

Person Re-identification in Surveillance Video

by

Abhishek Shah
202011017

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of

MASTER OF TECHNOLOGY
in
INFORMATION AND COMMUNICATION TECHNOLOGY
to

DHIRUBHAI AMBANI INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGY



June, 2022

Declaration

I hereby declare that

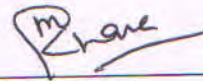
- i) the thesis comprises of my original work towards the degree of Master of Technology in Information and Communication Technology at Dhirubhai Ambani Institute of Information and Communication Technology and has not been submitted elsewhere for a degree,
- ii) due acknowledgment has been made in the text to all the reference material used.



Abhishek Shah

Certificate

This is to certify that the thesis work entitled "Person Re-identification in Surveillance Video" has been carried out by **Abhishek Shah (202011017)** for the degree of Master of Technology in Information and Communication Technology at *Dhirubhai Ambani Institute of Information and Communication Technology* under my supervision.



Dr. Manish Khare
Thesis Supervisor

Acknowledgments

The satisfaction that accompanies the successful partial completion of this thesis would be incomplete without mentioning those who made it possible. Furthermore, efforts would have been useless without their constant guidance and encouragement. I consider myself privileged to express gratitude and respect towards all those who have guided me through the thesis completion.

First, I would like to thank and show my sincere gratitude to my supervisor Dr. Manish Khare, for completing this thesis work. He was always there whenever I got stuck or needed any input. His Immense knowledge, professional expertise, and profound experience have enabled me to complete this research successfully. His idea of doing constructive weekly meetings helps a lot in completing the thesis in a given time. The feedback from the meetings surely helps me move forward in the right direction. Without his support and guidance, the completion of the thesis would not have been possible. I am honored to work under such brilliant guidance.

I also convey my sincere gratitude to my friends and family members for their constant support during my thesis work. I am grateful to the Dhirubhai Ambani Institute of Information and Technology(DAIICT) for giving me a fantastic opportunity to do the research work. Further, I want to thank DAIICT for providing constant hardware support, GPU support, and other helpful learning resources.

Contents

Abstract	v
List of Principal Symbols and Acronyms	vi
List of Tables	vii
List of Figures	viii
1 Introduction	1
1.1 Data Processing	2
1.2 General Taxonomy of Person Re-Identification	3
1.3 Motivation	4
1.4 Objective and Problem Statement	5
1.5 Organization of the Thesis	5
2 Literature Survey	6
2.1 Metric-learning based person Re-ID	6
2.2 Deep learning based person Re-ID	7
2.3 GAN-Based person Re-ID	8
2.4 Some Popular GAN-based Approaches	8
2.4.1 Pose Guided Person Generation Network	9
2.4.2 Feature Distilling Generative Adversarial Network	9
2.4.3 Pose Normalization Generative Adversarial Networks	10
3 Relevant Materials	12
3.1 Generative Adversarial Networks	12
3.1.1 Discriminator	13
3.1.2 Generator	13
3.1.3 Importance of GANs	13
3.2 Loss Function	14

4	Methodology	15
4.1	Discriminative and Generative Learning	15
4.1.1	Overview of DG-Net	15
4.1.2	Self-Identity Generation	16
4.1.3	Cross-Identity Generation	17
4.1.4	Major Architectural Components of DG-Net	17
4.2	Different Reconstruction Loss Functions	18
4.2.1	Mean Absolute Error(MAE)	18
4.2.2	Mean Squared Error(MSE)	18
4.2.3	Mean Rooted Absolute Error(MRAE)	19
4.2.4	Root Mean Square Error(RMSE)	19
4.2.5	Cosine Similarity(CS)	20
4.2.6	Multiplicative Loss Function(ML)	20
4.2.7	Huber Loss Function(Huber)	20
4.3	Proposed Method of Fusing Loss Functions	21
4.4	Implementation Details	22
5	Experimental Results and Analysis	23
5.1	Dataset Details	23
5.1.1	Market1501	23
5.1.2	DukeMTMC	24
5.2	Evaluation Metrics	25
5.2.1	Frechet Inception Distance	25
5.2.2	Mean Average Precision	25
5.3	Obtained Results using Market1501 Dataset	26
5.4	Obtained Results using DukeMTMC Dataset	31
5.5	Observations	32
6	Conclusion and Future Work	34
6.1	Conclusion	34
6.2	Future Work	34
	References	35

Abstract

The Person Re-identification (Re-ID) task has gained popularity in recent times. Researchers are continuously looking to improve the accuracy of the existing person Re-ID systems. Identifying the person from the surveillance footage can be an essential aspect of security concerns. Currently, there are many state-of-art person Re-ID systems available. Nowadays, Deep learning frameworks are adopted for designing Re-ID systems. Apart from deep learning-based approaches, the Generative Adversarial Networks (GAN) based approach also gained substantial interest in person Re-ID tasks. Augmentation of training data has significantly improved the performance of the system. Our primary objective is to analyze the effect of applying different reconstruction losses and their combinations on the GAN-based approach. The Discriminative and Generative Learning (DG-Net) based approach is chosen for carrying out this study from other existing GAN-based systems. DG-Net is currently considered benchmarked in the GAN-based method for person Re-ID. Analysis shows that the proposed idea of using a variety of reconstruction losses simultaneously significantly improves the existing system's performance. Using the proposed technique of fusing multiple Losses simultaneously, we achieved a massive performance gain of 20.57% over the current benchmarked approach on the Market1501 dataset. This report includes a thorough study of different loss functions and their effect on the generated images for performing person Re-ID task.

List of Principal Symbols and Acronyms

<i>CS</i>	Cosine Similarity
<i>DG – Net</i>	Discriminative and Generative Network
<i>FID</i>	Frechet Inception Distance
<i>GAN</i>	Generative Adversarial Network
<i>Huber</i>	Smooth L1 Loss
<i>MAE</i>	Mean Absolute Error
<i>mAP</i>	Mean Average Precision
<i>ML</i>	Multiplicative Loss
<i>MRAE</i>	Mean Root Absolute Error: Rooted MAE
<i>MSE</i>	Mean Square Error
<i>Re – ID</i>	Re-Identification
<i>RMSE</i>	Root Mean Square Error

List of Tables

5.1	Statistical Details of Market1501 Dataset	23
5.2	Statistical Details of Market1501 Dataset	24
5.3	Quantitative analysis of different loss functions	27
5.4	Quantitative analysis of the considered combinations of the loss Functions	27
5.5	Quantitative analysis of different loss functions for DukeMTMC Dataset	31

List of Figures

1.1	Example of different camera view points [1]	1
1.2	Steps involved in the Person Re-ID task [1]	2
1.3	Steps required to build a Person Re-ID system from scratch [1]	2
1.4	Various types of patch-wise data processing techniques to get the local representations of the different features [1]	3
1.5	General Taxonomy of the person Re-ID system [1]	4
2.1	General Approach for the metric-learning based Person Re-Identification [2]	7
2.2	Classification of different methods for Deep Learning based Person Re-Identification methods [3]	8
2.3	Architecture of use Guided Person Generation Network [4]	9
2.4	Architecture of Feature Distilling Generative Adversarial Network [5]	10
2.5	Architecture of Pose Normalization Generative Adversarial Networks [6]	11
3.1	Example Block Diagram of Generative Adversarial Network [7]	12
4.1	Architectural design of DG-Net [8]	16
5.1	Sample images taken from Market1501 dataset [9]	24
5.2	Sample images taken from DukeMTMC dataset [10]	24
5.3	Flowchart of how Mean Average Precision(mAP) is calculated during the testing	26
5.4	Side by side comparison of generated images using different loss functions. Value of the curly braces in the first row indicates the FID score for respective loss function.	28

5.5	Side by side comparison of generated images using different loss functions combined with the L1 Loss function. Value of the curly braces in the first row indicates the FID score for respective combination of the loss functions.	29
5.6	Effect of swapping Appearance and Structure codes of Images for different Loss Functions. The type of loss function is identified in curly braces in each section.	30
5.7	Effect of swapping Appearance and Structure codes of Images for considered combinations of Loss Functions. The type of loss function is identified in curly braces in each section.	31
5.8	Effect of swapping Appearance and Structure codes of Images for the DukeMTMC dataset. The type of loss function is identified in curly braces in each section.	32
5.9	Sample Comparison of generated image using L1 loss function (i.e., left image) and using with the combination of L1 and RMSE loss functions (i.e., right image)	33
5.10	Sample Comparison of generated image by swapping the appearance and structure codes using L1 loss function (i.e., output is shown in second column) and using with the combination of L1 and RMSE loss functions (i.e., output is shown in third column)	33

CHAPTER 1

Introduction

Person Re-identification (Re-ID) is crucial in a multi-camera environment, which is usual in surveillance videos. The main focus of these system types is to assign a stable id to the person appearing from the various non-overlapping camera views. The cameras from which these images have been taken often have significant intra-class variations due to the background, atmospheric changes, person's gestures or movements, different camera viewpoints, and many more causes. Examples of the various types of camera viewpoints are shown in Figure 1.1[1].



Figure 1.1: Example of different camera view points [1]

Person Re-ID involves three significant steps. The first important step is to detect the person from the frame captured by the surveillance camera. After detecting the person, it becomes essential to track the same person through all the frames. There is a need to establish a stable id for tracking each person. If a new person appears in the frame, the system must retrieve its id from the available database. All this is summarised in Figure 1.2.

