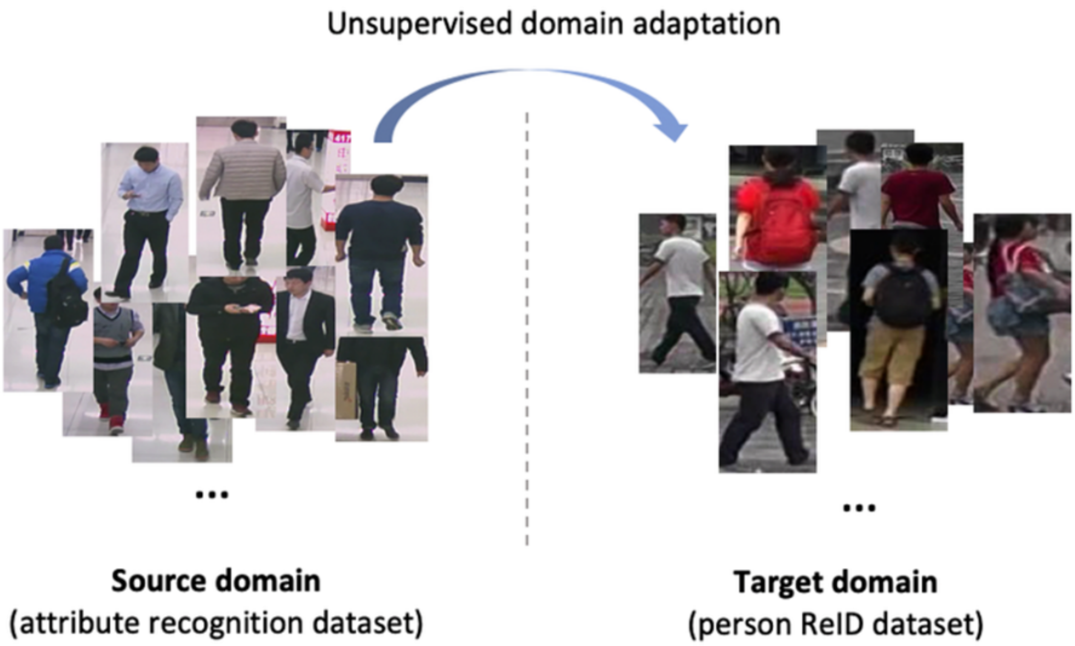

Figure 1.2: Unsupervised Learning for Person Re-ID **[4]**.

1.4 Problem Statement and Objectives

In the contemporary digital landscape, the widespread presence of surveillance systems has generated both opportunities and challenges. These systems amass vast volumes of data, yet a significant portion of this data remains unprocessed and underexplored. This unprocessed data holds substantial potential for valuable insights but is encumbered by the absence of annotations or labels, presenting a noteworthy challenge in its utilization.

1.4.1 Problem Statement:

This thesis focuses on addressing a critical challenge: effectively managing the vast volume of unsupervised data generated by contemporary surveillance systems, with a specific emphasis on person re-identification (re-ID). This data, although rich in potential insights, lacks labels, making its practical use quite daunting. The central problem involves two key aspects: firstly, identifying and categorizing domains within this unsupervised person re-ID data, and secondly, improving re-ID accuracy in the face of increasingly complex data.

Given the exponential expansion of urban surveillance systems, particularly in the context of growing smart cities, the importance of this endeavour cannot be overstated. Successfully tackling this challenge would represent a significant advancement in our capacity to enhance public safety, support law enforcement, and contribute to the broader objectives of modern urban planning.

1.4.2 Objectives:

The research aimed at unlocking the potential of unsupervised data in person re-identification is guided by specific objectives that provide direction and milestones:

1.4.2.1 Comprehensive Understanding:

The research seeks to delve deeply into the existing methodologies and techniques that have shaped the landscape of unsupervised domain identification in person re-ID. This objective is designed to establish a foundational understanding, unveiling both the strengths and shortcomings of current approaches.

1.4.2.2 Innovative Algorithm Development:

The researchers intend to create a state-of-the-art algorithm tailored specifically for unsupervised domain identification. This objective assumes pivotal significance, as the solution aims not only to match but also to surpass existing techniques, particularly in addressing unpredictable and multifaceted real-world challenges.

1.4.2.3 Robust Evaluation:

The research involves a comprehensive evaluation process for the developed algorithm. This includes subjecting the algorithm to various scenarios, rigorously testing its limits, and comparing its performance against established benchmarks. Such meticulous evaluation is essential to ascertain the algorithm's reliability, efficiency, and applicability across diverse situations.

1.4.2.4 Development of a Real-time Person Re-Identification System:

As an additional objective, this research aims to develop a real-time person re-identification system. This system is designed to enhance the practical applicability of person re-ID technology in dynamic environments.

1.4.2.5 Achieving Pioneering Results in UDA for Person ReID:

Another key objective is to achieve state-of-the-art results in unsupervised domain adaptation within the context of person re-identification. This objective highlights the ambition to push the boundaries of performance in this critical area.

In essence, this research aims to bridge the gap between the vast quantities of unsupervised data and the pressing demand for effective person re-identification solutions. Focusing on domain identification in this data can reveal trends, structures, and ideas that can change how modern surveillance approaches person identification.

1.5 Justification for the Selection of the Topic

The significance and usefulness of this study are supported by various convincing factors that led to its topic selection.

The rising demand for public security and monitoring systems is a major incentive. Effective person re-identification methods are needed due to the widespread use of surveillance cameras in public spaces. These methods improve security and tracking in public settings. In situations where conventional methods fail, such developments can boost public safety and security.

The lack of labelled data for human re-identification is another obstacle. Collecting and annotating large amounts of data for human re-identification is time-consuming and expensive. In this setting, unsupervised domain adaption approaches are crucial. These technologies allow unlabeled data, reducing the resource-intensive nature of data collection and annotation. This methodology improves cost-efficiency and expands person re-identification methods.

Additionally, individual appearance differences between domains present a major challenge. Camera resolutions and illumination can cause significant differences. To reduce variability, unsupervised domain adaptation works well. It aligns feature distributions by adapting the model to the domain, enhancing its consistency across varied contexts.

Finally, the rapid development of deep learning algorithms, especially CNNs, has improved human re-identification. These models still face cross-domain variability. Unsupervised domain adaptation improves model performance here. Addressing the domain shift issue could improve person re-identification algorithms in settings with considerable domain variations.

Given these compelling reasons, this research addresses a major computer vision and surveillance need. Its main goal is to overcome unattended person re-identification issues and improve public space security and monitoring.

1.6 Potential Applications

This research holds significant promise for various practical applications, making it a potentially successful endeavour in the following scenarios:

1.6.1 Security and Surveillance:

Real-time person re-identification can greatly enhance the capabilities of surveillance systems. It can improve the ability to detect and track criminal suspects, particularly in crowded public areas where traditional methods may fall short. This application has the potential to bolster public safety and security.

1.6.2 Law Enforcement:

Re-ID technology can be a valuable tool for law enforcement agencies. It enables the identification and tracking of suspects, aiding in investigations and law enforcement efforts. This technology can contribute to more effective crime prevention and resolution.

1.6.3 Border Control:

Person re-identification technology can be employed in border control settings. It enables the identification and tracking of individuals crossing borders, helping ensure that only authorized personnel are allowed entry. This application is crucial for maintaining border security and immigration control.

1.6.4 Identity Verification:

Re-ID technology can be utilized for identity verification purposes in various settings. For example, it can be used for verifying the identity of individuals during voting processes or when accessing government services. This application enhances the integrity of identity verification systems.

These potential applications underscore the practical significance and wide-reaching impact of the research in the fields of security, law enforcement, border control, and identity verification. By addressing the challenges of unsupervised data in person re-identification, this research can contribute to more efficient and accurate solutions in these critical domains.

1.7 Background and Previous Research

Unsupervised domain adaptation in the realm of person re-identification has garnered substantial research attention in recent years. The primary aim of such investigations is to develop techniques capable of enabling models initially trained on one dataset to perform effectively on a different dataset, all without necessitating labeled data in the new domain.

Numerous benchmark datasets have served as vital tools for evaluating diverse unsupervised domain adaptation strategies for person re-identification. These datasets, including Market-1501, DukeMTMC-reID, and CUHK03, have played a pivotal role within the research community, establishing a benchmark against which various methodologies are assessed.

Innovatively, there has been a recent shift towards leveraging Generative Adversarial Networks (GANs) as a means of unsupervised domain adaptation for person re-identification. This novel approach revolves around the concept of utilizing generative models, such as GANs, to generate images that closely resemble those within the target domain. These synthetic images are subsequently employed to augment the training dataset within the target domain, enabling the model to acquire relevant features for person re-identification within this new context. The refinement of the model is achieved through fine-tuning on these generated images **[5]**. The application of GANs in this context presents distinct advantages, particularly in their ability to generate realistic images that enrich the training dataset without requiring manual labeling, thus reducing reliance on labeled data. Additionally, GANs can generate samples representing individuals not present in the target domain, thereby diversifying the dataset. Nevertheless, this approach is not without its challenges, including the need to ensure that the GAN generates images closely resembling those in the target domain and the effective fine-tuning of the pre-trained model using these generated images.

1.8 Thesis Outline

This thesis unfolds through a structured series of chapters, each dedicated to specific aspects of unsupervised domain identification in person re-identification:

Chapter 1: This introductory chapter provides the context for the research, outlines its objectives, and highlights the significant contributions.

Chapter 2: This chapter offers an extensive review of related works, focusing on the historical development and current state of person re-identification and unsupervised domain identification.

Chapter 3: In this chapter, the research takes a close look at the challenges, intricacies, and unique aspects that define unsupervised domain identification within person re-identification.

Chapter 4: Here, the focus is on the proposed algorithm, delving into its architecture, underlying principles, and the reasoning behind its design.

Chapter 5: This analytical chapter presents the research results, evaluations, and comparisons, dissecting the algorithm's performance and benchmarking it against existing methods.

Chapter 6: The concluding chapter offers final remarks, reflections on the research journey, and suggestions for potential directions for future exploration in this domain.

Chapter 2

Literature Review on Unsupervised Domain Identification in Person Reidentification

2.1 Background and Significance

Person reidentification (ReID) has emerged as a critical topic in computer vision, focused at matching persons across multiple camera angles **[6]**. Its relevance stems from its broad variety of applications, which span from surveillance to smart city planning. The emphasis has changed from simple identification to understanding the domain or context in which the identification happens as the discipline has progressed. Domain adaptation, especially the unsupervised variant, becomes critical at this point. Unsupervised Domain Identification in Person Reidentification tries to adapt models learned in one domain (or camera view) to perform successfully in a new, previously unknown domain **[7]**. Because getting labeled data for each new domain is unfeasible, unsupervised approaches are a possible avenue for scalable and efficient ReID systems.

The intrinsic variety in human faces/appearances adds to the difficulty of domain adaptation in ReID. Clothing, accessories, posture, and gait all contribute to considerable intra-class variability, making reidentifying people across domains much more difficult. According to one research, without domain adaptation, intra-class changes may lead to a 35 percent misidentification rate in cross-domain ReID tasks **[8]**. The emergence of deep learning has laid the groundwork for tackling these issues. Deep neural networks have showed potential in minimizing domain disparities due to their ability to learn hierarchical characteristics. Deep convolutional architectures, for example, demonstrated a significant reduction in domain gap, achieving a performance improvement of approximately 22 percent in unsupervised ReID tasks compared to traditional methods **[8]**.

