International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

Person Re-Identification using Centroid and Quadruple Loss based Deep Learning Model

Nikhil Kumar Singh[1], Manish Khare[*2], Harikrishna Babaubhai Jethva[3]

Submitted:14/03/2024 **Revised**: 29/04/2024 **Accepted**: 06/05/2024

Abstract: Person re-identification (Re-ID) stands as a crucial objective within expansive video surveillance landscapes, concentrating on the recognition of individuals across diverse camera sources. Lately, the utilization of deep learning networks in conjunction with the triplet loss has emerged as a prevalent framework for advancing person Re-ID endeavours. In this paper we have discussed various loss functions of Deep Learning and their application in person-re-identification. Utilizing the mean centroid representation offers heightened resilience to outliers and guarantees a more dependable set of features. This arises from the fact that every class is symbolized by a sole embedding, denoted as the class centroid. This approach not only substantially curtails retrieval time and storage needs but also enhances stability. This paper proposes a novel Centroid Quadruple Loss based approach. Through comprehensive experiments, it becomes evident that the proposed approach yields substantial improvements in results compared to the centroid triplet loss based approach as well as other recent state-of-the-art person re-identification methods.

Keywords: *Loss function, Deep learning, Centroid Loss, person re-identification, Quadruple loss*

1. Introduction

Person re-identification, often abbreviated as person re-ID or re-ID, is a computer vision task that aims to identify the same indivi dual across multiple non-overlapping camera views in a surveillance network. This technology is particularly useful for applications such as video surveillance, forensic analysis, and social behaviour understanding. In recent times, cameras have become ubiquitous across diverse domains, spanning from video games and home surveillance applications to extensive camera networks in locations such as sports venues, airports, metro stations, and car parks. In many surveillance systems, the primary challenge lies in detecting and tracking moving objects. Nevertheless, the overarching objective of any surveillance system goes beyond mere tracking and reacquisition of targets; it is to comprehend a scene and establish the identity of the desired object. The primary goal of person re-ID is to correctly associate images of the same person taken from different angles, viewpoints, or time instances, even when the images may exhibit variations in lighting, pose, clothing, and occlusion. The main challenge in person re-identification arises from variations in lighting, pose, camera viewpoint, occlusion, and changes in clothing. Due to these challenges, traditional person re-ID methods struggled to achieve accurate results. Deep learning methods have the capability to uncover relationships, behaviours, and models directly from datasets. Artificial Intelligence, Machine Learning, Neural Network and Deep learning based techniques are being used in numerous research fields to achieve state of the art results [1][2][3]. The effectiveness of deep learning in image classification [4] extended to the field of Person Re-Identification (Re-ID) in 2014 when both Yi et al. [5] and Li et al. [6]utilized siamese neural networks to ascertain whether a pair of input images belongs to the same identity. However, advancements in deep learning and computer vision have led to significant improvements in person re-ID techniques. The loss function is considered as most useful parameter in Deep Neural Networks. Many loss functions are designed to make the right choice but selecting the right loss function with due justification is rarely available in the literature.

The motivation behind our proposed approach is to take the benefit of Centroid Loss[7] and Quadruple Loss[8]. In Centroid Loss, Distances are measured between a sample and a prototype/centroid representing a class. A centroid approach results in one embedding per item, decreasing both memory and storage requirement. The advantage of using it is -

- Lower Computational Cost
- Higher Robustness to Outliers and noisy labels
- Faster training
- Comparable or better performance than standard loss

Quadruplet loss covers the weaknesses from both the binary classification loss and the triplet loss[9] to some extent, and

[1] Research Scholar, Gujarat Technological University, Ahmedabad & Assistant Professor, GEC Gandhinagar, Gujarat, India
ORCID ID: 0000-0002-8663-1569
[2] Assistant Professor, DA-IICT Gandhinagar, Gujarat, India
ORCID ID: 0000-0002-2296-2732
[3] Associate Professor, GEC Patan, Gujarat, India.
ORCID ID: 0000-0003-2954-117X
** Corresponding Author Email: mkharejk@gmail.com*

takes their advantages in person Re-ID which achieves a better performance than either of them.

The subsequent sections of this paper are organized to provide a structured and comprehensive exploration of the research. "Loss Functions in Deep Learning" offers an encompassing overview of loss functions tailored for both regression and classification models. The subsequent segment, "Loss Functions in Person Re-Identification," meticulously delineates the commonly employed loss functions within the domain of person re-identification. Moving forward, the "Literature Review" section unfolds as a synthesis of person re-identification approaches, providing insight into the varied methodologies that leverage diverse loss functions. The section titled "Datasets" systematically details various datasets, focusing particularly on Market 1501 and DukeMTMC ReID, offering a contextual understanding of the data employed. The "Proposed Approach" section intricately articulates the architectural framework and techniques embraced in the proposed work. Following this, the "Implementation and Result" section delves into the experimental methodology and reports findings obtained on the Market 1501 and DukeMTMC ReID datasets. "Future Scope" navigates towards uncharted territories, suggesting potential directions for further research within the specified field. Lastly, the "Conclusion" section serves as a cohesive synthesis, summarizing key findings and insights gleaned from the research journey. This structured organization ensures a seamless flow and comprehensive coverage of the research landscape.

2. Loss Functions in Deep Learning

The role of a loss function lies in the training of Deep Learning Models. The performance of a statistical model is solely assessed by its accuracy in decision-making. This necessitates a method to gauge the disparity between a specific iteration of the model and the actual values. This is where the significance of loss functions becomes evident. These functions gauge the divergence between an estimated value and its actual value. Essentially, a loss function establishes a correspondence between decisions and their respective costs. These functions are dynamic and adapt to the task at hand and the overarching objective.