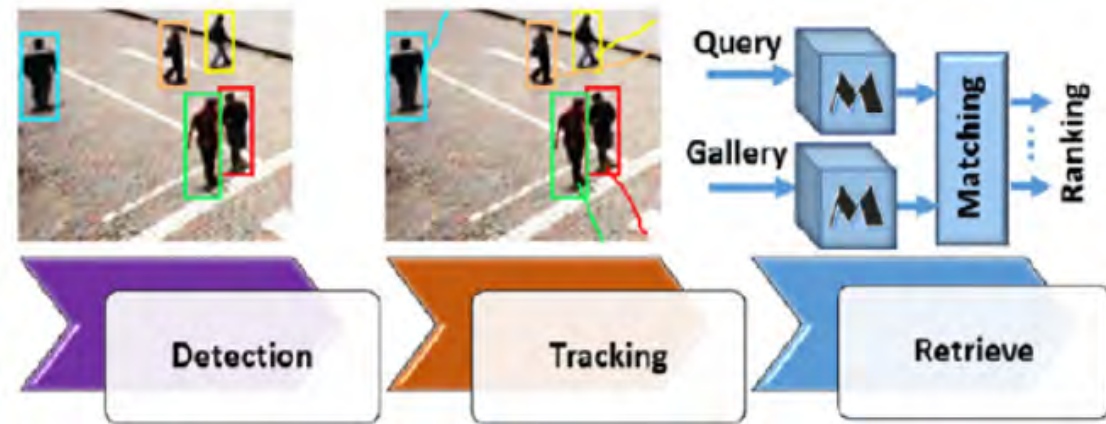


Figure 1.2: Steps involved in the Person Re-ID task [1]

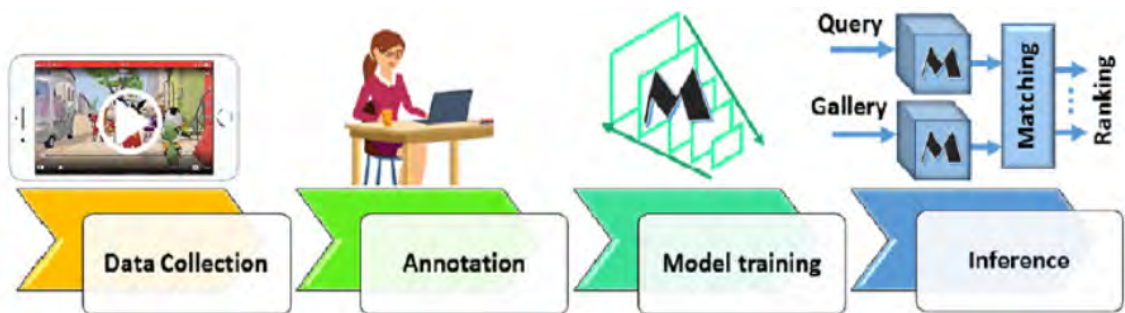


Figure 1.3: Steps required to build a Person Re-ID system from scratch [1]

If anyone wants to design a Person Re-ID system from scratch, the required steps are highlighted in Figure 1.3. Data collection becomes an important part when the system is designed from scratch. After completing annotating the data, any Person Re-ID methods can be applied.

1.1 Data Processing

After collecting and processing the dataset, the vital task is to extract the relevant features from the dataset. There are many ways to extract the features from a given person's image. Feature extraction can be performed in two ways. (i.e., Global-based processing and Patch-wise processing or local processing). Global-based processing methods focus on the topology of the cameras. In other words, the Actual location of the camera can play a significant role. For example, if there are two cameras at the entrance and the exit, it is evident that the person appears first at the entrance camera and then on the exit camera. Whereas, Patch-wise processing focuses on the minor details of the image. It can help discriminate the

intra-class samples. Some of the Patch-wise processing techniques are shown in Figure 1.4.



Figure 1.4: Various types of patch-wise data processing techniques to get the local representations of the different features [1]

1.2 General Taxonomy of Person Re-Identification

Person Re-ID systems have many points of view. The global idea of the overall taxonomy of the Person Re-ID systems is taken from [1]. Various aspects of Person Re-ID systems are beautifully summarized in Figure 1.5. If anyone wants to develop a new system or wants to start researching in the field of Person Re-ID domain, this taxonomy has proven essential. This taxonomy includes settings, learning types, approaches, context, query type, data modeling, and strategy. It gives a broader picture of Person re-identification in general. Also, it gives the researchers a new dimension of thinking for developing the person Re-ID models.

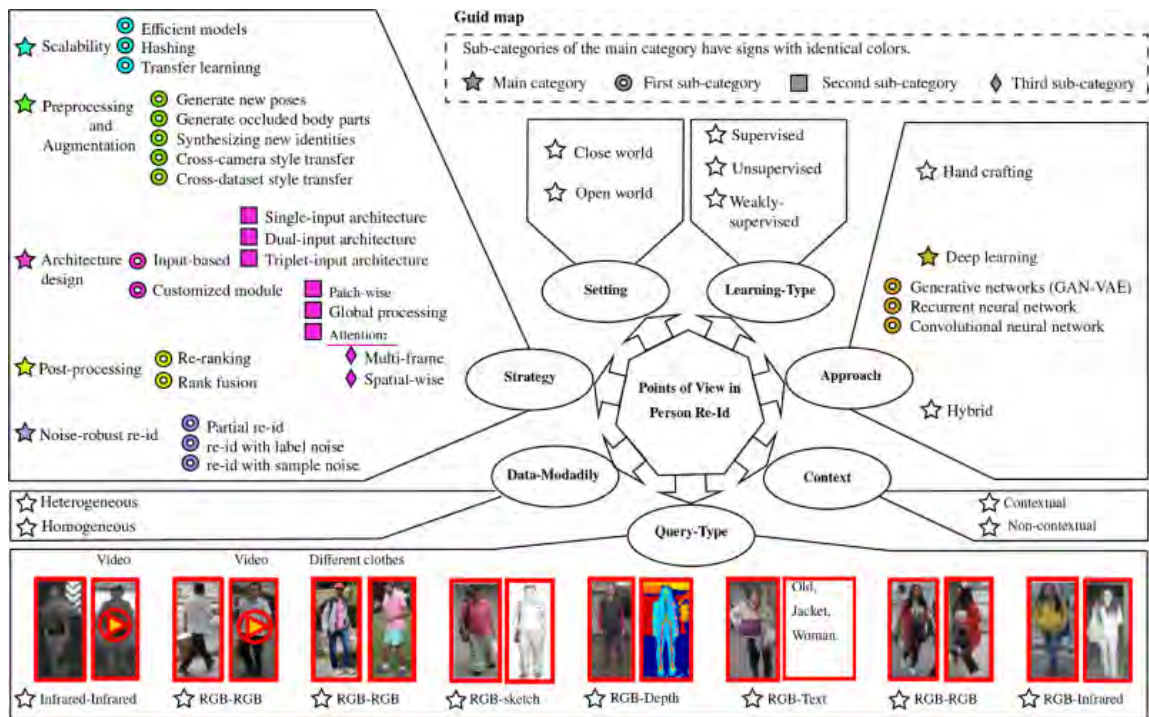


Figure 1.5: General Taxonomy of the person Re-ID system [1]

1.3 Motivation

Person Re-Identification becomes crucial when security is a concern. From the security perspective, only person detection and tracking are not sufficient. If the same person appears with different characteristics, these algorithms may fail to identify the person correctly. To deal with this issue, Person Re-Identification is essential. For Person Re-Identification, it becomes essential that the proposed approach is well generalized to handle different diversities caused by various characteristics of the person. There are many approaches available for doing the same. These days, Deep Learning based approaches are becoming popular.

The primary issue with training any Deep Learning based model is the availability of a massive amount of data. The model could not be generalized better if less data is provided. The data must have enough diversity. Hence, there is a need to augment the data. However, only data augmentation is not sufficient every time. To deal with this issue, Generative Adversarial Networks help create new images that try to mimic the characteristics of the original dataset. For Person Re-Identification, finding the dataset with enough diversity is challenging. Hence the GAN-based approaches become significantly crucial for performing the Person Re-Id tasks.

1.4 Objective and Problem Statement

The primary objective of this thesis/study is to enhance the performance of the person Re-Identification (Re-ID) task by exploring new techniques for the Generative Adversarial Networks(GAN) based approach. Many complexities are involved with the GAN-based approaches, such as dealing with different camera viewpoints, feature extraction, and other Re-ID-related features. It is often difficult when the GAN-based approaches are considered. To generalize the model for the GAN-based approaches in Person re-ID is an intricate task. The main task for any GAN-based approach is to generalize the training process from the available dataset and reconstruct the images with better quality that can be further used. It becomes essential how these images are processed. The inputs and putting constraints are the most vital task in any GAN-based approach.

Discriminative and Generative Network(DG-Net) [8] is currently a benchmarked GAN-based approach. This thesis work/study mainly focuses on improving the performance of DG-Net and improving the quality of the generated images.

1.5 Organization of the Thesis

The rest of the thesis is organized as follows. Chapter 2 contains a comprehensive literature survey in the person Re-ID domain. Chapter 3 gives insights into the technologies used in the method that is used in this thesis. Mainly, it gives the idea of Generative Adversarial Networks(GAN) and the Loss functions. In Chapter 4, the methodology is described in detail. It contains the details about Discriminative and Generative Networks and the proposed method. Chapter 5 discusses the obtained experimental results. Some of the critical observations are also highlighted in this chapter. In Chapter 6, the conclusion and the future directions are written. At the end of the thesis, all the references are included.