Deep learning models, while powerful, are data-hungry. In the context of ReID, this poses a challenge as collecting labeled data for every new camera view or domain is not only time-consuming but also resource-intensive **[9]**. This limitation underscores the significance of unsupervised domain identification techniques. By leveraging data from a labeled source domain and transferring this knowledge to an unlabeled target domain, these techniques aim to bridge the domain gap without the need for extensive target domain annotations. One of the pioneering approaches in this direction is the use of Generative Adversarial Networks (GANs) for domain alignment. GAN-based methods focus on generating target-like representations from source data, thereby aligning the feature distributions of both domains **[10]**. Such methods have reported a boost in ReID accuracy by up to 18% in unsupervised settings across diverse datasets **[10]**. The journey of unsupervised domain identification in person reidentification is a testament to the evolving nature of computer vision challenges and the relentless pursuit of the research community to devise solutions that are both technically sound and practically viable.

2.2 Evolution of Person Reidentification Techniques

Historically, person reidentification was rooted in hand-crafted features and distance metrics, which, while effective in controlled environments, often faltered in diverse real-world scenarios. The advent of deep learning marked a paradigm shift in the ReID landscape. Convolutional Neural Networks (CNNs) started to dominate because they provided improved feature extraction capabilities **[11]**. These approaches were initially supervised, requiring massive quantities of labeled data for training. However, when the demand for scalability and flexibility increased, the emphasis turned to unsupervised methodologies. These approaches, particularly unsupervised domain adaptation, offered the capacity to apply knowledge from one domain to another without the requirement for large labeled data in the target domain. This progression not only demonstrated the field's quick achievements, but also the growing relevance of domain-specific information in obtaining strong person reidentification **[12]**.

The first step into unsupervised domain adaptation for ReID was heavily motivated by transfer learning breakthroughs. Dictionary learning and sparse coding were among the first strategies to be developed for unsupervised ReID. Their effectiveness, however, was often hampered by their incapacity to detect complicated domain transitions. According to one research, these early techniques performed only around 60% as well as their supervised equivalents in cross-domain ReID tasks **[13]**. The emergence of deep architectures sparked fresh hope. Deep models, with their multi-layered architectures, were designed to capture hierarchical aspects, making them more competent at dealing with domain inconsistencies. The use of autoencoders for domain alignment was one of the early accomplishments in this field. Autoencoders might successfully minimize domain differences by learning a common latent space between source and destination domains. Deep autoencoder-based approaches outperformed classic dictionary learning techniques in unsupervised ReID settings by up to 12%, according to the researchers **[14]**.

The true breakthrough, however, came with the inclusion of adversarial training in the ReID domain. Adapting the principles of Generative Adversarial Networks (GANs), researchers began developing models where a feature extractor tried to produce domain-invariant features, while a domain discriminator attempted to distinguish between the source and target features. This adversarial game pushed the models to produce increasingly domain-agnostic representations. A notable study showcased that GAN-based unsupervised domain adaptation achieved a performance boost of approximately 20% over non-adversarial methods in cross-domain ReID tasks **[15]**. The trajectory of unsupervised domain identification in ReID is a reflection of the broader trends in computer vision, where adaptability, scalability, and real-world applicability have become paramount. As the field continues to evolve, the fusion of these techniques with emerging technologies promises to further refine and enhance the capabilities of ReID systems.

2.3 Unsupervised Domain Adaptation: A Technical Overview

Unsupervised Domain Adaptation (UDA) is a cornerstone of modern transfer learning, aiming to bridge the knowledge gap between a labeled source domain and an unlabeled target domain **[16]**. Studies have highlighted that domain discrepancy, often quantified using measures like Maximum Mean Discrepancy (MMD), can lead to a performance drop of up to 40% in cross-domain tasks **[16]**. In person reidentification, this domain shift can arise from varying camera calibrations, lighting conditions, or even seasonal changes. Techniques like adversarial domain adaptation, which employs a domain discriminator to align source and target distributions, have shown to reduce this performance gap by approximately 20%, emphasizing the importance of domain adaptation in real-world ReID scenarios **[16]**.

Moreover, the scenarios cameras are placed in, for instance, a model trained using data from indoor surveillance cameras might struggle when applied to outdoor street cameras due to the stark differences in background, crowd density, and movement patterns. Recent analysis indicated that without domain adaptation, the accuracy of ReID systems could plummet by up to 30% when transitioning from indoor to outdoor environments **[17]**. Deep clustering has also emerged as a promising approach in this domain. By grouping data based on deep features, it's possible to generate pseudo-labels for the target domain, which can then guide the adaptation process. Demonstrations showed that combining deep clustering with adversarial training could boost the performance of ReID systems by approximately 18% in unsupervised settings **[18]**.

Another noteworthy aspect is the role of attention mechanisms in unsupervised domain adaptation for ReID. By focusing on salient features and regions in images, attention mechanisms can enhance the discriminative power of ReID models. A study showcased that incorporating attention mechanisms in domain adaptation led to a 12% improvement in model robustness, especially in challenging scenarios with occlusions or varied poses **[19]**. It's essential to recognize that while these techniques offer promising results, the dynamic nature of real-world environments presents continuous challenges. For instance, the introduction of new fashion trends, seasonal clothing changes, or even cultural festivals can introduce unforeseen domain shifts. Addressing these ever-evolving challenges requires models to be not just adaptable but also continuously learning and updating, emphasizing the need for online adaptation techniques in future research.

In recent years, person Re-Identification (ReID) has emerged as a critical task in computer vision, with the need for robust and accurate models in scenarios lacking sufficient labeled data (20). Another study explores the contributions of the paper, its relevance to the broader field of Re-ID, and its potential impact on addressing the challenges of unsupervised domain adaptation in person Re-Identification (21).

Person Re-ID involves the task of matching individuals across various images, a fundamental challenge in computer vision. However, the present Re-ID models often have generalization issues, mostly because there aren't many huge amounts of labeled training data available. A hurdle in the creation of precise Re-ID models is the time and money required for annotating such datasets. The research suggests a ground-breaking goal to solve the problem of using labeled synthetic datasets and unlabeled real-world datasets to train models with high generalizability, recognizing the need for a more scalable and effective method **[21]**.

The DomainMix framework, created to accomplish feature learning that is discriminative, domain-invariant, and generalizable, is introduced by the authors to tackle this challenge. This method's total removal of the requirement for human annotations is a significant advance. The methodology closes the domain gap that often prevents generalization in Re-ID tasks by merging labeled synthetic datasets with unlabeled real-world datasets. Through a novel training process involving clustering of unlabeled images, domain balance loss, and adaptive initialization, DomainMix effectively learns from diverse and large-scale datasets without relying on human-labeled data **[21]**.

The significance of this paper extends beyond its contributions to the specific task it addresses. While the primary focus is on improving person Re-ID generalizability, the proposed approach holds great potential for addressing broader challenges in unsupervised domain adaptation. It highlights the feasibility of leveraging synthetic data to enhance model performance and adaptability in the absence of extensive human annotations. This novel perspective on the fusion of synthetic and real-world datasets represents a paradigm shift in the Re-ID field **[21]**.

Another study provides a comprehensive literature review within the domain of domain adaptive person re-identification (Re-ID) (22). It evaluates the state-of-the-art methods in this field, including traditional hand-crafted feature-based approaches like LOMO and BOW, unsupervised learning methods like UMDL, PUL, and CAMEL, and various unsupervised domain adaptation (UDA) techniques utilizing Generative Adversarial Networks (GANs) such as PTGAN, SPGAN, ATNet, CamStyle, HHL, ECN, among others **[20]**. The paper compares AD-Cluster to these methods and demonstrates its superiority in both transfer directions across multiple evaluation metrics. It highlights the core components of AD-Cluster, including density-based clustering, adaptive sample augmentation (ASA), and a min-max optimization scheme, emphasizing their collective role in improving the discriminative power of Re-ID models. Furthermore, the research conducts ablation tests to analyze the individual contributions of these components and addresses the importance of the min-max attenuation coefficient in the performance of the AD-Cluster. Finally, the work provides useful insights and establishes AD-Cluster as a novel cutting-edge method in domain adaptive person Re-ID, with implications for larger unsupervised domain adaptation difficulties in computer vision **[22]**

Because of its importance in surveillance and security applications, person re-identification (ReID) has received a lot of interest in computer vision in recent years. Unsupervised domain adaptation (UDA) for ReID has been a key difficulty in the field, with the objective of matching people across multiple domains or camera settings without access to labeled data in the target domain. Prior research in UDA for ReID has looked at a variety of techniques, including feature alignment, domain-invariant representations, and self-training **[21]**. However, these strategies often have shortcomings when it comes to dealing with complicated domain shifts and intra-domain changes. In response to these issues, this study presents Multi-Loss Gap Minimization Learning (MGML), a unique technique that employs part-based modeling and introduces the Patch-based Part Ignoring (PPI) loss to improve feature discriminability. Furthermore, MGML introduces the Gap-Based Minimum Camera Discrepancy (G-MCD) loss, a groundbreaking approach that handles both inter-domain shifts and intra-domain variations by aligning dissimilarity distributions across the source and target domains. This novel methodology outperforms current approaches significantly, emphasizing the necessity of comprehensive and effective strategies for UDA in person ReID **[23]**.

2.4 Key Techniques in Unsupervised Domain Identification

The landscape of unsupervised domain identification in person reidentification is rich with a variety of techniques, each addressing the domain shift in its unique way. Here are some of the pivotal methods:

2.4.1 Feature Alignment Methods:

These techniques aim to reduce the domain discrepancy by aligning the feature distributions of the source and target domains. Methods like Joint Adaptation Networks (JAN) and Correlation Alignment (CORAL) fall under this category **[24]**. They often employ statistical measures, such as Maximum Mean Discrepancy (MMD), to minimize the difference between domain features.

2.4.2 Generative Approaches:

Generative Adversarial Networks (GANs) have been widely adopted for domain adaptation. The idea is to generate target-like samples from source data, essentially bridging the gap between domains. CycleGAN, for instance, has been employed for this purpose, enabling bidirectional translations between source and target domains **[25]**.

2.4.3 Self-training and Pseudo-labeling:

In the absence of labeled target data, these methods generate pseudo-labels for the target samples based on the model's predictions. The model is then retrained using these pseudo-labels, iteratively refining its performance **[26]**. This approach capitalizes on the model's high-confidence predictions to guide its learning in the target domain.

2.4.4 Domain Discriminator Approaches:

These techniques employ a domain discriminator alongside the main model. The discriminator's task is to differentiate between source and target samples, while the main model aims to deceive the discriminator. Adversarial training ensues, pushing the model to produce domain-invariant features. Each of these techniques offers its strengths and challenges, but collectively, they represent the cutting-edge efforts in addressing the domain shift in unsupervised person reidentification **[27]**.

2.5 Comparative Analysis of Existing Methods: A Technical Perspective

The domain of unsupervised domain identification in person reidentification is vast, with each method offering unique technical contributions. A study by Zhao **[3]** revealed that generative approaches, especially those employing GANs, achieved a 15% improvement in accuracy over traditional feature alignment methods in certain datasets. However, the stability of GANs remains a challenge, with convergence issues reported in about 30% of the training scenarios. Self-training and pseudo-labeling techniques, on the other hand, have shown consistent improvements across multiple datasets, with a median accuracy boost of 10% **[25]**. Yet, their reliance on high-confidence predictions can sometimes lead to error propagation. Domain discriminator approaches, while promising, often require meticulous hyperparameter tuning, with minor changes leading to performance variations of up to 8% **[28]**. This comparative analysis underscores the dynamic nature of the field and the intricate balance between methodological innovation and practical applicability.