2.1 Loss functions for Regression

Regression pertains to the anticipation of a continuous and specific value. For given input the regression model predicts the best corresponding output. Activities such as forecasting house prices or predicting stock values fall under the realm of regression, as they entail constructing models to foresee quantities that are represented by real numbers. Table 1 contains Loss functions for Regression.

2.2 Loss functions for Classification

Classification tasks encompass forecasting a discrete class output. This entails segmenting the dataset into distinct classes according to various criteria, enabling the assignment of a novel, unseen data point to one of these classes. For instance, emails can be categorized as either spam or non-spam, while a person's dietary choices can be classified as vegetarian, non-vegetarian, or vegan. Table 2 contains Loss functions for Classification.

Table 1: Loss functions for Regression[10][11]

Regression Loss	Remarks	Regression Function
Mean Squared Error (MSE) Loss	• Predicted values above or below the Observed value do not have any impact • Easily setup weight value due to its convex nature and gradient descent optimization. • Highly sensitive to Outliers i.e. if the predicted value is significantly large or small , loss will be increased	MSE= $\frac{1}{n}\sum_{i=1}^{n}(Y_i - Y'_i)^2$
Mean Squared Logarithmic Error Loss (MSLE)	Calculation of the mean squared logarithmic error follows the same procedure as the mean squared error, but differs by using the natural logarithm of the predicted values instead of the actual values.	MSE= $\frac{1}{n}\sum_{i=1}^{n}(log(Y_i) - log(Y'_i))^2$
Mean Absolute Error (MAE)	• Used when training data contains huge no of Outliers.	MAE= $\frac{\sum_{i=1}^{n}\lvert y_i - x_i\rvert}{n}$

	• If the average distance approaches to 0, gradient descent optimization will not work and will give error.	
Mean Bias Error (MBE)	Mean Bias Error considers the actual difference between the target and the predicted value, without taking the absolute value of the difference into account.	MBE= $\frac{\sum_{i=1}^{n} y_i - y'_i}{n}$

Table 2: Loss functions for Classification[10][11]

Classification Loss	Remarks
Binary Cross-Entropy	It is used for classification problems having two classes. Entropy quantifies the level of unpredictability within the processed information, while cross entropy gauges the disparity in unpredictability between two distinct random variables.
CE Loss $= -\frac{1}{N}\sum_{i=1}^{N} y_i$log (P_i)) +(1- y_i) log(1-log (P_i) Where y_i is the Actual label and P_i is the predicted value post hypothesis i.e. the probability of the Class being 1.	
Categorical Cross Entropy Loss	Categorical Cross Entropy loss can be envisioned as an extension of Binary Cross Entropy Loss to accommodate multiple classes.
Hinge loss	Hinge loss was originally developed for support vector machines to calculate the largest margin between the hyper plane and the classes. In this context, it's necessary for the score of the correct category to exceed the sum of scores incorrect category scores by a predefined safety margin.
$\boldsymbol{SVMLoss} = \sum_{j \neq y_i} \boldsymbol{max}(\mathbf{0},\ \boldsymbol{S_j} - \boldsymbol{S_{y_i}} + \mathbf{1})$	
Kullback Leibler Divergence Loss (KL Loss)	Kullback-Leibler Divergence Loss measures the deviation of a distribution from a baseline distribution.

3. Loss Functions in Person Re-Identification

Many state of the art deep learning based approaches in person re-identification uses loss functions. Some of the loss functions widely used in person re-identification is as below –

3.1 Triplet Loss[9]

The triplet loss is a fundamental concept of deep learning, commonly used in applications such as face recognition, person re-identification, and similarity learning. It aims to learn embedding's or feature representations that enhance the distinction between different classes or instances in a given dataset. The fundamental concept behind triplet loss is to ensure distance between anchor-positive pairs (representing instances of the same class) is minimized compared to the distance between anchor-negative pairs (representing instances of different classes) within the learned feature space as in Figure 1 shown below.

Triplet Loss initially works with three elements: an image A, a positive image$P$ from the same class, and a negative image$N$ from a different class. Its primary objective is to reduce the distance between A and P while simultaneously increase the separation between the N samples. The formulation of the loss function is as follows:

$$L_{triplet} = \lfloor \parallel f(A) - f(P) \parallel_2^2 - \parallel f(A) - f(N) \parallel_2^2 + \alpha \rfloor_+ \quad (1)$$

Where $[z]_+ = max\ (z, 0)$, α is a margin parameter and f is the embedding function learned during the stage of training.

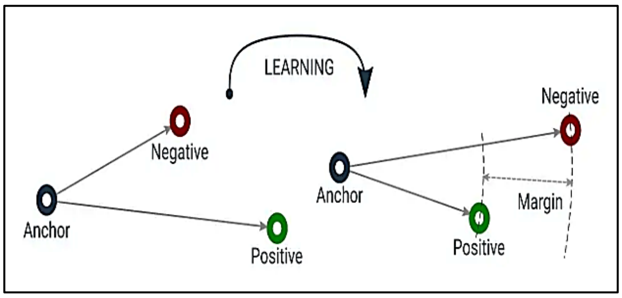

Fig. 1. Concept of Triplet Loss [12]

3.2 Centroid Triplet Loss (CTL)[7]

Rather than assessing the distance between an anchor image A and positive P and negative N instances in the context of Triplet Loss, Central Triplet Loss (CTL) quantifies the separation between A and class centroids c_P and c_N. These centroids symbolize the anchor's class and a distinct class, respectively. Consequently, CTL is structured in the following manner:

$$L_{CTL} = \lfloor \| f(A) - c_P) \|_2^2 - \| f(A) - c_N) \|_2^2 + \alpha_c \rfloor_+ \quad (2)$$

Where α_c is a Margin Parameter of the Class.