CHAPTER 2

Literature Survey

In this chapter, comprehensive idea of the research happening in the domain of person Re-ID are discussed. From 2008, the researchers showed considerable interest in developing personal Re-ID systems. Approximately person Re-ID systems are divided into two parts. Feature-based and the Metric-based. Feature-based approaches try to find an efficient representation of the person as a feature. In contrast, the primary focus of the Metric-based Approach is to develop an efficient metric to compute the similarity between two person's images[2]. The complete literature survey is divided based on the methods used to develop the person Re-ID systems. Some of the major works are cited in this paper. Recent works are mainly in the deep learning domain. Data augmentation using GANs and other deep learning-based hybrid approaches is famous for developing Re-ID systems.

2.1 Metric-learning based person Re-ID

This type of approach mainly deals with the similarity metric. In Figure 2.1, the idea of Metric-learning based approach is highlighted. In early 2000, The Mahalanobis distance [11] was introduced in the person Re-ID domain. It was used to measure the similarity in the Person Re-ID task. It tries to over-fit the model easily. To solve this issue of overfitting the model, the regularized independent metric is introduced in [12]. As the need for surveillance video increased, people put multiple cameras to make it more secure. Traditional metrics are not capable of dealing with these types of settings. In 2016, an asymmetric distance metric was introduced in [13] by Chen et al. Apart from cartesian systems, and An et al. introduced a metric based on hypergraph [14]. In [15], an improved version of the hypergraph method is introduced to deal with joint learning. Metric learning algorithms can be classified into majorly two types: classical metric learning algorithms and Deep-Learning based metric learning algorithms.

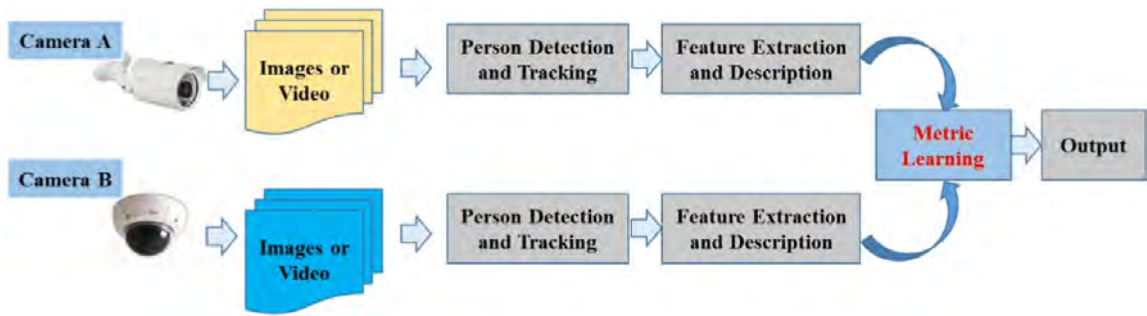


Figure 2.1: General Approach for the metric-learning based Person Re-Identification [2]

2.2 Deep learning based person Re-ID

Some of the very recent approaches to dealing with various aspects of Re-ID systems are highlighted in this section. These approaches are classified into different categories. These are well described in Figure 2.2. Also In [16], the limitations of stationary domain person Re-ID has been handled using a novel framework for knowledge representation named AKA (Adaptive Knowledge Accumulation). In unsupervised person Re-ID systems, intra-inter camera similarity computations are introduced to deal with the variations caused by the multiple cameras [17]. Fusing the inter-camera and intra-camera similarity has tremendously improved the performance of the person Re-ID system. Kecheng Zhang has introduced a grouping-based approach to improving the unsupervised person Re-ID [18]. It uses the idea of unsupervised domain adaptiveness. In other words, a system trained on some of the labeled domains can be applied to unlabeled domains without carrying out the annotations. Apart from these, many deep learning-based approaches are available for person re-identification frameworks. These can be found in one of the recent survey paper[19].

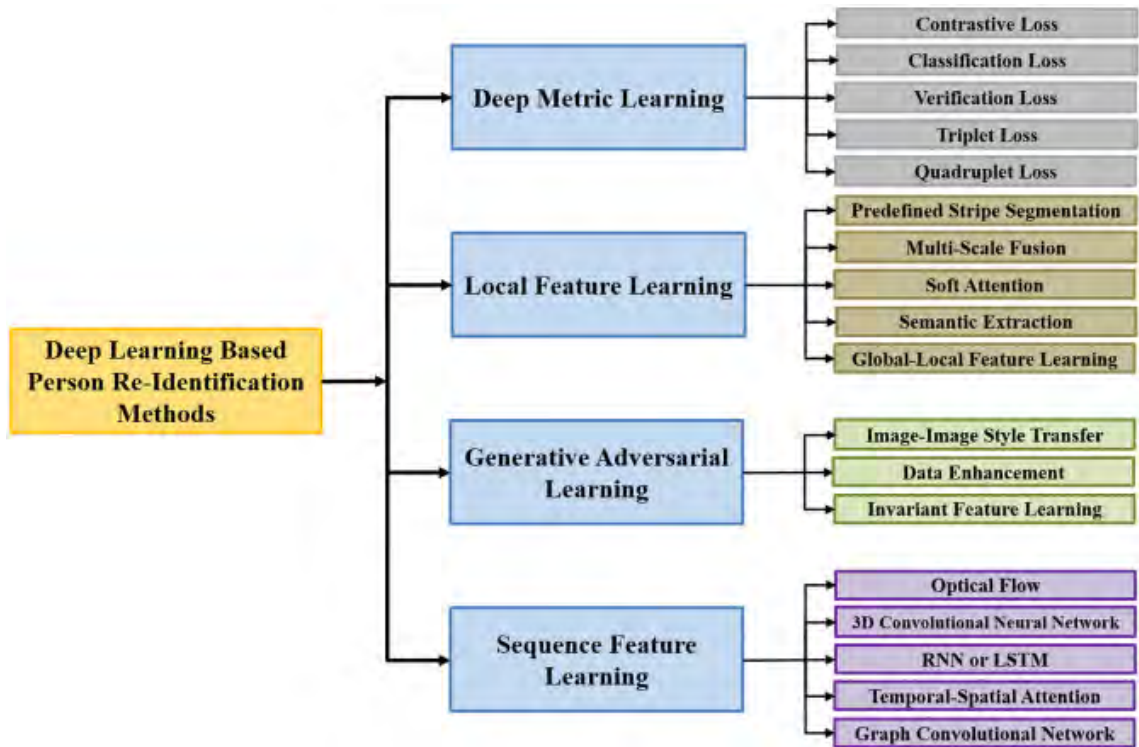


Figure 2.2: Classification of different methods for Deep Learning based Person Re-Identification methods [3]

2.3 GAN-Based person Re-ID

Nowadays, This Approach is a very active research area for improving the accuracies of the existing models. Zheng et al. was the first one who introduced the unconditional GANs to the person Re-ID.[20]. He used a deep convolutional generative adversarial network (DCGAN) to generate the samples. In [21], images are generated based on pose. Pose normalized GAN (PN-GAN) [6] has achieved significant accuracy in generating the images based on poses. Mostly the GAN-based approaches are unsupervised. Least Squares Generative Adversarial Networks (LSGANs) [22] is introduced to generalize the learning process well with the novel loss function (i.e., least-square loss function). After years of research, the latest DG-NET [8] has surpassed all the previous GAN-based approaches. More on DG-NET is discussed in the upcoming chapter.

2.4 Some Popular GAN-based Approaches

In this section, some of the popular GAN-based approaches are briefed. These were the revolutionary approaches in their respective times.

2.4.1 Pose Guided Person Generation Network

It is also considered one of the popular GAN-based approaches. They have introduced a concept of image generation based on the translation of poses. Pose Guided Person Generation Networks (PG^2) [4] mainly have two components in their architecture. The first component is for pose integration. It tries to translate the pose of one image to another and gives the intermediate output. This output is refined in the later section in an adversarial manner. The architectural details are given in Figure 2.2.

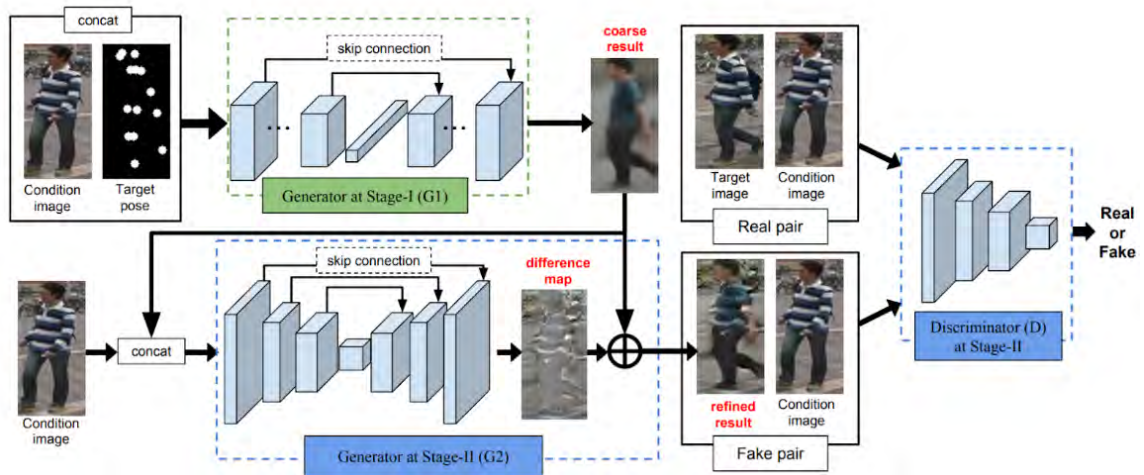


Figure 2.3: Architecture of Pose Guided Person Generation Network [4]

2.4.2 Feature Distilling Generative Adversarial Network

Before 2018, almost all the GAN-based approaches used additional pose-related information. It requires extra computations. To avoid this extra computation, in 2018, Yixiao Ge came up with the idea of using an identity-based approach. They introduced the Feature Distilling Generative Adversarial Network (FD-GAN) [5] in 2018. It has a siamese structure for tackling pose-related identifications. For better image generation, it uses the verification classifier. The architectural details of FD-GAN are shown in Figure 2.3.