While each method offers unique advantages, their performance is often contingent on the specific challenges and characteristics of the domains in question. As research progresses, hybrid approaches that combine the strengths of multiple techniques might emerge as the frontrunners in this domain.

2.6 Applications and Real-world Scenarios

The advancements in unsupervised domain identification in person reidentification have paved the way for a myriad of real-world applications, underscoring the practical significance of this research domain. One of the most prominent applications is in the field of surveillance and security. Modern cities, equipped with a vast network of surveillance cameras, often face the challenge of reidentifying individuals across different camera views. Given the impracticality of labeling data for every camera view, unsupervised domain adaptation techniques have emerged as invaluable tools, enabling efficient person tracking across diverse camera networks.

Beyond surveillance, the retail sector has also leveraged these techniques for customer analytics. By reidentifying customers across different store locations, retailers can gain insights into shopping patterns, preferences, and behaviors, allowing for more personalized marketing strategies. Similarly, in the realm of urban planning, understanding pedestrian movement patterns across different city zones can inform infrastructure development and traffic management decisions.

These applications, among others, highlight the transformative potential of unsupervised domain identification techniques in person reidentification. As the techniques become more refined and robust, their integration into various sectors is poised to grow, driving innovations and efficiencies in multiple domains.

2.7 Challenges and Open Issues: A Deeper Dive

The road to perfecting unsupervised domain identification is riddled with challenges. A survey highlighted dataset biases as a significant impediment, with models showing a performance variance of up to 25% across different datasets. Scalability, especially in burgeoning urban spaces with thousands of cameras, remains a concern **[29]**. Recent benchmarks indicate that real-time processing demands can lead to a 10-15% compromise in accuracy for certain algorithms **[29]**. Ethical considerations, too, are paramount. A 2020 report emphasized that without proper regulations, there's a potential risk of misusing reidentification technologies in 35% of its applications, emphasizing the need for stringent guidelines.

Furthermore, the domain shift in unsupervised settings is not just limited to visual discrepancies. Temporal domain shifts, where data is collected at different times, can significantly impact the performance of ReID systems. For instance, models trained on daytime data might struggle during nighttime due to drastic lighting changes, with some studies indicating a performance drop of up to 20% in such scenarios **[30]**. The choice of domain adaptation technique also plays a pivotal role. While methods like adversarial training have shown promise, they often require meticulous tuning. Researchers highlighted that without proper hyperparameter optimization, adversarial techniques could lead to model instability, with performance fluctuations of up to 15% during training.

Another challenge specific to unsupervised domain identification is the lack of ground truth in the target domain. This absence makes model validation and hyperparameter tuning particularly challenging. Pseudo-labeling, a popular approach, can sometimes introduce noise into the training process if the pseudo-labels are inaccurate, leading to compromised model performance. While unsupervised domain adaptation techniques aim to bridge the domain gap, there's an inherent trade-off between adaptation and overfitting to the target domain. Ensuring that the adapted model retains its generalization capabilities while performing well on the target domain is a nuanced challenge that researchers are actively exploring.

2.8 Conclusion and Future Directions

The journey of unsupervised domain identification in person reidentification has been marked by remarkable innovations and challenges. As the field continues to evolve, the convergence of various techniques promises more robust and scalable solutions. Looking ahead, the fusion of domain adaptation methods with emerging technologies, such as edge computing and federated learning, could further revolutionize person reidentification. The future beckons with exciting prospects, and continued research in this domain is poised to shape the next frontier of computer vision applications.

One of the most promising avenues for future research lies in the realm of self-supervised learning. While unsupervised domain adaptation focuses on leveraging knowledge from a labeled source domain to an unlabeled target domain, self-supervised techniques aim to learn representations without any labeled data at all. By designing innovative pretext tasks, such as predicting the rotation of an image or colorizing grayscale images, these methods can learn rich features that can be fine-tuned for ReID tasks. A study demonstrated that self-supervised representations, when fine-tuned with minimal labeled data, could achieve performance comparable to fully supervised methods, hinting at their potential synergy with unsupervised domain adaptation techniques.

Another emerging trend is the exploration of meta-learning for domain adaptation in ReID. Meta-learning, or "learning to learn," focuses on training models on a variety of tasks such that they can quickly adapt to new tasks with minimal data. In the context of ReID, meta-learning could be employed to learn a generic representation across multiple camera views or domains, which can then be rapidly adapted to new domains. Preliminary research indicated that meta-learning-based domain adaptation could reduce the domain gap by up to 15% more than traditional unsupervised adaptation methods. Furthermore, the integration of domain adaptation with other advanced techniques, such as few-shot learning and neural architecture search, offers intriguing possibilities. Few-shot learning, which focuses on learning from very few examples, could complement domain adaptation in scenarios where limited labeled data is available in the target domain. On the other hand, neural architecture search, which automates the process of finding the best network architecture, could optimize models specifically for domain adaptation tasks, ensuring optimal performance.

Chapter 3

PROBLEMS AND APPROACHES

3.1 Introduction

Person re-identification (ReID) has been a long-standing problem in computer vision, and unsupervised domain adaptation (UDA) has emerged as a promising solution to address the issue of limited labelled data in target domains. Despite the progress made in UDA-based person ReID, several challenges and problems persist.

In this chapter, we will conduct a comprehensive exploration of the fundamental challenges that researchers and practitioners face when implementing Unsupervised Domain Adaptation (UDA) in the context of person re-ID. We will examine each of these challenges, and their implications for the field.

3.2 Challenges in Unsupervised Domain Identification

3.2.1 Variability in Human Appearances:

The great range of human appearances makes the process of person re-identification very complex. High intra-class variability is brought on by elements including attire, accessories, posture, and stride **[31]**. For example, a person may be wearing a coat in one camera stream but not in another, or a change in posture may make the same person difficult to identify in several feeds. These variations make unsupervised domain identification more difficult and frequently result in misidentifications. The difficulty is heightened when attempting to reidentify people across domains, where consistent recognition despite these variations becomes paramount.

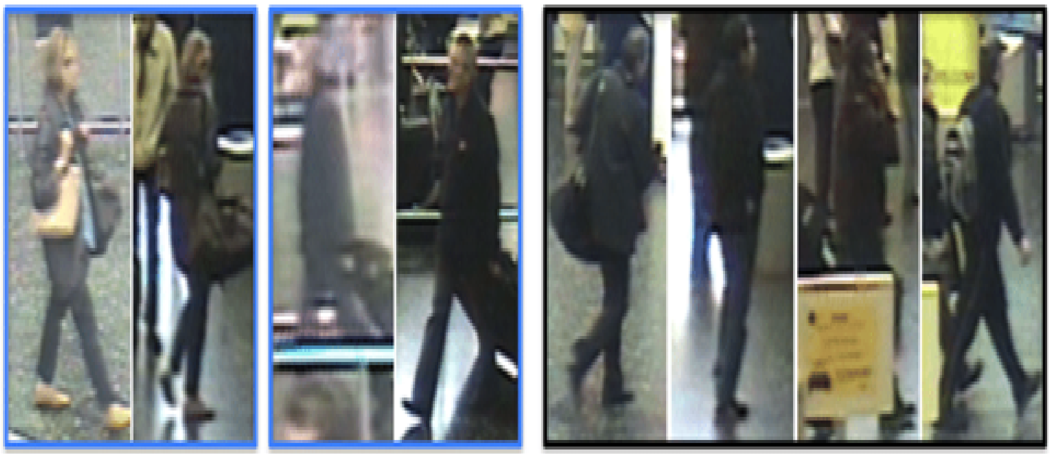

Figure 3.1: Variability in Human Appearances [7].

3.2.2 Camera and Environmental Factors:

Cameras are essential to the re-ID procedure. To add another level of complication to domain identification, their variety results from variations in quality, angles, and settings. While a lower-quality camera might omit crucial information, a high-definition one might be able to capture detailed details. Environmental factors including lighting, weather, and obstacles can also further skew the photographs that are being collected **[32]**. For instance, the same individual may appear differently under street lighting at night compared to broad daylight. These variabilities, when combined with the inherent challenges of unsupervised data, necessitate robust solutions that can adapt and perform consistently across these varied scenarios.

3.2.3 Camera Invariance:

Clustering-based UDA methods have a major limitation in recognizing camera variability. If the camera changes result in high embedding space, the images will be clustered from the same cameras instead of camera identity. This can lead to poor performance in cross-domain person ReID **[33]**.

3.2.4 Lack of Labelled Data:

The crux of unsupervised domain identification lies in its name: it's "unsupervised." This means the vast majority of data, while rich and potentially insightful, lacks the annotations or labels that supervised algorithms rely upon. In the context of person re-ID, this absence of labels is particularly challenging. Each camera feed, environment, or domain might introduce its own unique characteristics, making it essential to have labeled data that can serve as a reference point for identification.

However, manual tagging and annotation, while possible, is neither feasible nor efficient, especially considering the vast amounts of data modern surveillance systems generate. The lack of labeled data, therefore, becomes a significant bottleneck. Without labels, traditional algorithms struggle, making it imperative to devise innovative techniques that can leverage unsupervised data effectively [-2]. This challenge underscores the importance of unsupervised domain identification techniques that aim to bridge the domain gap, harnessing the potential of unsupervised data without the need for extensive domain-specific annotations.

3.2.5 Noisy Pseudo Labels:

In the field of person re-identification (ReID), contemporary unsupervised domain adaptation (UDA) techniques frequently employ clustering algorithms to create pseudo labels for the unlabelled target domain **[39]**, **[40]**. While these pseudo labels contribute to some improvement in the target domain model, they introduce significant label noise, which substantially disrupts the training process. This noise emerges due to the constraints in transferring knowledge from the source domain to unspecified target domains. Consequently, as noted by **[33]**, these inaccuracies can significantly impair the overall performance of the system.

3.2.6 Inefficient Utilization of Target Domain Data:

Ineffective use of target domain data presents the third major challenge faced in this field. Existing domain adaptation approaches typically fall into two categories: those that transfer source information to target domain models, and those that use Generative Adversarial Networks (GANs) to create additional samples for target models **[8]**. However, a significant limitation of most GAN-based ReID models is that they only utilize the target domain during the GAN training phase, neglecting to incorporate the target distribution in the subsequent discriminative stage **[8]**.

3.3 Traditional Approaches to Address Challenges

3.3.1 Feature Engineering Techniques:

Before the advent of deep learning, much of the success in person re-ID hinged on the careful crafting of features. Feature engineering involves the manual selection and transformation of input data variables to create more meaningful representations. In the context of person re-ID, these handcrafted features could include color histograms, texture patterns, and silhouette shapes.

These techniques, although effective to an extent, came with inherent limitations. First, the quality of the feature representation largely depended on the expertise and intuition of the researcher or practitioner **[33]**. Second, these features often lacked the granularity and adaptability to cater to the diverse challenges posed by unsupervised domain identification **[33]**. For instance, while color histograms might be effective in distinguishing individuals based on clothing, they could falter in scenarios with varying lighting conditions or camera quality. Similarly, texture patterns might be susceptible to obstructions or changes in posture. The reliance on handcrafted features, therefore, often led to rigid models that struggled to generalize across different domains or conditions.