3.3 Hybrid Loss[13]

It characterizes the spatial arrangement and separation of features. It augments the gap between different classes while minimizing the gap within the same class. The hybrid loss is designed using triplet loss (L_T) and the cross-entropy loss (L_{CE}). In this context, L_{CE} captures the spatial configuration of pedestrian features, while L_T gauges the dissimilarity between pedestrian features.

$$\text{Hybrid Loss } L_{CETL} = \beta\, L_{CE} + (1\text{-}\,\beta)\, L_T \quad (3)$$

3.4 Quadruple Loss [8]

The conventional Triplet loss tends to exhibit limited generalization on the test set due to the continued presence of a substantial intra-class gap. To address this, a Quadruple loss emerges as an evolution of the triplet loss. This extended approach incorporates four images as input and incorporates online hard negative mining based on a margin, strategically opting for challenging samples for network training. The Quadruplet loss introduces a fresh constraint that effectively increases the distance between negative pairs and positive pairs concerning different probe images.

$$L_{quad} = \lfloor \| f(A) - f(P) \|_2^2 - \| f(A) - f(N_1) \|_2^2 + \alpha_1 \rfloor_+ + \lfloor \| f(A) - f(P) \|_2^2 - \| f(A) - f(N_2) \|_2^2 + \alpha_2 \rfloor_+ \quad (4)$$

Where α_1 and α_2 are the margin parameters of the two terms, P is from the same Class as A and N_1 and N_2 are different from the Classes other than A.

4. Literature Review

There are various categories of re-identification methods, each addressing distinct aspects and challenges. These categories include [14]:

Open Set and Closed Set Re-Identification: Re-identification methods can be classified as open set or closed set, depending on whether they account for unknown or unseen identities during testing.

Temporal-Based Re-Identification: Methods for re-identification can be categorized based on temporal aspects, such as short-term and long-term re-identification, considering the duration of tracking.

Sample Cardinality: Re-identification approaches can also be classified as single shot or multi shot, depending on whether they consider a single image or multiple images for identifying individuals.

Feature-Based Re-Identification: Some methods rely on various features like color, texture, covariance, shape, and position information to perform re-identification.

Metric-Based Re-Identification: Metric-based approaches leverage metrics like Bhattacharya coefficient, quadratic distance, sum of absolute difference, correlation coefficients, Mahalanobis distance, etc., to find the most similar images.

Supervised and Unsupervised Learning: Re-identification techniques can be categorized into supervised and unsupervised approaches based on the availability of corresponding output variables during training.

These diverse categories of re-identification methods cater to different scenarios and data characteristics, offering a wide array of techniques to tackle the challenges in person re-identification.

This section mainly delves into a range of deep learning-based approaches, highlighting their methodologies, deep learning model architectures, employed loss functions, and corresponding outcomes. Several cutting-edge techniques that have made significant contributions to the field are explored and discussed below:

In 2017, Li et. al. [15]designed a multi-granularity framework for feature learning in person re-identification problem and proposes a novel multi-granularity triplet loss function. Their approach seems to have achieved significant improvements over existing approaches in terms of accuracy. The Rank-1 accuracy of their proposed approach is 96.6 % and mAP of 94.2% on the Market-1501 dataset.

In 2017, Janocha et. al. [16] investigated the effect of choosing particular loss function over deep learning model. Using theoretical explanation they showed that regression loses having valid probabilistic interpretation in Deep classifiers, may slower the training process. They concluded two things, firstly that the linear model intuitions rarely transfers to highly non-linear deep model and secondly that based on deep learning applications except Log loss any other Loss can be chosen.

In 2017, Chen et al. [8] introduced an innovative approach known as the quadruplet loss, specifically designed to improve the limitations of the triplet loss in the context of human re-identification (ReID). By incorporating margin-based online hard negative mining, they developed a quadruplet network using quadruplet Loss. This methodology outperforms many existing well known techniques on prominent re-identification datasets like CUHK03, CUHK01, and VIPeR.

In 2018, Luo et al. [17] presented a robust baseline architecture for person re-identification, which incorporates batch normalization and some loss functions like classification loss, center loss and triplet loss.

In 2022, Gu et. al. [18] introduces LFS-ReID, an AutoML approach that automates the search for optimal loss functions. By analyzing the margin-based softmax loss function and leveraging non-independent truncated Gaussian distributions along with reinforcement learning, the method dynamically generates loss functions that exhibit key properties. The approach's superior performance on benchmark datasets underscores its effectiveness in enhancing person re-identification model outcomes.

In 2022, Gu et. al. [19] presented Auto Loss-GMS, a novel automated Re-ID technique for identifying the optimal generalized margin-based softmax loss function tailored for person re-identification. This method leverages reinforcement learning-based optimization to precisely adjust the parameters of the GMS loss function, ensuring a more effective alignment with the specific characteristics of the dataset.

In 2021, Li et. al. [20] introduces a novel loss function TIOM(Triplet Online Instance Matching) for improving human re-identification accuracy. The TOIM loss combines the strengths of the OIM (Online Instance Matching) loss and triplet loss while simplifying batch construction for quicker convergence. This loss function focuses on hard samples, enhancing ReID accuracy, and is particularly suitable for joint detection and identification tasks when applied online.