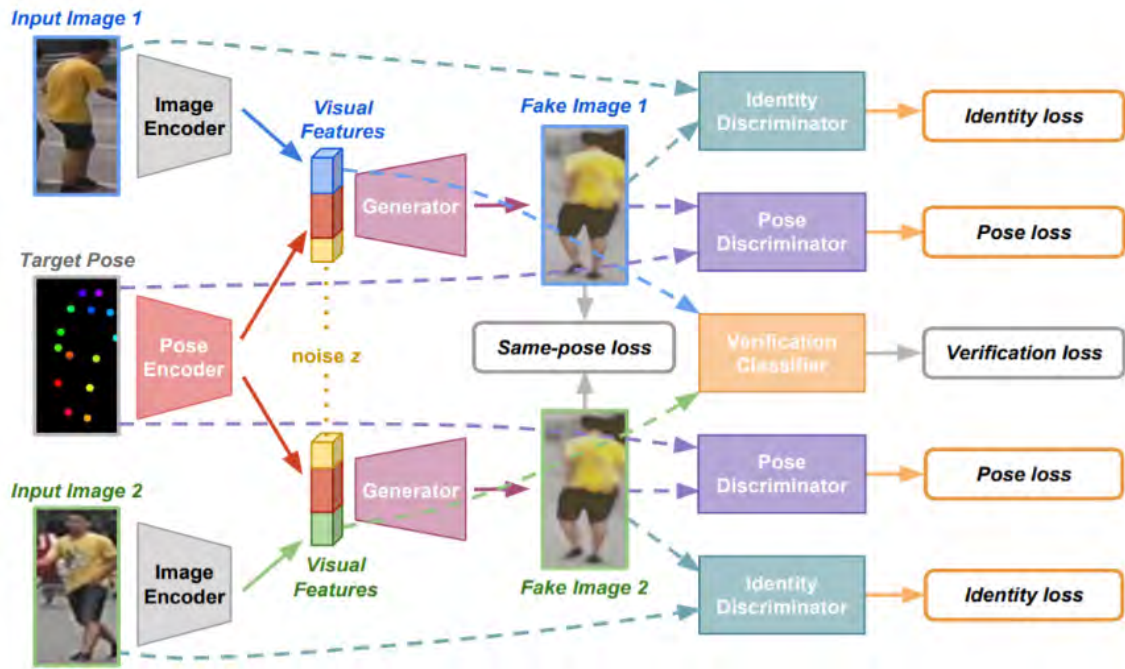


Figure 2.4: Architecture of Feature Distilling Generative Adversarial Network [5]

2.4.3 Pose Normalization Generative Adversarial Networks

Pose Normalization GAN(PNGAN)[6] has become popular as it tries to address two crucial aspects of feature extraction in person re-identification. (i.e., identity-sensitive related features and view-invariant features). They used the 8-canonical-based structure to represent the pose information. The structure of PNGAN is shown in Figure 2.4. In this, two networks are shown. The first network will focus on extracting features of different poses, and the other will focus on training those generated clusters of poses.

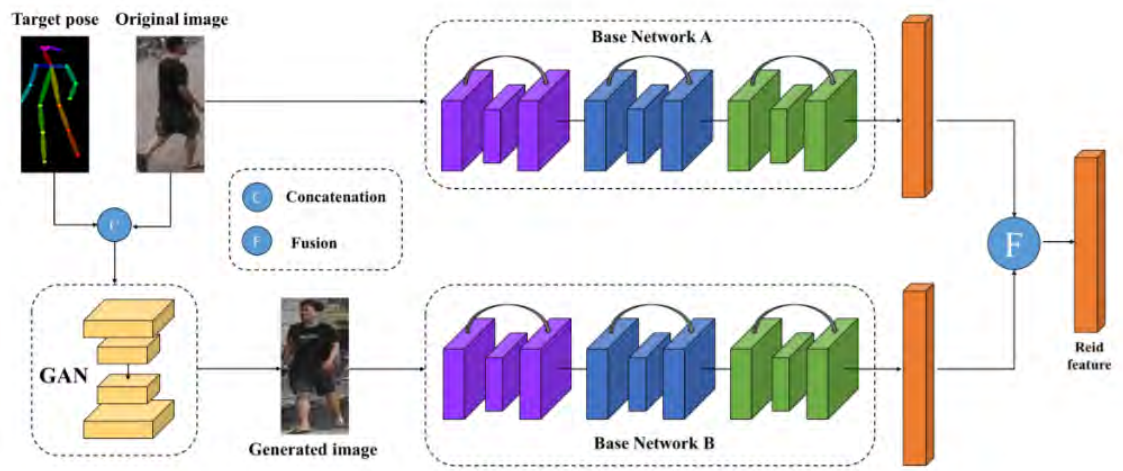


Figure 2.5: Architecture of Pose Normalization Generative Adversarial Networks [6]

CHAPTER 3

Relevant Materials

This chapter provides details about the essential technologies required for the proposed methodology. The Generative Adversarial Network(GAN) is introduced in this chapter. Also, the importance of the GANs is mentioned. Apart from this, loss functions are also discussed in brief. The primary usage of the loss function and its importance is also discussed.

3.1 Generative Adversarial Networks

Generative Adversarial Networks(GAN) are considered the most revolutionary addition to the neural network family. Thanks to Ian J. Goodfellow, who gave the world the idea about GANs in 2014. The goal of any GAN is to generate a new piece of content. The content can be images, audio, and video as well. GANs are now becoming popular choices for almost all the research domains due to their consistency. The significant usage of GANs is to generate more training samples to get more stability during the training. It causes the regularising effect for any neural network during training. It generally tries to map the probability distribution of the input with the generated output distribution. It takes sample distribution as an input and converts it into the target distribution.

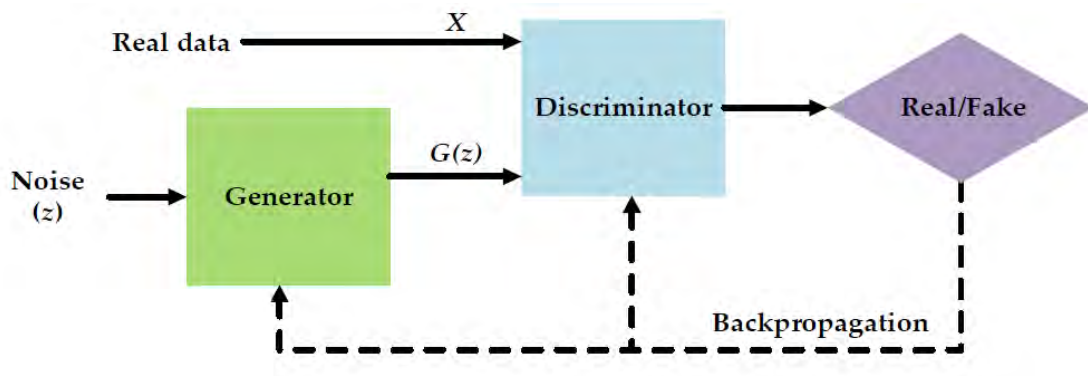


Figure 3.1: Example Block Diagram of Generative Adversarial Network [7]

In any GAN architecture, there are two primary components involved. The first is the generator, and the other is the discriminator. They work oppositely. Their goals are precisely the opposite. Training is done in an adversarial manner. They try to fool each other. The primary reason behind this adversarial training is to reduce the chance of overfitting the model. Figure 3.1 [7] describes the general idea of how these two components work together in an adversarial manner. More on the generator and discriminator are discussed in subsequent subsections.

3.1.1 Discriminator

The primary role of the discriminator is to tell whether the data is real or fake. The discriminator is trained on both actual and the generated images generated from the generator. Usually, the training of the discriminator starts when the generator is idle. In general, a significant focus of the GAN training is to ensure that the discriminator identifies the generated images (i.e., fake images) as the real ones. It can only happen when the generated images are close to the actual images. The training stops when the goal is achieved. In most cases, the discriminator is a binary classifier.

3.1.2 Generator

The primary role of the generator is to generate the images. The architecture of the generator varies from problem to problem. The training of the generator generally starts with the random noisy data. After several iterations, it tries to learn the distribution pattern from the input and generate the images close to the actual ones. The discriminator determines this. It updates its weights from the feedback got from the discriminator. Training converges when the generator can produce a better image matching the actual decision. In other words, training completes when the generator successfully fools the discriminator.