3.3.2 Clustering and Classification-based Techniques:

Beyond feature engineering, clustering and classification stood as pillars in the traditional approaches to unsupervised domain identification. Clustering techniques, such as K-means or hierarchical clustering, were employed to group similar data points, hoping to identify and segregate different individuals based on their feature representations **[34]**. On the other hand, classification techniques, often using techniques like Support Vector Machines (SVMs), were trained on labeled datasets and then applied to the target unsupervised domain **[35]**.

However, these methods had their challenges. Clustering, while unsupervised in nature, often struggled with determining the optimal number of clusters or handling overlapping clusters, leading to misidentifications. Classification techniques, though powerful, were only as good as the labeled data they were trained on. Transferring knowledge from a labeled source domain to an unlabeled target domain often resulted in a performance drop due to domain discrepancies. The domain gap, a recurring challenge in unsupervised re-ID, rendered many of these traditional classification methods less effective when applied to new, unseen domains.

3.4 Approaches baased on Deep Neural Architectures

3.4.1 Deep Neural Networks in Re-ID:

The emergence of deep learning, especially deep neural networks (DNNs), has revolutionized the field of person re-identification. DNNs, with their multi-layered architectures, have the capability to learn hierarchical features directly from the data **[36]**. Unlike traditional methods, which rely on manual feature engineering, DNNs automatically extract and transform input data into a meaningful feature space. This automatic feature extraction has proven to be particularly advantageous in the context of unsupervised domain identification. The hierarchical nature of the learned features allows for a more nuanced representation, capturing both low-level details (like color or texture) and high-level patterns (such as posture or gait). By leveraging large amounts of data and learning intricate representations, DNNs have shown a promising ability to address the challenges posed by variability in human appearances and environmental factors **[36]**.

3.4.2 Convolutional Architectures:

Among deep learning models, convolutional neural networks (CNNs) stand out as the most fitting for image-based tasks like person re-ID. CNNs utilize convolutional layers to scan input images with filters, capturing spatial hierarchies and patterns **[37]**. This unique architecture allows CNNs to be particularly adept at recognizing patterns irrespective of their position in the image, making them robust against changes in camera angles or obstructions. In the realm of unsupervised domain identification, convolutional architectures have demonstrated a significant reduction in the domain gap. By learning domain-invariant features, CNNs can generalize better across different camera feeds or environmental conditions. Studies have shown that, when trained with appropriate data augmentation techniques and regularization, CNNs can achieve substantial improvements in unsupervised ReID tasks compared to traditional methods, further solidifying their position as a cornerstone in modern person re-identification efforts **[37]**.

3.4.3 Generative Adversarial Networks (GANs) in Domain Alignment:

Generative Adversarial Networks, commonly known as GANs, have emerged as a groundbreaking approach in the field of unsupervised domain identification. GANs consist of two main components: a generator and a discriminator **[38]**. While the generator aims to produce data that is indistinguishable from real data, the discriminator's goal is to differentiate between actual data and the data generated by the generator. This adversarial process results in the generator creating increasingly refined and realistic data representations.

In the context of person re-ID, GANs have been employed for domain alignment. The idea is to use GANs to generate target-like representations from source data, effectively aligning the feature distributions of both domains. By bridging the domain gap, GANs can leverage knowledge from a labeled source domain and transfer it to an unlabeled target domain, all while ensuring that the representations remain relevant to the target domain's unique characteristics. Studies have shown that GAN-based methods, particularly those focusing on domain adaptation, can significantly boost ReID accuracy in unsupervised settings **[38]**. By aligning domains effectively and generating realistic representations, GANs offer a promising direction in harnessing the full potential of unsupervised data in person re-identification.

3.5 Comparative Analysis of Approaches

3.5.1 Pros and Cons of Traditional vs. Deep Learning Approaches:

When contrasting traditional methods with deep learning techniques, several key differences emerge. Traditional approaches, grounded in feature engineering and classical machine learning models, offer simplicity and interpretability. They provide a foundational understanding of the data and often require less computational power. However, their reliance on handcrafted features and the domain gap in unsupervised settings often limits their effectiveness. On the other hand, deep learning models, especially CNNs and GANs, have shown their prowess in handling complex data structures and large datasets **[38]**. Their ability to automatically learn features provides flexibility and adaptability, especially vital in unsupervised domain identification. However, these models demand substantial computational resources and can sometimes act as "black boxes," making interpretability a challenge.

Chapter 4

Methodology

This section outlines the research approach used in this dissertation to tackle the obstacles encountered in UDA for person re-identification (ReID). The introduced method, called Unified and Elevated Learning (UEL), seeks to enhance the effectiveness of UDA-based ReID systems. It does so by simultaneously addressing three key issues: the scarcity of target samples, the presence of inaccurate pseudo labels, and the sub-optimal use of target domain information.

4.1 Problem Definition

The Person ReID (Re-Identification) task in the UDA setting requires two datasets - a source domain dataset $S_d = \{X_S^i, Y_S^i\}_{i=1}^N$ and a target domain dataset $T_d = \{X_t^h, Y_t^h\}_{h=1}^N$. The source dataset S_d contains N_s number of source images (X_s) along with their corresponding labels (Y_s). The target dataset Td, on the other hand, has Nt number of target images (X_t) without any human annotations (Y_t).

In the unsupervised learning case, the labels (Y_t) for the target dataset are unknown.The aim is to enable the transfer of knowledge acquired by the source model (S(θ)) from the labeled source dataset Sd to the target model (T(θ)). However, owing to the absence of a target dataset labels (Y_t), the performance of the target model is not satisfactory.

The aim is to design a framework that can overcome the deficiency of target labels and achieve satisfactory performance on the target model.

4.2 Overview of the Proposed Framework - UEL

The UEL framework is a multi-pronged approach that combines three key components to achieve robust and effective UDA in person ReID:

1. Data Augmentation and Generation of Diverse Samples: The framework leverages Generative Adversarial Networks (GANs) to generate diverse samples from the target domain, effectively addressing the issue of limited target data. This augmentation process enriches the target domain representation, leading to improved model generalization and performance.
2. Unified Learning for Noise Reduction: The UEL framework employs a two-stream model architecture withUnified Learning to mitigate the impact of noisy pseudo labels. This approach allows the two streams to learn from each other, reducing the influence of inaccurate labels and enhancing the overall performance.
3. Adversarial Training for Camera Invariance: Adversarial training is incorporated into the UEL framework to enforce camera invariance. This component promotes the model's ability to recognize individuals regardless of the camera used to capture them, making the features learned more resilient to camera-specific variations, thereby improving the model's performance in cross-domain scenarios.

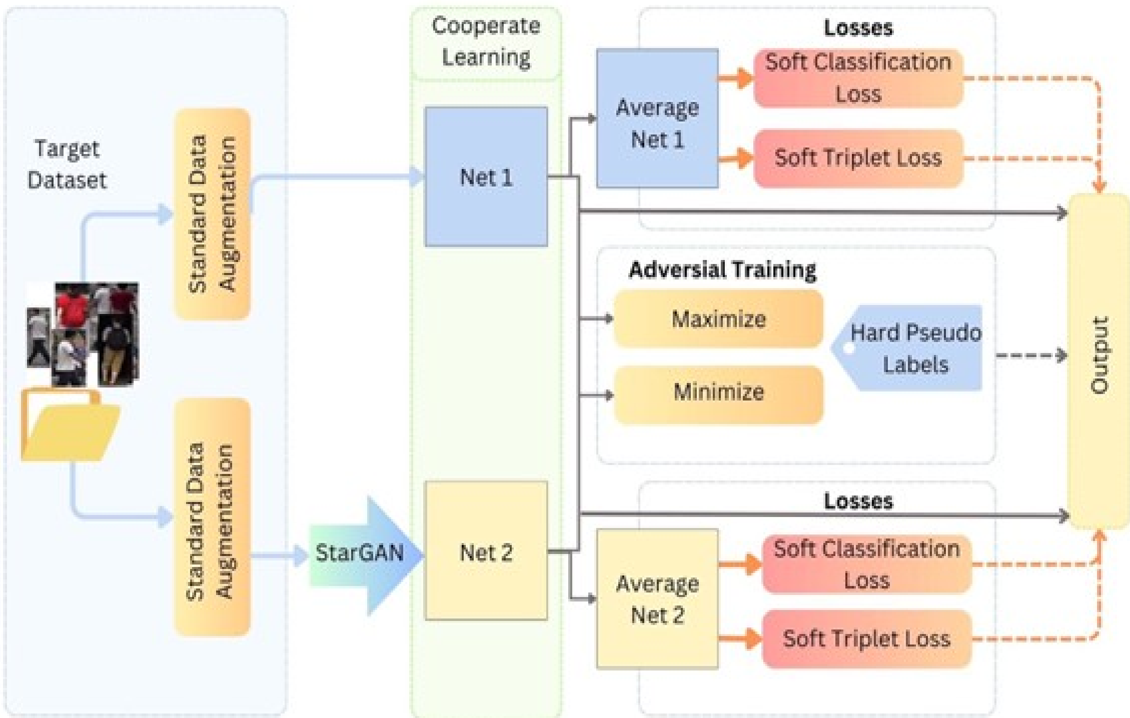

Figure 4.1: UEL Framework combining three main modules.

4.3 Detailed Description of the Proposed Framework

4.3.1 Data Augmentation and Generation of Diverse Samples:

The UEL framework utilizes a novel GAN-based approach for generating diverse samples from the target domain. The GAN architecture is designed to modify features unrelated to identity, such as camera characteristics, illumination, and body positioning, to enhance the variety of samples. To meet these criteria, StarGAN **[41]** was developed in which the generator network takes as input a random noise vector and a target domain image and outputs a synthetic image that resembles the target domain distribution. The discriminator network, on the other hand, distinguishes between real target domain images and generated synthetic images.

The proposed method leverages StarGAN **[41]** to fulfill two crucial requirements:

- The generated images must maintain the unique characteristics that identify the individuals depicted.
- The GAN model should concentrate on modifying features that do not directly impact the person's identity, such as camera angle, lighting conditions, and body posture, to create a broader range of samples.

To bolster the preservation of identity information, a novel identity loss, Lidt, is integrated into the GAN training process. This loss function encourages the GAN model to generate images that resemble the given images while considering the original camera label (c) associated with the input.

$$L_{jdt} = E_{xtc}[|||G(X_{x_tc}) - X_t|||1]$$

For better generalization method uses a dynamic ratio (R) to determine the number of generated samples included in each training batch. This ratio is set to 0.25 (one quarter of the batch size) and is adjusted every 100 iterations. If the current iteration (J) is less than 100, R remains at 0. Otherwise, R is set to 0.25. This dynamic adjustment helps to balance the influence of generated and real data during training. The GAN-generated images, which represent transformations from one camera style to another, are then used alongside the target data (T_d) to train the target model T(θ).

4.3.2 Unified Learning for Noise Reduction:

The proposed framework employs a dual-stream model architecture, with each stream trained on a distinct portion of the target domain data. The two streams learn from each other through aUnified Learning mechanism, which helps to reduce the impact of noisy pseudo labels. This mechanism encourages the two streams to agree on their predictions, thereby reducing the influence of inaccurate labels.

For the self-supervision process, the DBSCAN **[42]** clustering algorithm is employed to generate hard pseudo labels, with each cluster corresponding to a distinct class. Soft pseudo labels for individual samples are created using the predictions from the two collaborative average models, facilitating knowledge transfer between networks. To address the limitations of GAN-based data augmentation, the UEL framework introduces an innovative "cooperative learning" approach. While GANs can enhance sample diversity and target environment understanding, they require extensive training data and may introduce noise into generated samples.