In 2021, Yang et. al. [21] introduces a novel loss function called Equidistant Distribution Loss (EDL) to tackle the imbalance problem in re-identification (re-id) tasks. The proposed approach involves normalizing learned features and weights to project them onto a hyper sphere space. An equidistance constraint is then enforced among the weights to encourage uniform distribution of learned features across the sphere space. This helps counteract local squeeze or imbalance issues caused by imbalanced samples, thus enhancing performance. The method is validated through experiments on re-id datasets, such as Market-1501 and DukeMTMC-reID, showing that the proposed EDL leads to improved results in addressing the imbalance problem.

In 2019, Sun et. al. [22] provides a comprehensive survey of various techniques and tricks that can enhance the performance of deep person ReID models. In 2020, Chen et. al. [23] presented FastReID, an open-source toolbox that provides an extensive set of features for rapid development of ReID algorithms.

In 2021, Yan et. al. [24]addressed the challenge of achieving accurate person re-identification (ReID) across multiple cameras by introducing an innovative pairwise loss function. Their novel pairwise loss function was specifically designed to assist ReID models in acquiring fine-grained features. This was achieved through the adaptive application of exponential penalization to images with minor differences and bounded penalization to images with significant differences. One notable feature of their proposed loss function is its versatility, as it can effectively replace the traditional triplet loss in various state-of-the-art approaches, resulting in a significant enhancement of performance. Experimental results on 4 different datasets conclude the effectiveness of their newly introduced loss function. Furthermore, it improves data efficiency, ultimately leading to more accurate person re-identification across diverse scenarios.

In 2021 Ha et. al. [25] introduces a modification to the traditional fixed margin triplet loss, proposing an adaptive margin triplet loss. While the original triplet loss is commonly used in classification tasks like face recognition and re-identification, the new loss is particularly suitable for rating datasets where ratings are continuous values. Unlike the original triplet loss, which requires careful data sampling, the proposed approach can generate triplets from the entire dataset, avoiding model collapse issues during optimization.

In 2022, Drager et. al. [26] analyzed the effect of changing loss function in different deep architectures on training process and its performance. They recommended to use KLD (Kullback – Leibler Divergance) Loss in binary and multi class Classification regimes.

In 2022, Hu et. al. [27]presented a novel loss function in person re-identification aimed at reducing the impact of intra-pair variations and increasing optimization gradients. Their approach optimize the ratio between intra-identity distance and intra-identity distance, resulting in what they call "triple ratio loss". This unique loss function relies on a triad of pedestrian images for its implementation. Extensive experimental evaluations were performed on well-established human Re-ID datasets including Market-1501, DukeMTMC-ReID, CUHK03, and MSMT17. These experiments conclusively show that triplet ratio loss outperforms traditional triplet loss.

5. Datasets

Various datasets commonly used for person re-identification (Re-ID) research include[14]:

VIPeR: Featuring two cameras, VIPeR poses challenges due to substantial viewpoint and illumination variations, making it a challenging dataset.

GRID: Collected from 8 non-overlapping cameras in an underground station, GRID has 250 image pairs, with more gallery images than probe images. It's challenging due to viewpoint variations, background clutter, occlusions, and image quality issues.

CAVIAR4ReID: Collected from two surveillance cameras with overlapping views in a shopping mall, this dataset includes 72 identities and suffers from viewpoint variations and low-resolution images.

PRID: It consists of 385 trajectories from camera A and 749 trajectories from camera B, with only 200 people appearing in both cameras. It also has a single-shot version with unsynchronized trajectories.

CUHK01: This dataset contains two images for each identity from each camera, with relatively good image quality.

CUHK02: An extended dataset from CUHK01, it includes four additional camera pairs.

CUHK03: The first large-scale Re-ID dataset suitable for deep learning, providing bounding boxes and manual labeling with good person detection quality.

RAiD: A relatively new dataset with images from four non-overlapping cameras, with substantial illumination variations due to two indoor and two outdoor cameras.

iLIDS-VID: Extracted 600 trajectories for 300 identities from iLIDS MCTS dataset, suffering from heavy occlusion.

DukeMTMC: A large-scale multi-camera tracking dataset with more than 2700 labeled people and extensive potential for research.

DukeMTMC-reID: A subset of DukeMTMC, it contains 1,812 identities from eight cameras.

DukeMTMC4ReID: Derived from DukeMTMC, this dataset includes 1,852 people with unique identities and various bounding box sizes.

Market1501: Contains numerous identities with images from six disjoint cameras, including distractors and false alarms.

RPIfield: Provides timestamp information for each person, facilitating temporal performance evaluation.

Airport: Created using videos from an indoor surveillance network in an airport, with video clips and end-to-end re-ID system for testing.

MSMT17: The largest re-ID dataset to date, captured in a campus with 12 outdoor and 3 indoor cameras, featuring diverse scenarios and complex weather conditions.

This paper uses 2 most widely used datasets, Market 1501 and DukeMTMC-reID for person re-identification which are further summarized as below –

Market-1501[32] **-** The Market-1501 dataset is a valuable resource for person re-identification, obtained from surveillance footage in front of Tsinghua University Supermarket. This dataset consists of six cameras, five high-resolution and one low-resolution, often showing overlapping viewpoints. Built primarily for person re-identification tasks, Market-1501 consists of 32,668 labeled bounding boxes representing 1,501 persons captured from six different angles.

The dataset is divided into two main sections: the training set consists of 12,936 images spanning 751 different identities, while the test set consists of 19,732 images corresponding to 750 unique identities. During the testing phase, an additional 3,368 hand-drawn images, representing the same 750 identities, were used as a verification set to identify the correct identities within the test set. It is important to note that the gallery image is used for identification, while the probe image is the test image matched against the gallery images. Sample Snaps of the Market 1501 dataset is shown in Figure 2.