3.1.3 Importance of GANs

These days, data augmentation has become necessary in almost all cases. Nowadays, traditional data augmentation techniques are not sufficient because they try to augment the same actual data differently. Moreover, GAN tries to generate the new images, not in the actual dataset but to resemble the same. With the use of more data, the model can be generalized better. GANs are useful in reinforcement learning as well. Apart from these, GANs can be used to fill up the missing

data by generating them. GAN is used in a wide range of applications, such as Image-to-image translation, Text to image translation, Image inpainting, Image super-resolution, and Face Generation. With the constant improvements in the GAN architectures, the quality of the generated images is also improving. GAN is proven one of the most effective neural network family innovations.

3.2 Loss Function

Loss functions are an integral part of any neural network. It helps to quantify the error while predicting the output. Formally a loss function can be defined as a metric to find the cost between the actual and predicted values.[23] Broadly, the loss function can be categorized into two types (i.e., Classification loss and Regression loss). As the GAN tries to minimize the distance between the actual distribution and the generated distribution, in GANs, the loss functions quantify the desired distance. Loss functions can be linear as well as non-linear. Many non-linear loss functions exist, such as Root Mean Square Error, Cosine Similarity, cross-entropy loss, and many others. Some of them are used in this thesis work. They are discussed in detail in the Methodology section.

CHAPTER 4

Methodology

This chapter describes the methodology used for carrying out the study. This paper [8] is chosen as the baseline paper to carry out this research work. Apart from many GAN-based approaches, the Discriminative and Generative Networks(DG-Net) [8] based approach is benchmarked. It is trying to propose a novel framework that combines the data generation and Re-ID learning features. DG-Net and proposed scheme are discussed in the following subchapters.

4.1 Discriminative and Generative Learning

4.1.1 Overview of DG-Net

An effective GAN-based architecture named Discriminative and Generative Learning (DG-Net)[8] has been considered baseline. It has a unique generative module in which encoders decompose the input image (i.e., person's image) into two spaces (i.e., Appearance Space and The Structure Space). Clothing styles, shoe styles, textures, and other look-related features are considered in the appearance space. The size of the person's body, hair, pose, background, position, and camera angle/viewpoint is considered in the structure spaces. The design of DG-Net can produce high-resolution images and generate a diverse set of images. It does not require any additional data to produce the images. In other GAN-based approaches, the pose is given as an additional input to give the model more stability. However, DG-Net does not require any additional data to produce the images. It can handle intra-class variations and other pose-related diversities. It uses self-identity generation and cross-identity generation to generate images. Architectural details are given in Figure 4.1[8].

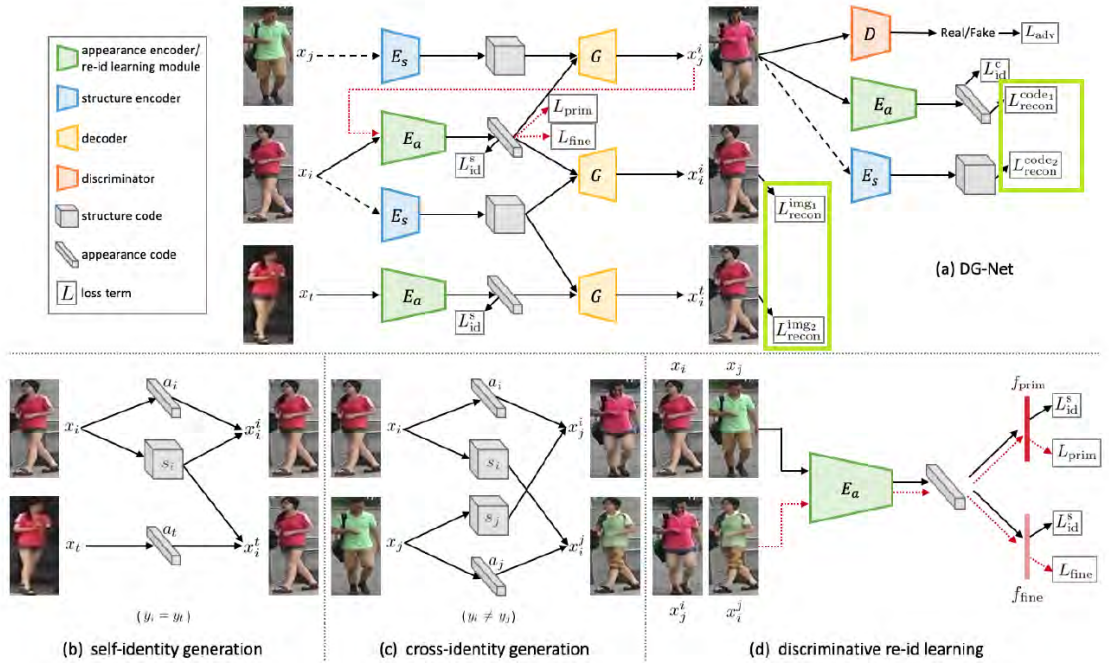


Figure 4.1: Architectural design of DG-Net [8]

4.1.2 Self-Identity Generation

In the Self-Identity generation, images are generated having the same identity. Apart from generating the same image from the training sample, there is also another image with the same identity given as an input to this module. Loss is calculated between generated image and the actual image. The equations for the same are given below.

$$\mathcal{L}_{recon}^{img1} = \mathbb{E} \left[\|x_i - G(a_i, s_i)\|_{LossType} \right] \quad (4.1)$$

$$\mathcal{L}_{recon}^{img2} = \mathbb{E} \left[\|x_i - G(a_t, s_i)\|_{LossType} \right] \quad (4.2)$$

Here,

$G \rightarrow$ Generator

$x_i \rightarrow$ Input Image

$a_i \rightarrow$ Appearance code of the input image x_i

$a_t \rightarrow$ Appearance code of some other image x_t which has the same identity as x_i

$s_i \rightarrow$ Structure code of the input image x_i

$LossType \rightarrow$ Type of the loss function

In equation 4.1, reconstruction from the same image is shown, while in equa-

tion 4.2, reconstruction from the other image with the same identity is shown.

4.1.3 Cross-Identity Generation

Cross-Identity generation focuses on generating images of two different identities. The crucial factor for generating the images from the two different identities is to identify the correct appearance code and the structure code from the respective input images. Due to this reason, the reconstruction constraints are put on the generated codes instead of the generated images. Equations for the same are shown below.

$$\mathcal{L}_{recon}^{code_1} = \mathbb{E} \left[\left\| a_i - E_a (G (a_i, s_j)) \right\|_{LossType} \right] \quad (4.3)$$

$$\mathcal{L}_{recon}^{code_2} = \mathbb{E} \left[\left\| s_j - E_s (G (a_i, s_j)) \right\|_{LossType} \right] \quad (4.4)$$

Here,

$G \rightarrow$ Generator

$E_a \rightarrow$ Appearance Encoder

$E_s \rightarrow$ Structure Encoder

$a_i \rightarrow$ Appearance code of the input image x_i

$s_j \rightarrow$ Structure code of the input image x_j

$LossType \rightarrow$ Type of the loss function

In the base paper, the L1 norm is used as a *LossType*. In this study, there are many combinations of losses covered. These are discussed in the following sub-chapters.

4.1.4 Major Architectural Components of DG-Net

The structure of DG-Net contains mainly four subblocks named structure encoder, decoder, discriminator, and the appearance encoder. These are also shown in Figure 4.1. Appearance Encoder is the slightly modified architecture of the ResNet50 [24] which is pre-trained on ImageNet [25]. In that output, the Adaptive Max Pooling layer is considered as an appearance code for the respective image. A structure encoder is a shallow convolutional neural network (CNN) based architecture. For the discriminator, the architecture is taken from the multi-scale PatchGAN [26]. Detailed architectural details of all these components are given in [8].

4.2 Different Reconstruction Loss Functions

Reconstruction loss can be defined as the loss between the actual values and their reconstructed values. For this study, for Self-Identity generation, reconstruction loss is used to optimize the generated images, whereas, in a cross-identity generation, it is used to optimize the generated appearance and the structure codes. All the considered loss functions are briefly discussed in the subsequent sections. For all the loss functions, instead of writing the above equations again, their representative equations are shown. In all the equations, x and y are used to denote the inputs. Inputs for the self-identity generation are $(x_i, G(a_i, s_i))$ and $(x_i, G(a_t, s_i))$. Whereas, inputs for the cross-identity generations are $(a_i, E_a(G(a_i, s_j)))$ and $(s_j, E_s(G(a_i, s_j)))$ for all the equations.