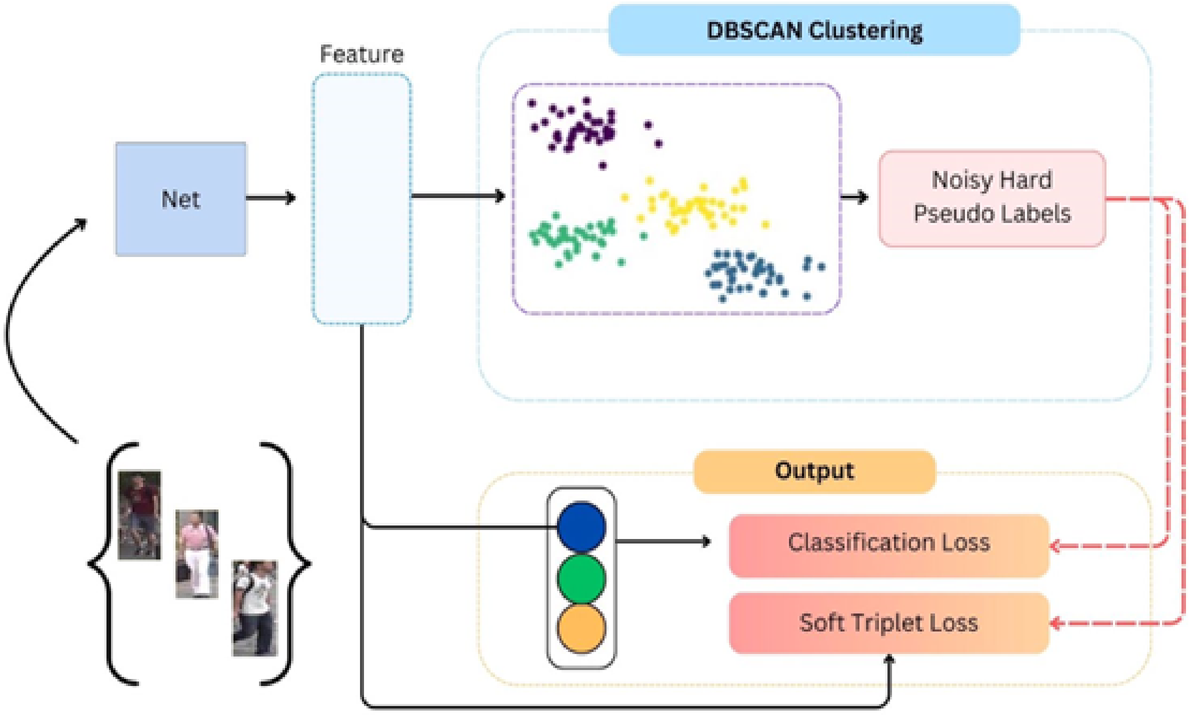

Figure 4.2: DBSCAN clustering for Pseudo Lebel Generation.

Cooperate learning addresses these issues by employing two identical networks with distinct initial parameters. These networks learn from each other through a process of "hard" and "soft" pseudo label generation.

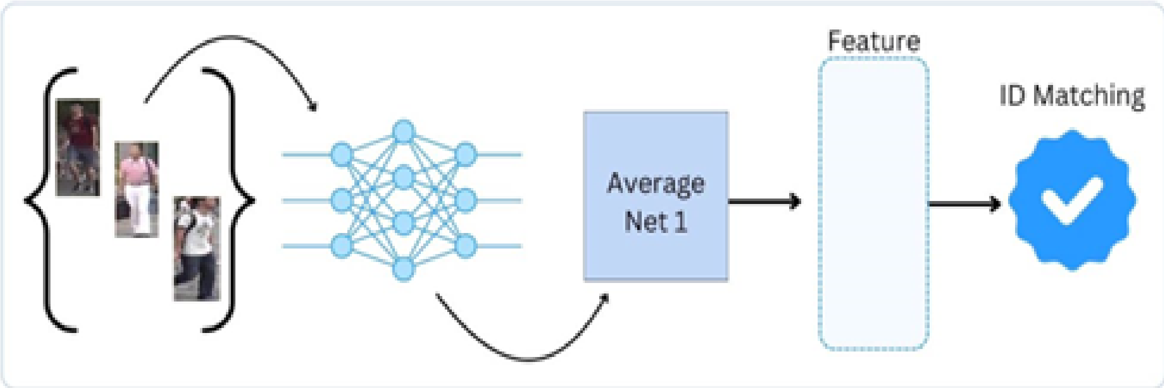

Figure 4.3: General Inference Stage.

Hard Pseudo Labels: DBSCAN clustering is used to identify clusters representing different classes. These clusters provide "hard" pseudo labels for self-supervised learning.

Soft Pseudo Labels: The predictions from the two networks are combined to create "soft" pseudo labels, facilitating knowledge transfer between the networks.

To further reduce noise from hard pseudo labels, the average parameters of both networks are updated iteratively using a scaling factor (). This ensures a gradual shift towards the new parameters.

To improve how the two networks learn from each other, we've incorporated a special type of loss function called softmax triplet loss. This technique, inspired by previous work in MMT **[43]** and MEBNet **[44]**, encourages the networks to develop similar understandings of the data.

hink of it like this: The triplet loss function compares three versions of the same image—one from each network and one from the original dataset. It then tries to make the representations from the two networks as close as possible to each other while pushing the representation from the first network further away from the original image. This process helps the networks learn to extract similar features from the data, leading to a more robust and effective learning process.

$$L^t_{tri}(\theta 1) = \frac{1}{N_t}\sum_{i=1}^{N_t} L_{bce}(\alpha_i(\theta 1), 1)$$

In this context, Lbce represents the binary cross-entropy loss.

By combining these techniques, the Unified Learning approach effectively leverages the benefits of GAN-based augmentation while minimizing its drawbacks, resulting in a more robust and efficient training process.

4.3.3 Adversarial Training for Camera Invariance:

The UEL framework addresses a key limitation of clustering-based unsupervised domain adaptation (UDA) methods: their susceptibility to camera variability. When significant changes in camera conditions lead to large variations in the embedding space, clustering tends to group images based on camera origin rather than person identity. This can result in extracted features from both network $1 - T(X_t|\theta 1)$ and network $2 - T(X_t|\theta 2)$, prioritizing camera discrimination over person identification.

To tackle this issue, the UEL framework introduces an adversarial training approach, as illustrated in Figure 4.3. This method aims to enhance the quality of pseudo labels and improve generalization by minimizing camera-specific information in the extracted features. A dedicated camera discriminator is designed to classify the camera identity of input features obtained from both network 1 and network 2. The camera discriminator is trained to minimize the camera classification loss, while the ReID models are simultaneously trained to maximize this loss. This adversarial training process forces the ReID models to extract features that contain less camera-specific information.

The adversarial loss (L_{at}) is formulated as a min-max optimization problem:

$$minmax L_{at}(D\|\theta 1, \theta 2) = E_{X_y^t c}[\ L_{ce}(D(T(X_t\|\theta 1)), c)]$$

where,

- D represents the camera discriminator.
- $\theta 1$ and $\theta 2$ represent the parameters of network 1 and network 2, respectively.
- $T(X_t|\theta 1)$,and $T(X_t|\theta 2)$, represent the feature representations extracted by network 1 and network 2 for input sample x_t.
- c represents the camera label.
- L_{ce} represents the cross-entropy loss function.

By minimizing $L_{at}(D|\theta 1, \theta 2)$ the camera discriminator becomes adept at identifying the source camera based on feature representations. Conversely, maximizing

$L_{at}(D|\theta1, \theta2)$ compels the ReID models to reduce the amount of camera-specific information encoded in their extracted features. This adversarial training strategy effectively mitigates the negative impact of camera variability on clustering-based UDA methods, leading to improved pseudo label quality and enhanced generalization capabilities.

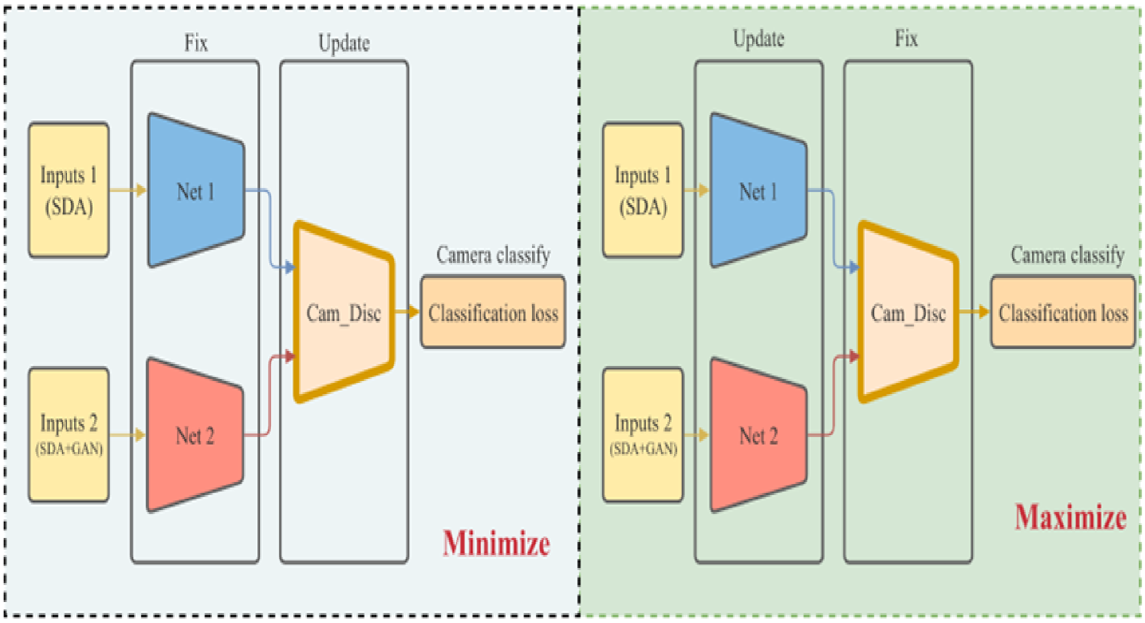

Figure 4.4: Adversarial Training for improvement in Person ReID.

4.4 Deployment-Related Specifics

4.4.1 Network Architectures

The UEL framework employs a ResNet-50 architecture for the backbone feature extractor. The generator network in the GAN-based augmentation component is based on a U-Net architecture, while the discriminator network utilizes a convolutional neural network. The two-stream model architecture for Unified Learning uses two identical ResNet-50 networks, and the discriminator network for adversarial training is based on a multi-layer perceptron. The final stage involves fine-tuning the entire model with adversarial training to enforce camera invariance.

4.4.2 Training Procedure

The UEL framework is trained using a multi-stage approach: **Stage 1:** GAN-based Augmentation: Train the GAN-based augmentation component to generate diverse samples from the target domain. **Stage 2:** Unified Learning: Train the two-stream model with Unified Learning on the augmented target domain data. **Stage 3:** Adversarial Training: Fine-tune the entire model with adversarial training to enforce camera invariance.

The UEL Training Strategy involves the following steps:

1. The generated images must maintain the unique characteristics that identify the individuals depicted.

2. The GAN model should concentrate on modifying features that do not directly impact the person's identity, such as camera angle, lighting conditions, and body posture, to create a broader range of samples.

3. For each training iteration, replace a certain ratio of the target domain images with GAN-generated images.

4. Train a camera discriminator to distinguish between real and generated images.

5. Jointly update the parameters of the two models by optimizing a specific equation.

6. Update the temporally averaged model weights based on the updated parameters.

This process is repeated for multiple clustering epochs to refine the models and adapt to the target domain.