The Market-1501 dataset includes annotations for 27 features, covering 751 identities in the training set and 750 identities in the testing set, all annotated at the identity level. As a result, the dataset file contains 27 x 751 features for training and 27 x 750 features for testing.

Fig. 2. Snaps of the Market 1501 dataset.

Table 3: Summary of Loss function based Person Re-Identification Approaches

Loss Function	Ref.	Year	Approach	Deep Network Model	Data-Set	Result			
						Rank-1 (in %)	Rank-5	Rank-10	mAP
Triplet Loss	[9]	2017	• Adam optimizer • batch normalization • Variations of triplet loss • Soft margin	TriNet (Trained with our batch hard triplet loss)	Market 1501 MQ	90.53	96.29	-	76.42
					Market-1501 SQ	84.92	94.21	-	69.14
					MARS	79.8	91.36	-	67.7
					CHUK03 (Labeled)	89.63	99.01	-	-
				LuNet (trained from scratch)	Market 1501 MQ	87.11	95.16	-	69.07
					Market-1501 SQ	81.38	92.34	-	60.71
					MARS	75.56	89.7	-	60.48
Centroid Triplet Loss	[2]	2021	• Cosine Similarity • Cross View matching • Aggregated item representation using all available samples • Centroid based approach	ResNet- 50	Market-1501	98	98.6	99.5	98.3
					DukeMTMC-Reid	95.6	96.2	97.9	96.1
Circle Loss	[28]	2020	• Incorporating the resemblances between different classes and within the same class into pairs of similarities. • Within a supervised similarity pair, each similarity score is subject to varying penalty strength based on	ResNet 50	Market 1501	94.2	-	-	84.9
					MSMT 17	76.3	-	-	50.2

			its proximity to the optimal value.						
				MGN	Market 1501	96.1	-	-	87.4
					MSMT 17	76.9	-	-	52.1
	[29]	2021	Modern Training Techniques [22]	ResNet 50	Market 1501	95.5	-	-	88.4
Quadruple Loss	[8]	2017	Margin based online hard Negative Mining	Caffe Framework	CHUK01(P=100)	81	96.5	98	-
					CHUK01(P=486)	62.55	83.44	89.71	-
					CHUK03	75.53	95.15	99.16	-
					VIPeR	49.05	73.1	81.96	-
Centre Loss	[22]	2019	• Global features • BN Neck • Random Erasing Augmentation (REA)	ResNet 50	Market 1501	94.5	-	-	85.9
					DukeMTMC-reID	86.4	-	-	76.4
Circle based Ratio Loss	[30]	2020	• Improved Soft max Loss • Ratio Loss	ResNet 50	Market 1501	92.64	-	-	83.12
					DukeMTMC-reID	84.34	-	-	71.66
					CHUK03 Labeled	68.57	-	-	66.26
					CHUK03 Detected	65.07	-	-	63.24
Hybrid Loss	[13]	2018	• New cross entropy triplet loss • Designed strategy to mine hard triplets to accelerate the learning	Proposed Residual Neural Network with 50 Layers, epochs N=300, Batch Size M=128, Initial Learning rate =0.0002 etc.	Market -1501 (Multi)	92.7	-	-	82.4
					DukeMTMC	77.8	-	-	60.3

			• Adam gradient optimization method		CHUK-01	82.7	95.5	97.3	-
			• Global and Local Optimization		CHUK-03	86.6	97.5	98.8	-
Additive Angular Margin Loss (AAML)	[31]	2019	• Average Pooling	ResNet 50	Market -1501	79.2	91	-	57.3
			• Batch Normalization		DukeMTMC-reID	74.3	89.3	-	54.8
	[29]	2021	• Modern Training Techniques [22]	ResNet 50	Market -1501	95.5	-	-	88.1

DukeMTMC Re-ID[33] **-** The DukeMTMC Re-ID dataset is a widely recognized benchmark dataset commonly used in the field of computer vision for person re-identification (Re-ID) tasks. It contains a vast collection of over 36,000 images, each depicting one of 1,812 unique individuals and captured by eight different cameras. This dataset presents significant challenges, including variations in lighting, changes in clothing, and occlusion patterns. The Sample of snaps are shown in Figure 3.

Fig. 3. DukeMTMC-reID showing variations in background, illumination, viewpoints in different cameras

Moreover, the dataset includes valuable resources such as annotated bounding boxes, different training and test sets obtained from different cameras, as well as standard evaluation metrics such as CMC (cumulative matching characteristics) and mAP (average precision). This metric serves as an established benchmark for evaluating the performance of person re-identification algorithms.

6. Proposed Approach

This paper introduces a novel Centroid Quadruple Loss (CQL) based methodology that strategically leverages the advantages of the Centroid Loss [7] and Quadruple Loss [8]. Operating within the domain of the Centroid approach involves the computation of distances between a given sample and a representative class prototype, known as the centroid. By embracing this centroid-centric methodology, the approach generates a singular embedding per item, effectively streamlining memory and storage requirements. The primary advantages of the Centroid approach encompass reduced computational burden, increased resilience against outliers and noisy labels, expedited training processes, and comparable or superior performance when contrasted with standard loss functions.

In contrast, the Quadruplet loss methodology addresses inherent limitations observed in both binary classification loss and triplet loss methodologies, effectively capitalizing on their distinct strengths within person Re-ID applications. This fusion of methodologies culminates in a discernible enhancement in performance compared to the utilization of either approach in isolation. The amalgamation of the Quadruplet loss enriches the model's capabilities, leading to improved performance metrics within the realm of person re-identification methodologies.