4.2.1 Mean Absolute Error(MAE)

It calculates the mean of the difference between the actual and obtained values. Sometimes this error function is referred to as the linear loss function or the L1 loss function. The representative equation for the MAE is shown in equation 4.5. The range of the MAE is $[0, \infty]$. Lower values are considered better.

$$\mathcal{L}_1(x, y) = (1/n) \left(\sum_{i=1}^{i=n} |x_i - y_i| \right) \quad (4.5)$$

Here,

$x \rightarrow$ Input data

$x_i \rightarrow i^{th}$ sample of the input data

$y \rightarrow$ Generated data

$y_i \rightarrow i^{th}$ sample of the generated data

$n \rightarrow$ Total number of pixels in the image

4.2.2 Mean Squared Error(MSE)

It calculates the mean of the squared of the actual and obtained values. This loss function is mainly referred to as the L2 Loss function. Equation 4.6 shows the formula used for MSE. The range of the MSE is $[0, \infty]$. Lower values are considered better.

$$\mathcal{L}_2(x, y) = (1/n) \left(\sum_{i=1}^{i=n} |x_i - y_i|^2 \right) \quad (4.6)$$

Here,

$x \rightarrow$ Input data

$x_i \rightarrow i^{th}$ sample of the input data

$y \rightarrow$ Generated data

$y_i \rightarrow i^{th}$ sample of the generated data

$n \rightarrow$ Total number of pixels in the image

4.2.3 Mean Rooted Absolute Error(MRAE)

It is the variant of the MAE. In other words, this can be defined as a normalized MAE. Instead of using absolute difference, the square root of that absolute difference is taken. Sometimes it is referred to as rooted MAE. It is defined in equation 4.7. The range of the MRAE is $[0, \infty]$. Lower values are considered better.

$$MRAE(x, y) = (1/n) \left(\sum_{i=1}^{i=n} \sqrt{|x_i - y_i|} \right) \quad (4.7)$$

Here,

$x \rightarrow$ Input data

$x_i \rightarrow i^{th}$ sample of the input data

$y \rightarrow$ Generated data

$y_i \rightarrow i^{th}$ sample of the generated data

$n \rightarrow$ Total number of pixels in the image

4.2.4 Root Mean Square Error(RMSE)

It is a popular choice among researchers for the loss function. The equation for the RMSE Loss function is shown in equation 4.8. The range of the RMSE is $[0, \infty]$. Lower values are considered better.

$$RMSE(x, y) = \sqrt{(1/n) \left(\sum_{i=1}^{i=n} \sqrt{|x_i - y_i|^2} \right)} \quad (4.8)$$

Here,

$x \rightarrow$ Input data

$x_i \rightarrow i^{th}$ sample of the input data

$y \rightarrow$ Generated data

$y_i \rightarrow i^{th}$ sample of the generated data

$n \rightarrow$ Total number of pixels in the image

4.2.5 Cosine Similarity(CS)

It calculates the similarity between two vectors. It calculates the cosine angle between two vectors. The representative equation is shown in equation 4.9. The range of the CS is $[0, 1]$. In this, higher value is considered better. The value is more towards 1, the images are more similar.

$$CS(x, y) = \frac{\left(\sum_{i=1}^{i=n} x_i y_i \right)}{\left(\sqrt{\sum_{i=1}^{i=n} x_i^2} \sqrt{\sum_{i=1}^{i=n} y_i^2} \right)} \quad (4.9)$$

Here,

$x \rightarrow$ Input data

$x_i \rightarrow i^{th}$ sample of the input data

$y \rightarrow$ Generated data

$y_i \rightarrow i^{th}$ sample of the generated data

$n \rightarrow$ Total number of pixels in the image

4.2.6 Multiplicative Loss Function(ML)

It is an experimental loss function used for this study. The definition for the multiplicative loss function is shown in equation 4.10. Due to multiplication, the values are becoming too big, and the results are not satisfactory. All these are discussed in the Results chapter. The range of the multiplicative loss is $[0, \infty]$. Lower values are considered better.

$$ML(x, y) = (1/n) \left(\sum_{i=1}^{i=n} |x_i * y_i|^2 \right) \quad (4.10)$$

Here,

$x \rightarrow$ Input data

$x_i \rightarrow i^{th}$ sample of the input data

$y \rightarrow$ Generated data

$y_i \rightarrow i^{th}$ sample of the generated data

$n \rightarrow$ Total number of pixels in the image

4.2.7 Huber Loss Function(Huber)

Huber Loss combines the Mean Absolute Error (MAE) and the Mean Square Error(MSE) with some criteria. It tries to balance both. The absolute difference is

used to decide between MAE and MSE. It is compared with the δ value. δ is the controlling parameter. Definition for the Huber loss is shown in equation 4.11. Huber loss approximates to MSE when $\delta \rightarrow 0$ and MAE when $\delta \rightarrow \infty$.

$$Huber(x, y) = \begin{cases} (1/n) \left((1/2) \sum_{i=1}^n |x_i * y_i|^2 \right) & \text{if } |x_i - y_i| \leq \delta, \\ (1/n) \left(\sum_{i=1}^n \delta (|x_i * y_i| - (1/2)\delta) \right) & \text{if } |x_i - y_i| > \delta, \end{cases} \quad (4.11)$$

Here,

$x \rightarrow$ Input data

$x_i \rightarrow$ i^{th} sample of the input data

$y \rightarrow$ Generated data

$y_i \rightarrow$ i^{th} sample of the generated data

$n \rightarrow$ Total number of pixels in the image

$\delta \rightarrow$ controlling factor (i.e., threshold)

4.3 Proposed Method of Fusing Loss Functions

It is analyzed that using the single loss function for reconstruction does not give better results. We performed several experiments considering multiple loss functions together. The base paper only uses the L1 loss function as a reconstruction loss function. The issue with the L1 loss function is that it cannot give enough non-linearity to the training process. As a result, many artifacts have occurred in the generated images. Hence, we proposed to use more than one reconstruction loss simultaneously.

Combinations are chosen in such a way that they have enough linear and non-linear components. To achieve this goal, we have combined all the non-linear loss functions with the linear loss function (i.e., MAE Loss function). The combinations we have considered for this study are listed below.

1. MAE(\mathcal{L}_1) + MSE(\mathcal{L}_2)
2. MAE(\mathcal{L}_1) + RMSE
3. MAE(\mathcal{L}_1) + MRAE
4. MAE(\mathcal{L}_1) + CS

Apart from the above combinations, we have also tried the combination of losses with the weights (i.e., $\lambda_1 * \text{linear_loss_function} + \lambda_2 * \text{non-linear_loss_function}$).

We chose the cosine similarity loss function for the non-linear loss to test this hypothesis, and 0.4 and 0.6 are chosen for λ_1 and λ_2 , respectively. The reason behind choosing these values is to give more importance to the non-linear loss function. Getting more importance can result into the better optimisation of that loss function. It was not able to surpass the baseline results. Hence for the study, only the above combinations are considered. Results are more than evident to claim that using more than one loss function has consistently performed better than the usual choice of selecting only L1 loss using the above fusing technique.

4.4 Implementation Details

The whole implementation is done using Python. PyTorch is used as the major framework to deal with creating the model, train the model and testing the model. For creating visual grids, PIL library is used. To calculate the FID score, the external FID library created by Zhedong Zheng is used as it is available publicly. To calculate the value of various loss functions, NumPy library is used. Also, Torch is used to create the tensors. For the training of the DG-Net, SGD [27] and Adam [28] optimiser are used at the various stages. The whole training details are described in the [8].

CHAPTER 5

Experimental Results and Analysis

In this section, details of all the experiments are mentioned. All the experiments are performed on the GeForce-RTX 2080 Ti GPU and Xenon-based CPU. At first, details of the dataset are included. After that, all the experiments are discussed in detail. In the end of this chapter, some of the critical observations are highlighted.

5.1 Dataset Details

5.1.1 Market1501

Market1501 dataset [9] is used to evaluate the performance of the system. This dataset is considered the benchmark dataset for evaluating the Person Re-ID Systems. It is available publicly, with one thousand five hundred-one identities captured from six different cameras. The whole dataset has two parts (i.e., training and testing). The distribution of the images among these two parts is mentioned in Table 5.1. This dataset is used to evaluate the proposed claim. The dataset contains a diverse set of images collected from multiple surveillance cameras. An example of the sample dataset is shown in Figure 5.1.

Table 5.1: Statistical Details of Market1501 Dataset

Market1501	Train	Query	Gallery
Number of IDs	751	750	751
Available number of images	12936	3368	15913



Figure 5.1: Sample images taken from Market1501 dataset [9]

5.1.2 DukeMTMC

Duke Multi-Tracking Multi-Camera Re-Identification (DukeMTMC-reid) [29] is another dataset used for this study for comparison purposes. This dataset is subset of the bigger dataset of DukeMTMC [30]. The images are taken from eight different cameras. And it has 702 identities available. Other statistical details are given in Table 5.2. This dataset does not have very diverse set of images.

Table 5.2: Statistical Details of Market1501 Dataset

DukeMTMC	Train	Query	Gallery
Number of IDs	702	702	702
Available number of images	16522	2228	17661



Figure 5.2: Sample images taken from DukeMTMC dataset [10]

5.2 Evaluation Metrics

5.2.1 Frechet Inception Distance

Frechet Inception Distance (FID) [31] score is mainly designed for Generative Adversarial networks. It quantifies how the generated dataset is close to the original data distribution. Sometimes FID score is referred as Wasserstein-2 distance [32]. A lower FID score indicates better image quality. Contraversouly, a higher FID score indicates that generated images are not close to the original images. FID score is currently the widely chosen evaluation metric to evaluate the quality of synthesized images.