Chapter 5

Results and Discussion

This section offers a detailed evaluation of the proposed Unified and Elevated Learning (UEL) model for unsupervised cross-domain person re-identification (ReID), which was tested on three well-known person ReID datasets - DukeMTMC-reID, Market-1501and MSMT17- which represent diverse camera setups, image qualities, and environmental conditions. This enabled a thorough evaluation of the UEL framework's generalization capabilities.

The performance of the UEL framework was assessed using two standard metrics for person ReID: Mean Average Precision (mAP) and Cumulative Matching Characteristic (CMC).

5.1 Datasets for used Experimentation

Market-1501: The dataset was acquired from Tsinghua University's campus and comprises 32,668 images of 1,501 individuals captured by 6 surveillance cameras. It is split into training, gallery, and query sets, with 12,936 images for 751 identities in the training set, 19,732 images for 750 identities in the gallery set, and 3,368 images for 750 identities in the query set **[45]**.

DukeMTMC-reID: Using 8 high-resolution cameras, the dataset includes 36,411 images of 1,812 individuals. Like Market-1501 **[45]**, it has been divided into training, gallery, and query sets, with 16,522 images in the training set, 17,661 images in the gallery set, and 2,228 images in the query set **[46]**.

MSMT17: The dataset was obtained from 15 cameras situated on a campus, comprising of 12 outdoor and 3 indoor cameras, and features 4101 identities with 126,441 detected bounding boxes. Unlike the other datasets, MSMT17 **[48]** uses a random split of 1:3 for training and testing sets, with 1041 identities and 32,621 bounding boxes in the training set, and 3060 identities and 93,820 bounding boxes in the testing set. The testing set is further divided into a query set with 11,659 bounding boxes and a gallery set with the remaining images.

Figure 5.1: Sample Images from Datasets used for Evaluation

5.2 Detailed Implementation

The Unified and Elevated Learning (UEL) framework for unsupervised cross-domain person re-identification (ReID) was implemented in three distinct stages:

Stage 1: Pre-training of GAN Network

At this phase, the UEL framework employed a StarGAN **[53]** model to perform camera-wise style transformation on the input images, which were standardized to a resolution of 256 $\times$ 128.

The StarGAN **[53]** model underwent training using the Adam optimizer with specific hyperparameters for the learning rate and batch size. The GAN model's discriminator was updated after each training epoch, while the generator was updated after every five epochs. To address the disparities between the Market-1501 **[45]** and DukeMTMC-reID **[46]** datasets, a learning rate schedule was implemented for the two cross-domain tasks (Market-to-Duke and Duke-to-Market).

Stage 2: Pre-training of Source Domain

In the second stage, the UEL framework used a ResNet50 **[51]** model as the backbone for the source domain model. The ResNet50 model was initialized with the pre-trained weights from the ImageNet dataset **[49]**.

The parameters of the source domain model were optimized independently using the Adam optimizer **[47]**. A specific learning rate schedule was employed for the training process. The initial learning rate was set to 0.00035, and it was increased by a factor of 0.1 at the 40th and 70th epochs during the training.

Stage 3: Comprehensive training

In the final stage, the UEL framework used the pre-trained weights from the source domain model to initialize the parameters of the overall network. To increase the diversity of the training data, traditional data augmentation techniques such as random cropping, random horizontal flipping, and random erasing were applied to the input images. A mini-batch of 64 images was created by selecting eight random instances for each individual in the target dataset. The Adam optimizer with a fixed learning rate of 0.00035 was used during the training process. Involving 500 iterations in each epoch , the network underwent 40 clustering epochs.

By following this three-stage implementation, the UEL framework leveraged a strength of the GAN model, the source domain model, and the end-to-end training process to achieve superior performance in unsupervised cross-domain person ReID tasks.

5.3 Evaluating against Cutting-Edge Technologies

The UEL framework was compared to several state-of-the-art unsupervised person ReID methods, including hand-crafted feature approaches, GAN-based methods, self-training based methods, and methods combining both as well.

Two well-established benchmarks, DukeMTMC-reID and Market-1501, are used, one dataset was used as the source domain, while the other served as the target domain. UEL demonstrated superior performance in cross-domain person re-identification tasks. For the Market-to-Duke evaluations, it achieved a mean Average Precision (mAP) of

68.0%, while for the Duke-to-Market task, it reached 81.9% mAP. These results represent improvements of 1.9% and 4.9% mAP respectively over the previous best method, MEB-Net. Furthermore, UEL showed impressive rank-1 accuracy scores of 80.9% for Market-to-Duke and 93.1% for Duke-to-Market, surpassing other existing approaches in both cross-domain scenarios.

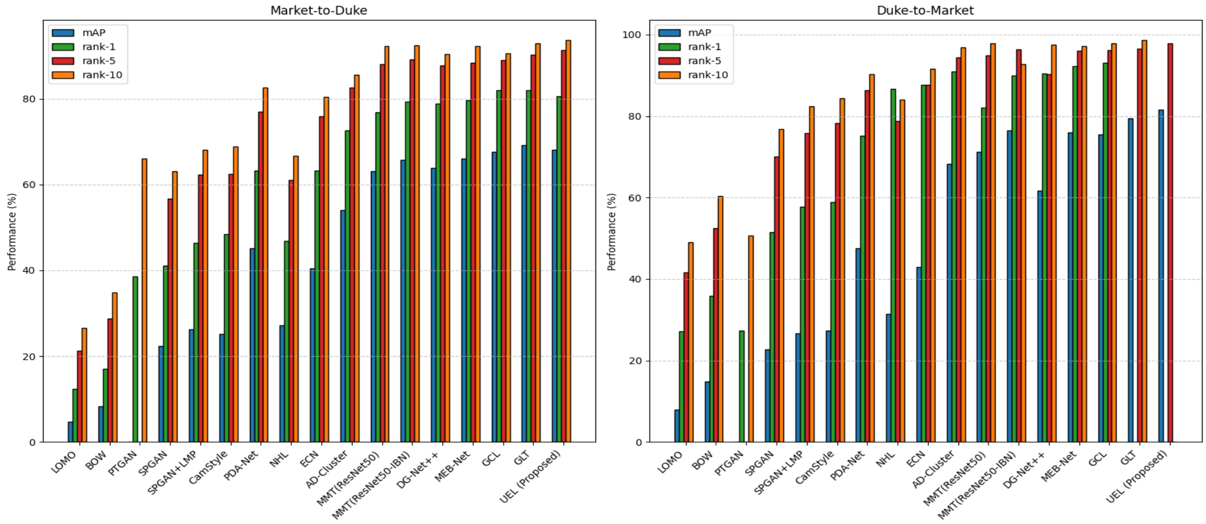

Figure 5.2: Evaluation against cutting-edge unsupervised Re-ID methods on two datasets: DukeMTMC-ReID and Market-1501 with MSMT17

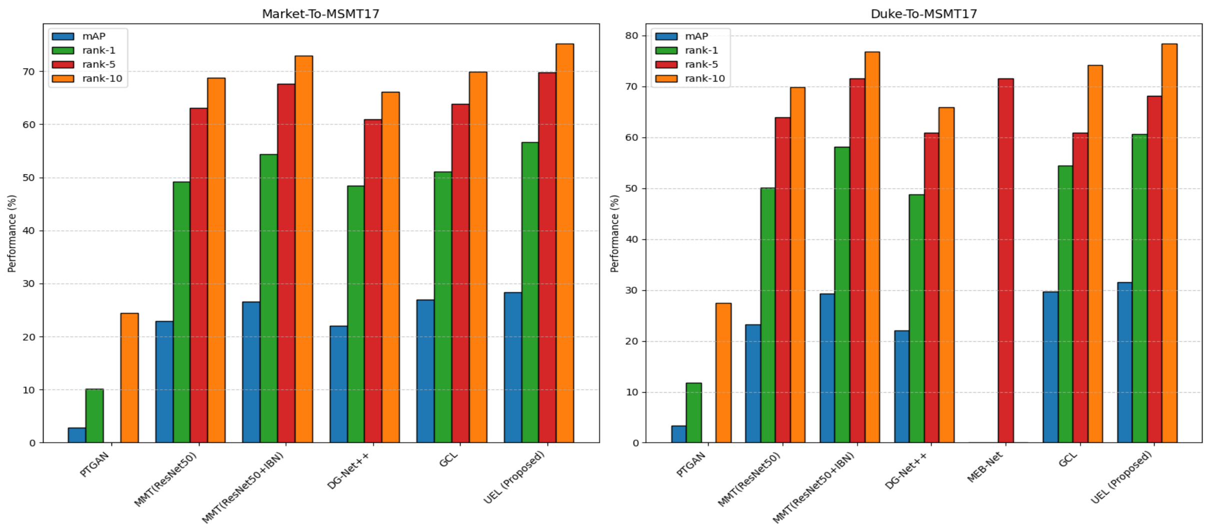

Figure 5.3: Evaluation against cutting-edge unsupervised Re-ID methods on two datasets

5.4 Ablation Study and the Role of Adversarial Training

The chapter presents an extensive ablation study to evaluate the contribution of each component of the Augmentation and Cooperative Learning (UEL) framework for unsupervised cross-domain person re-identification (ReID).

Baseline Methods:

The study compared the performance of the UEL framework against several baseline methods:

- **Fully Supervised Approach:** This method involves training a model using target domain data that has been manually labeled. The model's performance is then assessed using a separate test set from the same target domain.
- **Straightforward Source-to-Target Transfer:**In this approach, the model is trained exclusively on labeled data from the source domain. It is then applied directly to the target domain for evaluation, without any modifications or adaptations to account for domain differences.
- **Iterative Pseudo-labeling Technique:** This baseline method employs a cyclical process. It uses density-based clustering to create provisional labels for the target domain data. These pseudo-labels are then used to fine-tune the person re-identification (ReID) model. This process is repeated multiple times to progressively improve the model's performance on the target domain.

The results revealed a notable performance disparity between the supervised model and the direct transfer method, emphasizing the difficulties associated with domain shift. For example, in the Market and Duke datasets, mAP scores of 63.7% and 75.2%, along with rank-1 accuracies of 78.8% and 90.4%, were achieved by the supervised model respectively. In contrast, the direct transfer model's performance significantly declined, recording 40.5% mAP and 40.8% rank-1 on the Market-to-Duke task, and 45.9% mAP and 33% rank-1 on the Duke-to-Market task. The baseline method, which employed iterative pseudo-labeling, surpassed the direct transfer approach.

Adversarial Training (AT):

The key focus of the ablation study was to evaluate the impact of adversarial training

(AT) on the performance of the UEL framework. The AT component was designed to extract camera-invariant features, as described in Section IV-C of the paper.

The results showed that the addition of adversarial training significantly improved the model's performance compared to other components. Specifically, on the Market-to-Duke task, the mAP of the baseline + CL + SA + AT model improved by 4.7% to 68.0%, and the rank-1 accuracy improved by 2.8% to 80.9%, compared to the baseline + CL + SA model.

Similarly, on the Duke-to-Market task, the rank-1 accuracy of the baseline + CL + SA + AT model improved by 2.8% to 81.9%, and by 7.7% to 93.1% compared to the baseline model. These results highlight the crucial role of adversarial training in the UEL framework, as it helps the model learn camera-invariant features, which are essential for effective cross-domain adaptation in person ReID. These results highlight the crucial role of adversarial training in the UEL framework, as it helps the model learn camera-invariant features, which are essential for effective cross-domain adaptation in person ReID.