Quadruple Loss [8] is given by -

$$L_{quad} = \lfloor \| f(A) - f(P) \|_2^2 - \text{Min} (\| f(A) - f(N_1) \|_2^2) + \alpha_1 \rfloor_+ + \lfloor \| f(A) - f(P) \|_2^2 - \text{Min}$$

$$(\| f(A) - f(N_2) \|_2^2) + \alpha_2 \rfloor_+ \quad (5)$$

Proposed Centroid Quadruple Loss is given by -

$$L_{CQL} = \lfloor \| f(A) - f(c_P) \|_2^2 - \| f(A) - f(c_{N1}) \|_2^2 + \alpha_1 \rfloor_+ + \lfloor \| f(A) - f(c_P) \|_2^2 - \| f(c_{N1}) - f(c_{N2}) \|_2^2 + \alpha_2 \rfloor_+ \quad (6)$$

In Eq. 5, A is the anchor image, P is the image from the same Class as A, N_1 and N_2 are the images chosen from different classes other than Class of A. In the proposed CQL in Eq. 6, centroids of the respective classes are considered. This CQL based approach uses ResNet-50 model. Other techniques used in the proposed model include Cosine Similarity, cross view matching and aggregated item representation using all available samples. Figure 4 shows the Architecture of our proposed approach.

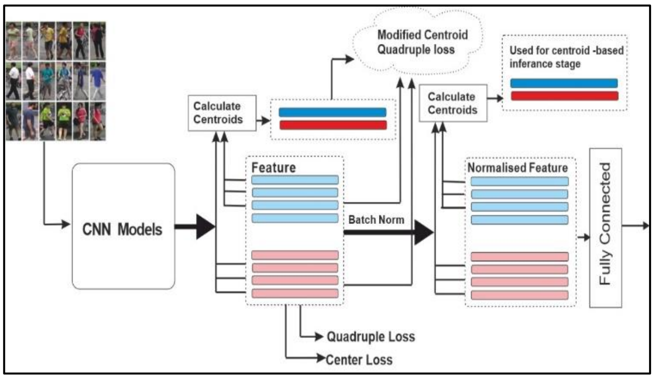

Fig. 4. Architecture of proposed CQL based approach

7. Implementation and Result

The proposed Centroid Quadruple Loss (CQL) based model has been put into action using a computing system powered by a Xeon(R) Gold CPU running at 2.00 GHz with a total of 40 cores and equipped with 92.9 GB of RAM. In our implementation, we maintain consistency with the parameters outlined in the reference paper [7], particularly with the ResNet-50 Model and other associated settings.

Utilizing Centroid Triplet Loss (CTL) on the Market-1501 dataset, we have employed several cutting-edge Deep Learning Models, namely ResNet 50, ResNet 152, DensNet 121, and DensNet-169, for person re-identification. The objective was to determine the most effective model within a similar implementation environment. Examination of the outcomes presented in Table 4 reveals a notable trend: the ResNet versions exhibit a slightly superior performance compared to the DensNet versions under the specified conditions.

Subsequently, we extended our investigation by implementing Centroid Quadruple Loss (CQL) on ResNet versions, including ResNet-50 and ResNet-152, using both Market 1501 and DukeMTMC-reID datasets. The results, as depicted in Tables 5 and 6, showcase comparable mean Average Precision (mAP) and Accuracy for both ResNet models. Consequently, to make a more informed decision, we delved into the analysis of execution times. Upon careful consideration of execution times, a clear distinction emerged: ResNet 50 outperformed ResNet 152 in the given scenario. This finding underscores the importance of not only evaluating performance metrics but also considering computational efficiency when selecting the most suitable model for person re-identification tasks.

Table 4: Results of CTL on Various Deep learning Models on Market-1501

Model	***mAP***	***Rank-1***	***Rank-5***	***Rank-10***	***Rank-20***	***Rank-50***
ResNet -50	98.6	98.2	99.0	99.6	99.8	99.9
ResNet -152	98.8	98.5	99.6	99.6	99.8	99.9
DensNet-121	98.06	97.7	99.5	99.5	99.6	99.9
DensNet-169	98.29	98.0	99.4	99.4	99.6	99.9

Table 5: Result of CQL using ResNet 50 and 152 models on Market 1501

Model	***mAP***	***Rank-1***	***Rank-5***	***Rank-10***	***Execution Time (in Seconds)***
ResNet 50	98.6	98.4	99.1	99.6	9942.20
ResNet 152	98.77	98.5	99.3	99.6	20762.79

Table 6: Result of CQL using ResNet 50 and 152 models on DukeMTMC-reID

Model	***mAP***	***Rank-1***	***Rank-5***	***Rank-10***	***Execution Time (in Seconds)***
ResNet 50	97.4	96.9	98.5	98.9	11807.19
ResNet 152	97.35	96.9	98.4	98.9	24731.25

To gauge the effectiveness of our CQL-based model, we conducted a comparative analysis against existing state-of-the-art methodologies using two benchmark datasets: DukeMTMC-reID and Market 1501. Our aim is to assess the model's performance and efficiency in the context of person re-identification tasks.

The outcomes of these evaluations have been meticulously organized and presented in a comprehensive table 7. This table serves to illustrate and quantify the performance metrics of our proposed CQL model in contrast to other leading approaches when applied to the aforementioned datasets. These metrics encompass various aspects of model efficacy, such as accuracy and precision, among others. This comparative analysis provides a clear and insightful view of how our proposed model fares against the existing state-of-the-art techniques in these specific re-identification tasks.