5.2.2 Mean Average Precision

FID score is used to evaluate the quality of the generated images. However, Mean Average Precision (mAP) will tell how well the model is generalized on that dataset. Mathematically the mAP can be defined as follows.

$$mAP = \frac{1}{N} \sum_{i=1}^{i=N} AP_i \quad (5.1)$$

Here,

$N \rightarrow$ Total number of queries

$AP_i \rightarrow$ Average precision of the i^{th} query

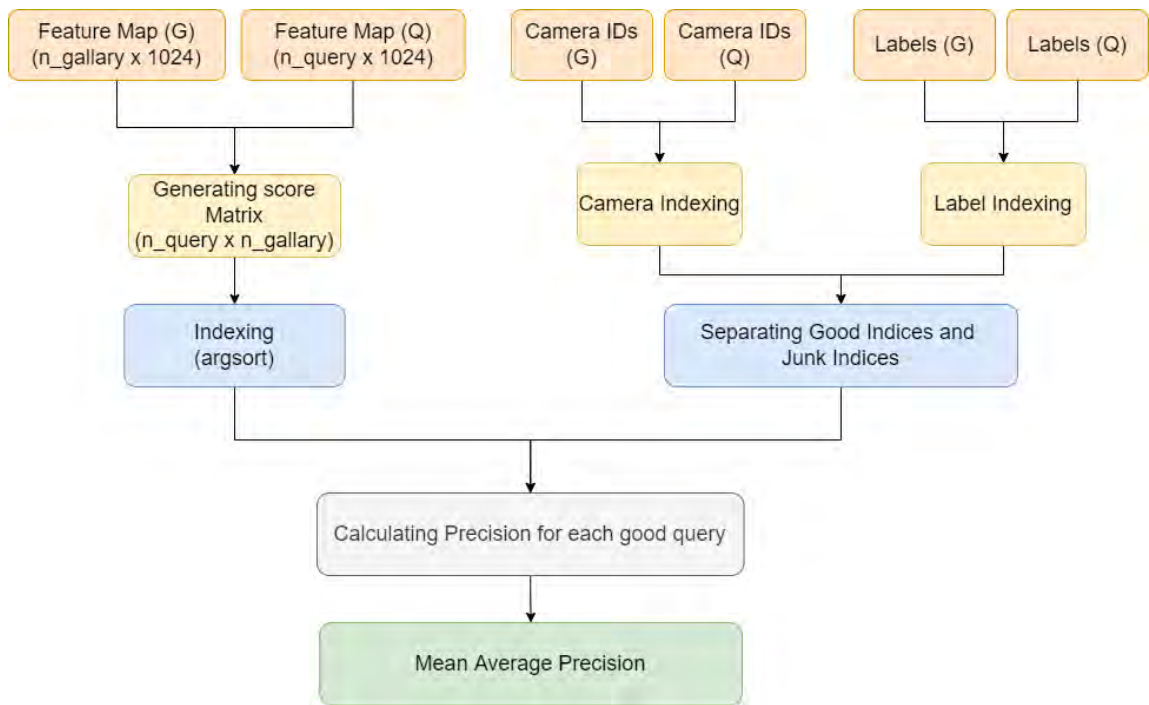


Figure 5.3: Flowchart of how Mean Average Precision(mAP) is calculated during the testing

In all the experiments, the procedure for calculating mAP values is elaborated in Figure 5.3. Feature map, camera ids, and the corresponding labels are taken from the gallery and query. Camera ids and labels help find good indices. Those ids will be considered in the good indices that are taken from different cameras and presented in the query index. Rest are considered junk indices. After performing Indexing on the good indices, average precision is calculated for each query. At last, the mean of all of them will give the final mAP value.

5.3 Obtained Results using Market1501 Dataset

The experiments are performed on the GeForce-RTX 2080 Ti GPU and Xenon-based CPU. All the Quantitative results with the individual loss function are shown in Table 5.2. Results with the combinations of loss functions are shown in Table 5.3. It is clear from both the table that if we use a combination of linear and non-linear loss functions simultaneously, the system’s performance improves. Also, the FID score consistently increases, meaning the generated image quality increases. In Table 5.2, the first row contains the results for the initially used method (i.e., using L1 Loss only). The performance gain is visible if any non-linear loss function is combined with the linear loss function (i.e., L1 Loss

function). The maximum performance gain is achieved when the RMSE loss function is combined with the L1 loss function, which is 20.57 % in the FID score and 1.19 % in the mAP value.

Table 5.3: Quantitative analysis of different loss functions

Reconstruction Loss	FID Score	Rank@1 (%)	Rank@5(%)	Rank@10(%)	mAP(%)
L1	19.68	91.09	96.73	97.80	75.96
Smooth L1(Huber)	25.17	90.65	96.73	98.01	75.83
MSE(L2)	21.40	90.29	96.88	98.31	74.75
MRAE	21.73	91.18	96.91	98.16	75.73
Cos	37.95	90.11	96.35	97.62	74.15
RMSE	19.65	91.33	96.79	98.01	75.19
ML	319.82	68.35	84.47	89.07	43.77

Table 5.4: Quantitative analysis of the considered combinations of the loss Functions

Reconstruction Loss	FID Score	Rank@1 (%)	Rank@5(%)	Rank@10(%)	mAP(%)
0.4*L1 + 0.6*Cos	20.55	90.32	96.70	97.74	74.81
L1 + MSE(L2)	17.99	92.37	97.45	98.56	78.63
L1 + MRAE	18.79	91.03	96.82	97.98	77.13
L1 + RMSE	15.63	91.42	97.06	98.16	76.87
L1 + Cos	17.35	91.45	96.44	97.95	76.92

Figure 5.4 visually shows how the generated images differ when the different loss functions are used. At the top, the FID score is mentioned in the curly braces. It is observed that when only non-linear loss functions are used (i.e., apart from L1 loss), the results are not that satisfying. The visual artifacts in the generated images are more in non-linear loss functions than a linear loss function. Also, when the multiplicative loss is used, the model could not generate the images well, which are available in the last column of Figure 5.4. Compared with only single losses, Figure 5.5 compares the combinations of various non-linear loss functions with the linear loss function(i.e., the L1 loss function). The results clearly say that combinations of loss functions can generate more realistic images than the single loss function. One thing is also noticeable here that it may happen that FID score is better but quality of the image is not better. It is because FID score is calculated based on the entire dataset. Some images can have degraded quality. Some of them can be spotted in the Figure 5.5 as well. (i.e., In the case of L1 + cos, FID score is better but image quality of third image is better in case of L1).

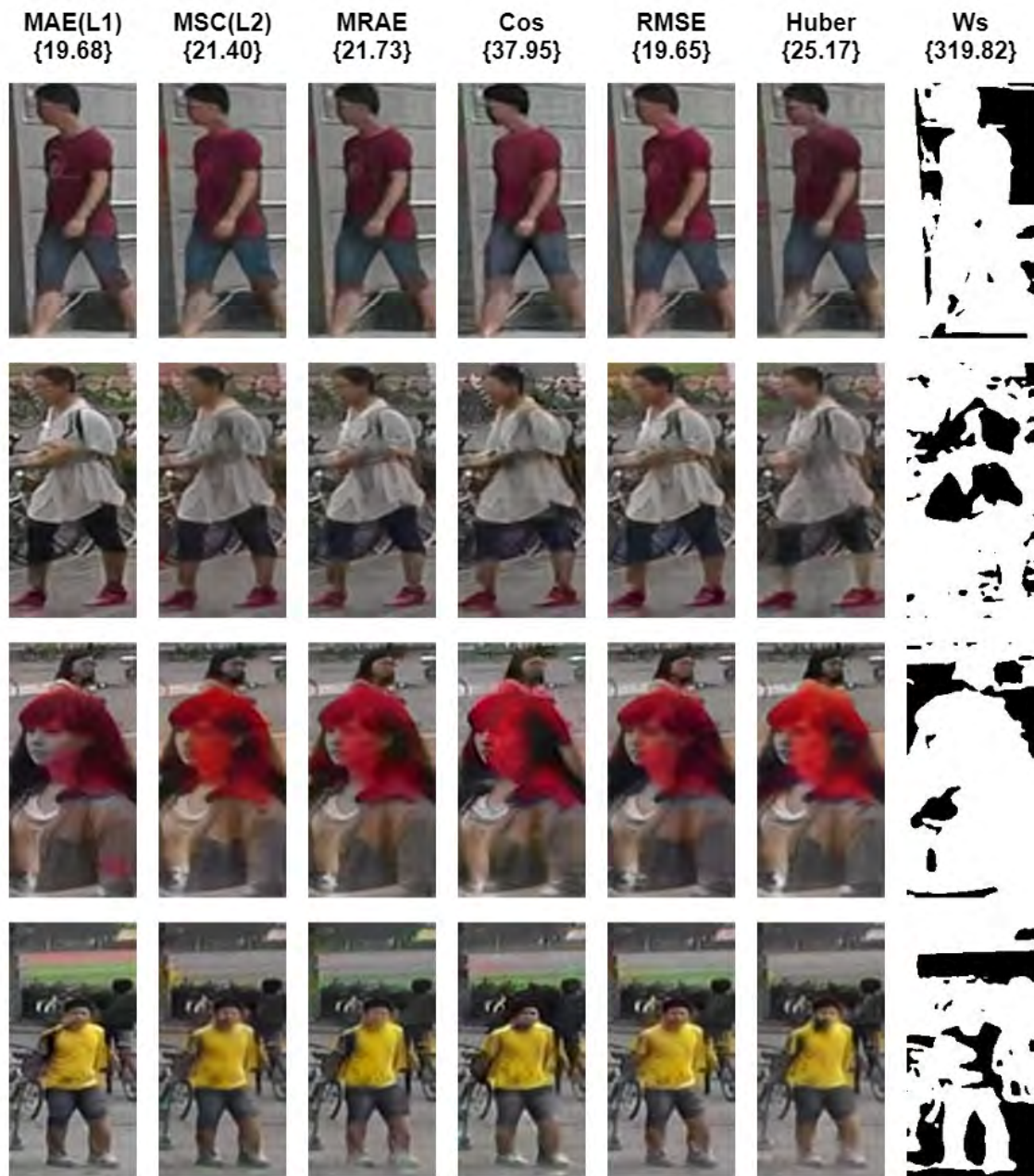


Figure 5.4: Side by side comparison of generated images using different loss functions. Value of the curly braces in the first row indicates the FID score for respective loss function.