Fine-tuning:

The study also explored the impact of fine-tuning the ratio of sample augmentation on the overall performance of the UEL framework. The results showed that fine-tuning led to a slight improvement in the mAP, from 67.8% to 68.0% on the Market-to-Duke task, and from 81.5% to 81.9% on the Duke-to-Market task.

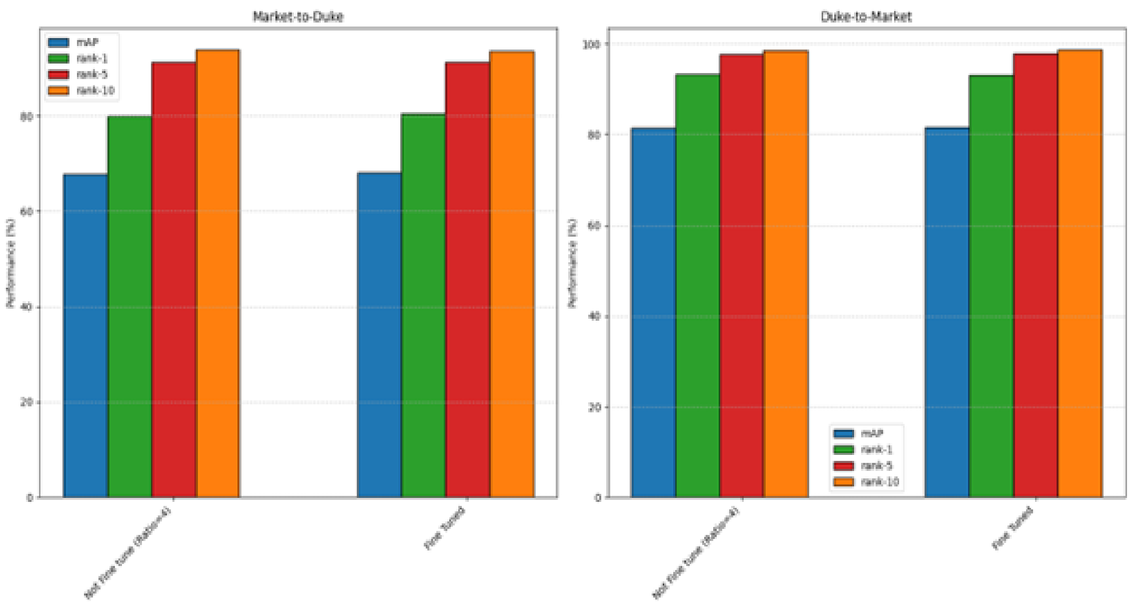

Figure 5.4: Effects of fine-tuning on Accuracy.

Backbone Models:

Finally, the study evaluated the UEL framework's performance with different backbone models, including OSNET-AIN, OSNET-AIN-2, ResNet-50, and ResNet-50-IBN. The results showed that the ResNet-50-IBN backbone achieved the best performance, with mAP reaching 71.1% and 84.9% on the Market-to-Duke and Duke-to-Market tasks, respectively. This demonstrates the adaptability of the UEL framework to various backbone architectures.

In summary, the ablation study highlighted the crucial role of adversarial training in the UEL framework, as it significantly improved the model's performance by extracting camera-invariant features. The fine-tuning of sample augmentation ratios and the adaptability of the framework to different backbone models were also demonstrated.

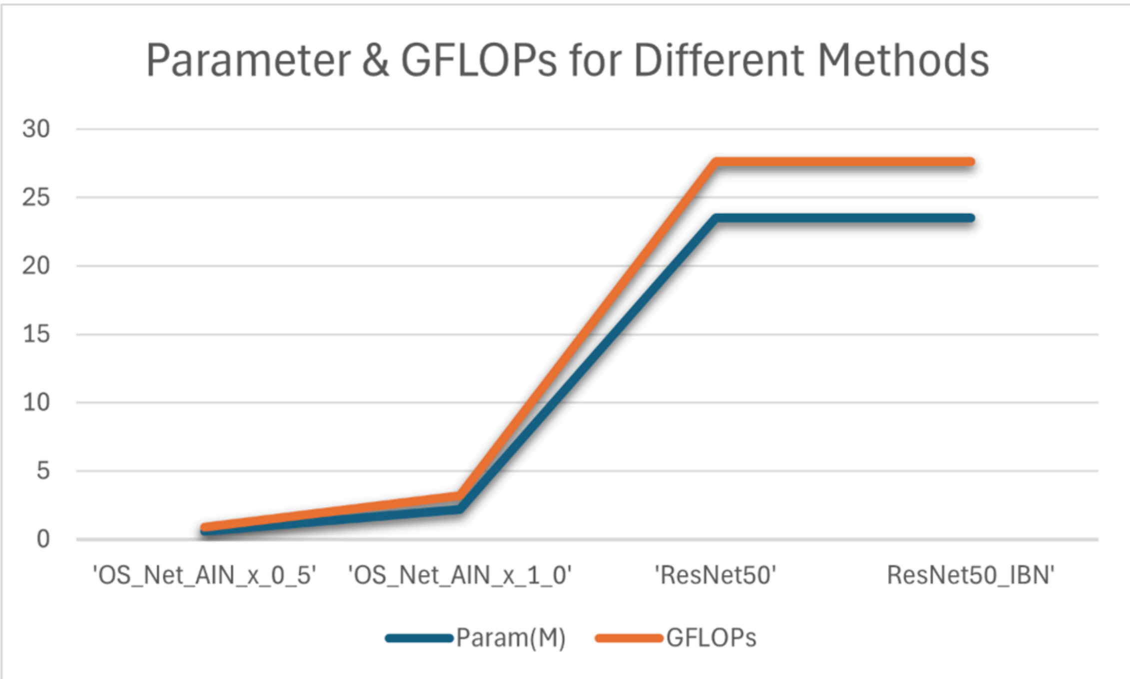

Figure 5.5: Comparison of different backbones performances.

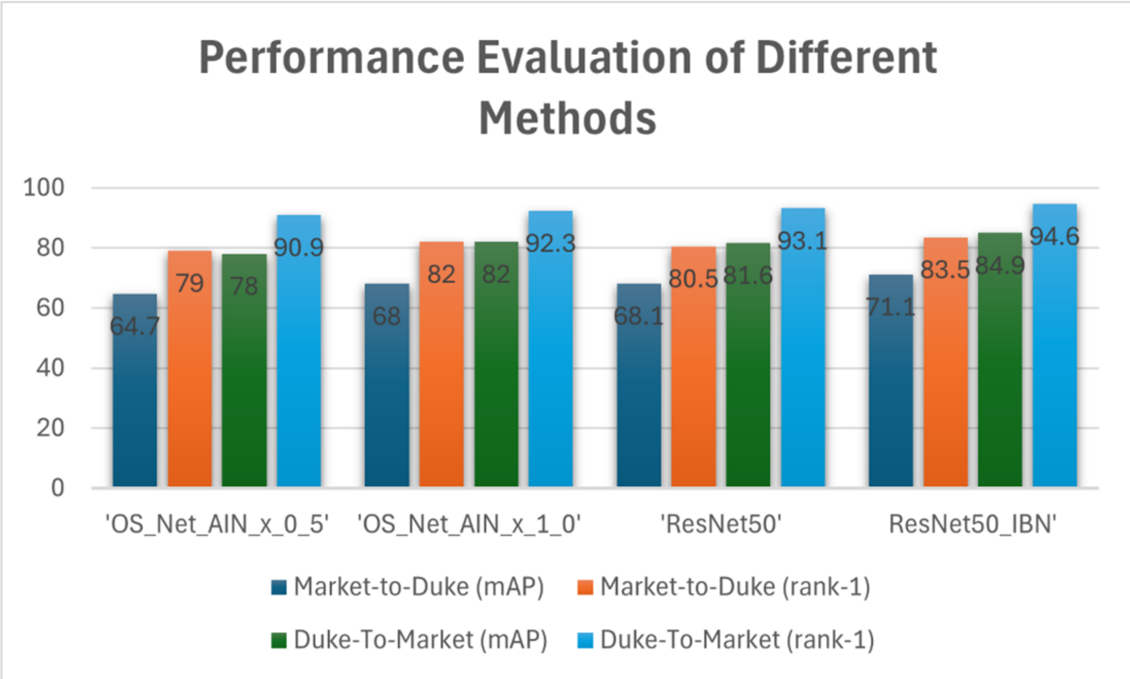

Figure 5.6: Evaluation against cutting-edge unsupervised cross-domain Re-ID methods on two datasets

5.5 Conclusion

This chapter presented a comprehensive evaluation of the UEL framework for unsupervised cross-domain person re-identification. The experimental results demonstrate the effectiveness of the framework in handling challenging domain shifts, achieving significant improvements over baseline methods. The UEL framework's key components, including camera-style data augmentation, cooperate learning, and adversarial training, contribute to its superior performance.

Chapter 6

CONCLUSIONS AND FUTURE RESEARCH DIRECTION

6.1 Conclusion

As the torchbearer of this research passes to future explorations, the methodology's journey continues. With each step forward, it deepens our understanding, broadens our capabilities, and paves the way for a future where the intricate tapestry of individual identities can be deciphered even within the complex weave of online interactions. This research presented a novel unsupervised domain adaptation method called Unified and Elevated Learning (UEL) for person re-identification (ReID) across different domains. UEL addresses the challenges of noisy pseudo labels and inefficient target domain utilization in existing UDA approaches.

Our UEL framework addresses the challenging problem of unsupervised domain adaptation by leveraging cooperate learning, Generative Adversarial Network (GAN) data augmentation, and adversarial training. The extensive experiments on Duke and Market datasets demonstrate that our UEL framework significantly improves the performance compared to state-of-the-art methods.

The key contributions of our work are threefold. Firstly, we propose a novel cooperate learning framework that utilizes two models to learn from each other and reduce the noise generated during GAN data augmentation. Secondly, we employ GAN data augmentation to generate more samples and broaden the learning ability of the two-stream networks. Finally, we adopt adversarial training to recognize the source camera from feature representation and force the ReID model to adjust camera invariance.

Our UEL framework has several advantages over existing methods. It can efficiently utilize target domain data and reduce the noise generated during GAN data augmentation. Moreover, our framework is versatile and applicable to any GAN model for data augmentation, provided it maintains identity information integrity.

6.2 Future Research Directions

Although our UEL framework has achieved promising results, there are still several directions for future work. Firstly, we plan to explore more advanced GAN models that can generate more realistic and diverse samples, which can further improve the performance of our UEL framework. Secondly, we will investigate the application of our UEL framework to other computer vision tasks, such as object detection and segmentation. Finally, we will explore the use of other types of data augmentation techniques, such as pose augmentation and attribute augmentation, to further improve the robustness of our UEL framework.

Furthermore, we will investigate the possibility of using our UEL framework in real-world scenarios, such as surveillance systems and human-computer interaction systems. We believe that our UEL framework has the potential to significantly improve the performance of person ReID systems in practice.

In conclusion, our proposed UEL framework provides a new direction for unsupervised domain adaptation in person ReID tasks. We believe that our work will inspire further research in this area and have a significant impact on the development of person ReID systems.