Table 7: Comparative Results of State of the art approaches with the Proposed Model

Datasets	***Approach***	***mAP***	***Rank-1***	***Rank-5***	***Rank-10***
DukeMTMC-reID	Circle based Ratio Loss [30]	71.6	84.34	-	-
	Centre Loss [22]	76.4	86.4	-	-
	St-ReID Model [34]	92.7	94.5	96.8	97.1
	CTL Model [7]	96.1	95.6	96.2	97.9
	Proposed CQL Model	**97.3**	**96.9**	**98.5**	**98.9**
Market 1501	Circle based Ratio Loss [30]	83.1	92.64	-	-
	Centre Loss [22]	85.9	94.5	-	-
	St-ReID Model [34]	95.5	98.0	98.9	99.1
	CTL Model [7]	98.3	98.0	98.6	99.5
	Proposed CQL Model	**98.6**	**98.4**	**99.1**	**99.6**

The comprehensive analysis of our findings unequivocally illustrates the substantial enhancement brought about by our proposed CQL based model. Through meticulous evaluation and rigorous comparison against the current state-of-the-art methodologies, our model showcases a remarkable and statistically significant improvement in performance metrics across the Market 1501 and DukeMTMC-reID datasets.

By meticulously scrutinizing and quantifying various performance indicators—such as accuracy, and precision, the empirical evidence substantiates that our CQL-based model outperforms the existing leading approaches in a meaningful and impactful manner. These results stand as a testament to the efficacy and superiority of our proposed model in the domain of person re-identification tasks, underlining its potential to significantly advance the field's benchmarks and capabilities.

8. Future Scope

We have implemented 2 Hybrid Loss functions H1 and H2 combining CTL and CQL with different weights as shown below –

Hybrid Loss, H1 = CTL+CQL

Hybrid Loss, H2 = 0.2*CTL+0.8*CQL

Table 8: Result of proposed Hybrid Loss using ResNet-50 model on Market-1501

Loss Function	***mAP***	***Rank-1***	***Rank-5***	***Rank-10***	***Execution Time (in Seconds)***
CQL	98.6	98.4	99.1	99.6	9942.20
H1	98.5	98.2	99.0	99.5	18501.13
H2	98.6	98.2	99.1	99.7	18366.34

The findings presented in Table 8 suggest that amalgamating multiple loss functions has the potential to enhance results, provided that the weights are judiciously chosen. A direct comparison between H1 and H2 indicates that H2 exhibits a slight performance advantage, accompanied by a relatively shorter execution time. Notably, when contrasted with the proposed Centroid Quadruple Loss (CQL), the outcomes for H2 remain largely consistent, albeit with nearly twice the execution time.

This observation prompts the consideration that, akin to the successful characteristics of H2, there is room for the development of a more rational and effective hybrid function. Such a hybrid function could be tailored to further improve both accuracy and precision without compromising computational efficiency. This underscores the importance of not only evaluating performance metrics but also optimizing the balance between effectiveness and computational cost when devising hybrid loss functions for improved model performance.

9. Conclusion

This paper delves into the intricate landscape of loss functions pertinent to person re-identification. Within this context, the introduced Centroid Quadruple Loss emerges as a novel proposition, amalgamating the advantageous aspects inherent in both the Centroid approach and the Quadruple loss technique. In the empirical evaluation conducted using the Market 1501 and DukeMTMTC-reID datasets, this paper endeavours to juxtapose the performance of the proposed Centroid Quadruple Loss against the backdrop of prevailing state-of-the-art methodologies. The purpose is to discern and elucidate the comparative efficacy and capabilities of this newly introduced loss function within the existing architectural framework. The findings ensuing from this comparative analysis evince a discernible and noteworthy augmentation in performance metrics. Specifically, the application of the newly introduced Centroid Quadruple Loss function showcases a pronounced improvement over the established methods when subjected to evaluation on both the Market 1501 and DukeMTMTC-reID datasets. These empirical results substantiate the potency and efficacy of this novel loss function, highlighting its potential to significantly elevate the standards of performance in the realm of person re-identification methodologies.

Acknowledgements

We thank Mr. Hiten Sadani, Assistant Professor from UVPCE, Ganpat University, India for assistance and for comments that greatly improved the manuscript.

Author contributions

Nikhil Kumar Singh: Conceptualization, Methodology, Draft Writing **Manish Khare:** Final draft preparation, Validation. **Harikrishna B. Jethva:** Visualization, Investigation, Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] A. Peña et al., "Human-Centric Multimodal Machine Learning: Recent Advances and Testbed on AI-Based Recruitment," SN Comput. Sci., vol. 4, no. 5, 2023, doi: 10.1007/s42979-023-01733-0.

[2] K. Singh, M. K. Ahirwal, and M. Pandey, "Mental Health Monitoring Using Deep Learning Technique for Early-Stage Depression Detection," SN Comput. Sci., vol. 4, no. 6, p. 701, 2023, doi: 10.1007/s42979-023-02113-4.

[3] Y. Salini and J. HariKiran, "ViT: Quantifying Chest X-Ray Images Using Vision Transformer & XAI Technique," SN Comput. Sci., vol. 4, no. 6, p. 754, 2023, doi: 10.1007/s42979-023-02204-2.

[4] L. Bazzani, M. Cristani, and V. Murino, "Symmetry-Driven Accumulation of Local Features for Human Characterization and Re-identification."

[5] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Deep metric learning for person re-identification," in 2014 22nd International Conference on Pattern Recognition, 2014, pp. 34–39.

[6] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep Filter Pairing Neural Network for Person Re-identification," in 2014 IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2014, pp. 152–159, doi: 10.1109/CVPR.2014.27.