Figure 5.5: Side by side comparison of generated images using different loss functions combined with the L1 Loss function. Value of the curly braces in the first row indicates the FID score for respective combination of the loss functions.

As discussed in the methodology section, the person's appearance and structure are considered for generating the images. Images are generated using one

person's structure data combined with the other person's appearance data and vice versa. Figure 5.6 shows the effect of swapping appearance and structure data by using different loss functions. In Figure 5.6, The top left images are the original input images, and the rest are generated using respective loss functions (i.e., Mentioned in the curly braces). In contrast, Figure 5.7 compares the combinations of the non-linear loss function with the linear loss function. In both the figures, the first two rows show the generated images for each section of images by swapping respective appearance and structure data (except for the top-left section).



Figure 5.6: Effect of swapping Appearance and Structure codes of Images for different Loss Functions. The type of loss function is identified in curly braces in each section.



Figure 5.7: Effect of swapping Appearance and Structure codes of Images for considered combinations of Loss Functions. The type of loss function is identified in curly braces in each section.

5.4 Obtained Results using DukeMTMC Dataset

The same set of experiments are carried out for the DukeMTMC dataset on the same hardware. The results obtained are not up to the mark. For the DukeMTMC dataset, a fusion of losses is not working well. It could not be able to regenerate the colors well. It is evident that by looking at Figure 5.8, non-linear loss or any combination of linear and nonlinear loss cannot be able to produce better results for DukeMTMC dataset. Quantitative results for the same are mentioned in Table 5.5. This can be caused because of the limitation caused by the dataset itself.

Table 5.5: Quantitative analysis of different loss functions for DukeMTMC Dataset

Reconstruction Loss	FID Score	Rank@1 (%)	Rank@5(%)	Rank@10(%)	mAP(%)
L1	20.84	81.19	90.71	93.18	67.33
L1 + L2	97.17	10.80	22.13	22.01	23.71
L1 + MRAE	86.40	14.37	23.71	26.31	26.34



Figure 5.8: Effect of swapping Appearance and Structure codes of Images for the DukeMTMC dataset. The type of loss function is identified in curly braces in each section.

5.5 Observations

It is hard to see the difference in the two images' quality from the figures shown in section 5.3. In Figure 5.9 and Figure 5.10, the sample output is enlarged. It gives a better idea of how well the technique of combining loss functions produces images close to the actual images. We can see the clear difference in Figure 5.9 and Figure 5.10. In Figure 5.9, it is shown that the shoulder part is more detailed when the loss functions are combined. In Figure 5.10, it is observed that using a combination of loss functions can generate a more detailed image. A clear difference is shown in the hair part and the shoe part. These observations are only valid for the Market1501 dataset.



Figure 5.9: Sample Comparison of generated image using L1 loss function (i.e., left image) and using with the combination of L1 and RMSE loss functions (i.e., right image)



Figure 5.10: Sample Comparison of generated image by swapping the appearance and structure codes using L1 loss function (i.e., output is shown in second column) and using with the combination of L1 and RMSE loss functions (i.e., output is shown in third column)

CHAPTER 6

Conclusion and Future Work

6.1 Conclusion

The primary target of this study was to improve the quality of the person Re-Identification dataset and try to enhance the performance of the person Re-ID systems using those datasets. Results showed that the proposed scheme of combining non-linear losses with the linear loss massively increases the performance of the existing GAN-based approach for the Market1501 dataset. As DukeMTMC dataset contains many false identities and also the dataset has not enough diversity, the same approach is not worked well on the DukeMTMC dataset. Results are evident enough to say that the proposed technique is consistent with almost all the non-linear losses. Using the proposed method to the existing GAN-based approach, a considerable performance gain of 20.57 % AND 3.51 % was achieved on the FID score and the mAP, respectively, on the Market1501 dataset.

6.2 Future Work

There is a scope for improvement in finding better combinations of loss functions or the individual loss function for the DukeMTMC dataset. Further, this study can be extended by adding more loss functions. Also, there is further scope to improve the quality of generated images by accomplishing necessary architectural changes.

References

- [1] Ehsan Yaghoubi, Aruna Kumar, and Hugo Proença. Sss-pr: A short survey of surveys in person re-identification. *Pattern Recognition Letters*, 143:50–57, 2021.
- [2] Guofeng Zou, Guixia Fu, Xiang Peng, Yue Liu, Mingliang Gao, and Zheng Liu. Person re-identification based on metric learning: a survey. *Multimedia Tools and Applications*, 80(17):26855–26888, Jul 2021.
- [3] Hugo Oliveira, José Machado, and Joao Tavares. Re-identification in urban scenarios: A review of tools and methods. *Applied Sciences*, 11:10809, 11 2021.
- [4] Liqian Ma, Xu Jia, Qianru Sun, Bernt Schiele, Tinne Tuytelaars, and Luc Van Gool. Pose guided person image generation. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- [5] Yixiao Ge, Zhuowan Li, Haiyu Zhao, Guojun Yin, Shuai Yi, Xiaogang Wang, and Hongsheng Li. Fd-gan: Pose-guided feature distilling gan for robust person re-identification, 10 2018.
- [6] Xuelin Qian, Yanwei Fu, Tao Xiang, Wenxuan Wang, Jie Qiu, Yang Wu, Yuguang Jiang, and Xiangyang Xue. Pose-normalized image generation for person re-identification, 2018.
- [7] Jie Feng, Xueliang Feng, Jiantong Chen, Xianghai Cao, Xiangrong Zhang, Licheng Jiao, and Tao Yu. Generative adversarial networks based on collaborative learning and attention mechanism for hyperspectral image classification. *Remote Sensing*, 12:1149, 04 2020.
- [8] Zhedong Zheng, Xiaodong Yang, Zhiding Yu, Liang Zheng, Yi Yang, and Jan Kautz. Joint discriminative and generative learning for person re-identification, 2021.

- [9] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1116–1124, 2015.
- [10] Weijian Deng, Liang Zheng, Guoliang Kang, Yezhou Yang, Qixiang Ye, and Jianbin Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 994–1003, 2018.
- [11] R. De Maesschalck, D. Jouan-Rimbaud, and D.L. Massart. The mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, 50(1):1–18, 2000.
- [12] Qi Meibin, Wang Yunxia, and et.al. Tan Shengshun. The mahalanobis distance. *Pattern Recognition and Artificial Intelligence*, 29(6):511–518, 2016.
- [13] Ying-Cong Chen, Wei-Shi Zheng, Jian-Huang Lai, and Pong C. Yuen. An asymmetric distance model for cross-view feature mapping in person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(8):1661–1675, 2017.
- [14] Le An, Xiaojing Chen, and Songfan Yang. Person re-identification via hypergraph-based matching. *Neurocomputing*, 182:247–254, 2016.
- [15] Xibin Zhao, Nan Wang, Yubo Zhang, Shaoyi Du, Yue Gao, and Jianguang Sun. Beyond pairwise matching: Person reidentification via high-order relevance learning. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8):3701–3714, 2018.
- [16] Nan Pu, Wei Chen, Yu Liu, Erwin M. Bakker, and Michael S. Lew. Lifelong person re-identification via adaptive knowledge accumulation, 2021.
- [17] Shiyu Xuan and Shiliang Zhang. Intra-inter camera similarity for unsupervised person re-identification, 2021.
- [18] Kecheng Zheng, Wu Liu, Lingxiao He, Tao Mei, Jiebo Luo, and Zheng-Jun Zha. Group-aware label transfer for domain adaptive person re-identification, 2021.
- [19] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven C. H. Hoi. Deep learning for person re-identification: A survey and outlook, 2021.
- [20] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro, 2017.

- [21] Liqian Ma, Xu Jia, Qianru Sun, Bernt Schiele, Tinne Tuytelaars, and Luc Van Gool. Pose guided person image generation. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- [22] Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks, 2017.
- [23] Christophe Pere. What are loss functions? <https://towardsdatascience.com/what-is-loss-function-1e2605aeb904>, May 2022.
- [24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. 7, 12 2015.
- [25] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.
- [26] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, 2017.
- [27] Wenbin Jiang, Geyan Ye, Laurence T. Yang, Jian Zhu, Yang Ma, Xia Xie, and Hai Jin. A novel stochastic gradient descent algorithm based on grouping over heterogeneous cluster systems for distributed deep learning. In *2019 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, pages 391–398, 2019.
- [28] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.
- [29] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *Proceedings of the IEEE International Conference on Computer Vision*, 2017.
- [30] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *European Conference on Computer Vision workshop on Benchmarking Multi-Target Tracking*, 2016.

- [31] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium, 2018.
- [32] Jason Brownlee. How to implement the frechet inception distance (fid) for evaluating gans, Oct 2019.