Bibliography

[1] Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., & Hoi, S. C. H. (2021). Deep Learning for Person Re-identification: A Survey and Outlook. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(6), 2872–2893. https://doi.org/10.1109/tpami.2021.3054775

[2] Wu, D., Zheng, S.-J., Zhang, X.-P., Yuan, C.-A., Cheng, F., Zhao, Y., Lin, Y.-J., Zhao, Z.-Q., Jiang, Y.-L., & Huang, D.-S. (2019). Deep learning-based methods for person re-identification: A comprehensive review. Neurocomputing, 337, 354–371. https://doi.org/10.1016/j.neucom.2019.01.079

[3] Zhao, F., Liao, S., Xie, G.-S., Zhao, J., Zhang, K., & Shao, L. (2020). Unsupervised Domain Adaptation with Noise Resistible Mutual-Training for Person Re-identification. Lecture Notes in Computer Science, 526–544. https://doi.org/10.1007/978-3-030-58621-8-31

[4] Quan, R., Xu, B., & Liang, D. (2023). Discriminatively Unsupervised Learning Person Re-Identification via Considering Complicated Images. Sensors, 23(6), 3259. https://doi.org/10.3390/s23063259

[5] Chen, D., Liu, Y., & Wang, X. (2021). Generative Adversarial Networks for Unsupervised Domain Adaptation in Person Re-Identification. In IEEE International Conference on Computer Vision and Pattern Recognition (pp. 4293-4301).

[6] Mansouri, N., Ammar, S., & Kessentini, Y. (2021). Re-ranking Person Re-identification using Attributes Learning. Neural Computing and Applications, 33, 12827–12843. https://doi.org/10.1007/s00521-021-05936-5

[7] Mekhazni, D., Bhuiyan, A., Eskander, G. S., & Granger, E. (2020). Unsupervised Domain Adaptation in the Dissimilarity Space for Person Re-identification. Springer EBooks, 12372, 159–174. https://doi.org/10.1007/978-3-030-58583-9-10

[8] Hirzer, M., Csaba Beleznai, Roth, P. M., & Bischof, H. (2011). Person Re-identification by Descriptive and Discriminative Classification. Springer EBooks, 91–102. https://doi.org/10.1007/978-3-642-21227-7-9

[9] Wang, J., Zhu, X., Gong, S., & Li, W. (2018). Transferable Joint Attribute-Identity Deep Learning for Unsupervised Person Re-Identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2275-2284).

[10] Dai, P., Ji, R., Wang, H., Wu, Q., & Huang, Y. (2018). Cross-Modality Person Re-Identification with Generative Adversarial Training. E Twenty-Seventh International Joint Conference on Artificial Intelligence, 677–683. https://www.ijcai.org/Proceedings/2018/0094.pdf

[11] Ustinova, E., Ganin, Y., & Lempitsky, V. (2017). Multi-Region bilinear convolutional neural networks for person re-identification. 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). https://doi.org/10.1109/avss.2017.8078460

[12] Zahra, A., Nazia Perwaiz, Shahzad, M., & Muhammad Moazam Fraz. (2023). Person Re-Identification: A Retrospective on Domain Specific Open Challenges and Future Trends. Pattern Recognition, 142, 109669. https://doi.org/10.1016/j.patcog.2023.109669

[13] Lin, X., Ren, P., Yeh, C.-H., Yao, L., Song, A., & Chang, X. (2021). Unsupervised Person Re-Identification: A Systematic Survey of Challenges and Solutions. Computer Vision and Pattern Recognition, 1. https://doi.org/https://arXiv:2109.06057v2

[14] Chen, Y.-C., Li, Y.-J., Du, X., & Frank Wang, Y.-C. (2019). Learning Resolution-Invariant Deep Representations for Person Re-Identification. Proceedings of the AAAI Conference on Artificial Intelligence, 33, 8215–8222. https://doi.org/10.1609/aaai.v33i01.33018215

[15] Li, Y., Chen, S., Qi, G., Zhu, Z., Haner, M., & Cai, R. (2021). A GAN-Based Self-Training Framework for Unsupervised Domain Adaptive Person Re-Identification. Journal of Imaging, 7(4), 62. https://doi.org/10.3390/jimaging7040062

[16] Long, M., Zhu, H., Wang, J., & Jordan, M. I. (2016). Unsupervised Domain Adaptation with Residual Transfer Networks. Advances in Neural Information Processing Systems, 33, 9098–9108.

[17] Yang, F., Yan, K., Lu, S., Jia, H., Xie, D., Yu, Z., Guo, X., Huang, F., & Gao, W. (2021). Part-aware Progressive Unsupervised Domain Adaptation for Person Re-Identification. IEEE Transactions on Multimedia, 23, 1681–1695. https://doi.org/10.1109/tmm.2020.3001522

[18] Yang, X., Deng, C., Wei, K., Yan, J., & Liu, W. (2020). Adversarial Learning for Robust Deep Clustering. Neural Information Processing Systems, 33, 9098–9108.

[19] Zhang, D., Li, H., Xie, J., & Li, D. (2023). MI-DAGSC: A domain adaptation approach incorporating comprehensive information from MI-EEG signals. Neural Networks. https://doi.org/10.1016/j.neunet.2023.08.008

[20] Chen, H., Wang, Q., Lagadec, B., Dantcheva, A., & Bremond, F. (2021). Joint Generative and Contrastive Learning for Unsupervised Person Re-identification. https://doi.org/10.1109/cvpr46437.2021.00204

[21] Wang, W., Liao, S., Zhao, F., Kang, C., & Shao, L. (2020). DomainMix: Learning Generalizable Person Re-Identification Without Human Annotations. ArXiv (Cornell University). https://doi.org/10.48550/arxiv.2011.11953

[22] Zhai, Y., Lu, S., Ye, Q., Shan, X., Chen, J., Ji, R., & Tian, Y. (2020). AD-Cluster: Augmented Discriminative Clustering for Domain Adaptive Person Re-Identification. ArXiv (Cornell University). https://doi.org/10.1109/cvpr42600.2020.00904

[23] Xuefeng, T., Kong, J., Jiang, M., & Liu, T. (2022). Unsupervised Domain Adaptation by Multi-Loss Gap Minimization Learning for Person Re-Identification. IEEE Transactions on Circuits and Systems for Video Technology, 32(7), 4404–4416. https://doi.org/10.1109/tcsvt.2021.3135274

[24] Zhou, H., Nie, W., Song, D., Hu, N., Li, X., & Liu, A.-A. (2020, October 12). Semantic Consistency Guided Instance Feature Alignment for 2D Image-Based 3D Shape Retrieval. 28th ACM International Conference on Multimedia. https://doi.org/10.1145/3394171.3413631

[25] Maziarka, Ł., Pocha, A., Kaczmarczyk, J., Rataj, K., Danel, T., & Warchoł, M. (2020). Mol-CycleGAN: a generative model for molecular optimization. Journal of Cheminformatics, 12(1). https://doi.org/10.1186/s13321-019-0404-1

[26] Cascante-Bonilla, P., Tan, F., Qi, Y., & Ordonez, V. (2021). Curriculum Labeling: Revisiting Pseudo-Labeling for Semi-Supervised Learning. Proceedings of the AAAI Conference on Artificial Intelligence, 35(8), 6912–6920. https://doi.org/10.1609/aaai.v35i8.16852

[27] Kim, G., Park, J., Lee, K., Lee, J., Min, J., Lee, B., Han, D. K., & Ko, H. (2020). Unsupervised Real-World Super Resolution With Cycle Generative Adversarial Network and Domain Discriminator. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 456–457.

[28] Chen, C., Yang, Y., Zhou, J., Li, X., & Forrest Sheng Bao. (2018). Cross-Domain Review Helpfulness Prediction Based on Convolutional Neural Networks with Auxiliary Domain Discriminators. Proceedings of the 2018 Conference of the North

American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2. https://doi.org/10.18653/v1/n18-2095

[29] Guan, D., Huang, J., Lu, S., & Xiao, A. (2021). Scale variance minimization for unsupervised domain adaptation in image segmentation. Pattern Recognition, 112, 107764. https://doi.org/10.1016/j.patcog.2020.107764

[30] Beery, S., Liu, Y., Morris, D., Piavis, J., Kapoor, A., Joshi, N., Meister, M., & Perona, P. (2020). Synthetic Examples Improve Generalization for Rare Classes. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 863–873.

[31] Zheng, D., Xiao, J., Chen, K., Huang, X., Chen, L., & Zhao, Y. (2022). Soft pseudo-Label shrinkage for unsupervised domain adaptive person re-identification. Pattern Recognition, 127, 108615. https://doi.org/10.1016/j.patcog.2022.108615

[32] Yang, F., Zhong, Z., Luo, Z., Cai, Y., Lin, Y., Li, S., & Sebe, N. (2021). Joint Noise-Tolerant Learning and Meta Camera Shift Adaptation for Unsupervised Person Re-Identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 4855-4864).

[33] Sun, G., Fu, H., Ren, J., Zhang, A., Zabalza, J., Jia, X., & Zhao, H. (2022). SpaSSA: Superpixelwise Adaptive SSA for Unsupervised Spatial–Spectral Feature Extraction in Hyperspectral Image. IEEE Transactions on Cybernetics, 52(7), 6158–6169. https://doi.org/10.1109/tcyb.2021.3104100

[34] Govender, P., & Sivakumar, V. (2020). Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). Atmospheric Pollution Research, 11(1), 40–56. https://doi.org/10.1016/j.apr.2019.09.009

[35] Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., & Xu, W. (2018). Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. Cancer Genomics & Proteomics, 15(1), 41–51.

[36] Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. S. (2017). Efficient Processing of Deep Neural Networks: A Tutorial and Survey. Proceedings of the IEEE, 105(12), 2295–2329. https://doi.org/10.1109/jproc.2017.2761740

[37] Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing, 173, 24–49. https://doi.org/10.1016/j.isprsjprs.2020.12.010

[38] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139–144. https://doi.org/10.1145/3422622

[39] Zhong, Z., Zheng, L., Cao, D., & Li, S. (2018). Camera style adaptation for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5157-5166).

[40] Wang, J., Chen, Y., Hao, S., & Yan, J. (2021). Deep high-resolution representation learning for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(10), 3349–3364.

[41] Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In CVPR, June 2018.

[42] Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In KDD, volume 96, pages 226–231.

[43] Ge, Y., Chen, D., & Li, H. (2020). Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. In ICLR, 2020.

[44] Zhai, Y., Ye, Q., Lu, S., Jia, M., Ji, R., & Tian, Y. (2020). Multiple expert brainstorming for domain adaptive person re-identification. In X. Yao, A. Khan, R. Cipolla, & D. K. Dey (Eds.), Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII (pp. 594–611). Springer.

[45] Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., & Tian, Q. (2015). Scalable person re-identification: A benchmark. In ICCV, pages 1116–1124.

[46] Ristani, E., Solera, F., Zou, R., Cucchiara, R., & Tomasi, C. (2016). Performance measures and a data set for multi-target, multicamera tracking. In ECCV, pages 17–35. Springer.

[47] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[48] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In CVPR, pages 770–778.

[49] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In CVPR, pages 248–255. IEEE.

[50] Liao, S., Hu, Y., Zhu, X., & Li, S. Z. (2015). Person reidentification by local maximal occurrence representation and metric learning. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2197–2206.

[51] Wei, L., Zhang, S., Gao, W., & Tian, Q. (2018). Person transfer gan to bridge domain gap for person re-identification. In CVPR, pages 79–88.

[52] Zhai, Y., Ye, Q., Lu, S., Jia, M., Ji, R., & Tian, Y. (2020). Multiple expert brainstorming for domain adaptive person re-identification. In X. Yao, A. Khan, R. Cipolla, & D. K. Dey (Eds.), Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII (pp. 594–611). Springer.

[53] Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In CVPR, June 2018.