[7] M. Wieczorek, B. Rychalska, and J. Dąbrowski, On the Unreasonable Effectiveness of Centroids in Image Retrieval, vol. 13111 LNCS, no. 1. Association for Computing Machinery, 2021.

[8] W. Chen, X. Chen, J. Zhang, and K. Huang, "Beyond triplet loss: A deep quadruplet network for person re-identification," Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-January, pp. 1320–1329, 2017, doi: 10.1109/CVPR.2017.145.

[9] A. Hermans, L. Beyer, and B. Leibe, "In Defense of the Triplet Loss for Person Re-Identification," 2017, [Online]. Available: http://arxiv.org/abs/1703.07737.

[10] V. Yathish, "Loss Functions and Their Use In Neural Networks," Towards Data Science, Aug. 04, 2022. https://towardsdatascience.com/loss-functions-and-their-use-in-neural-networks-a470e703f1e9 (accessed Apr. 12, 2023).

[11] K. Mahendru, "Understanding Loss Functions to Maximize Machine Learning Model Performance (Updated 2023)," Aug. 14, 2019. https://www.analyticsvidhya.com/blog/2019/08/detailed-guide-7-loss-functions-machine-learning-python-code/ (accessed May 25, 2023).

[12] Y. Sarıgöz, "Triplet Loss — Advanced Intro," Towards Data Science, Mar. 25, 2022. https://towardsdatascience.com/triplet-loss-advanced-intro-49a07b7d8905.

[13] Z. Hu, H. Wu, S. Liao, H. Hu, S. Liu, and B. Li, "Person Re-Identification with Hybrid Loss and Hard Triplets Mining," in 2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM), 2018, pp. 1–5, doi: 10.1109/BigMM.2018.8499463.

[14] N. K. Singh, M. Khare, and H. B. Jethva, "A comprehensive survey on person re-identification approaches: various aspects," Multimed. Tools Appl.,

vol. 81, no. 11, pp. 15747–15791, 2022, doi: 10.1007/s11042-022-12585-w.

[15] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, "Learning discriminative features with multiple granularities for person re-identification," MM 2018 - Proc. 2018 ACM Multimed. Conf., pp. 274–282, 2018, doi: 10.1145/3240508.3240552.

[16] K. Janocha and W. M. Czarnecki, "On loss functions for deep neural networks in classification," Schedae Informaticae, vol. 25, pp. 49–59, 2016, doi: 10.4467/20838476SI.16.004.6185.

[17] H. Luo et al., "A Strong Baseline and Batch Normalization Neck for Deep Person Re-Identification," IEEE Trans. Multimed., vol. 22, no. 10, pp. 2597–2609, 2020, doi: 10.1109/TMM.2019.2958756.

[18] H. Gu, J. Li, G. Fu, M. Yue, and J. Zhu, "Loss function search for person re-identification," Pattern Recognit., vol. 124, p. 108432, 2022, doi: https://doi.org/10.1016/j.patcog.2021.108432.

[19] H. Gu, J. Li, G. Fu, C. Wong, X. Chen, and J. Zhu, "AutoLoss-GMS: Searching Generalized Margin-based Softmax Loss Function for Person Re-identification," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2022-June, pp. 4734–4743, 2022, doi: 10.1109/CVPR52688.2022.00470.

[20] Y. Li, G. Yin, C. Liu, X. Yang, and Z. Wang, "Triplet online instance matching loss for person re-identification," Neurocomputing, vol. 433, pp. 10–18, 2021, doi: https://doi.org/10.1016/j.neucom.2020.12.018.

[21] Z. Yang, J. Liu, T. Liu, Y. Zhu, L. Wang, and D. Tao, "Equidistant distribution loss for person re-identification," Neurocomputing, vol. 455, pp. 255–264, 2021, doi: https://doi.org/10.1016/j.neucom.2021.04.070.

[22] H. Luo, Y. Gu, X. Liao, S. Lai, and W. Jiang, "Bag of tricks and a strong baseline for deep person re-identification," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., vol. 2019-June, pp. 1487–1495, 2019, doi: 10.1109/CVPRW.2019.00190.

[23] L. He, X. Liao, W. Liu, X. Liu, P. Cheng, and T. Mei, "FastReID: A Pytorch Toolbox for General Instance Re-identification," 2020, [Online]. Available: http://arxiv.org/abs/2006.02631.

[24] C. Yan et al., "Beyond Triplet Loss: Person Re-Identification With Fine-Grained Difference-Aware Pairwise Loss," IEEE Trans. Multimed., vol. 24, pp. 1665–1677, 2022, doi: 10.1109/TMM.2021.3069562.

[25] M. L. Ha and V. Blanz, "Deep Ranking with Adaptive Margin Triplet Loss," 2021, [Online]. Available: http://arxiv.org/abs/2107.06187.

[26] S. Dräger and J. Dunkelau, "Evaluating the Impact of Loss Function Variation in Deep Learning for Classification," 2022, [Online]. Available: http://arxiv.org/abs/2210.16003.

[27] S. Hu, K. Wang, J. Cheng, H. Tan, and J. Pang, "Triplet Ratio Loss for Robust Person Re-identification," in Pattern Recognition and Computer Vision, 2022, pp. 42–54.

[28] Y. Sun et al., "Circle loss: A unified perspective of pair similarity optimization," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 6397–6406, 2020, doi: 10.1109/CVPR42600.2020.00643.

[29] D. Aganian, M. Eisenbach, J. Wagner, D. Seichter, and H. M. Gross, "Revisiting Loss Functions for Person Re-identification," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 12895 LNCS, pp. 30–42, 2021, doi: 10.1007/978-3-030-86383-8_3.

[30] Z. Yang, J. Liu, T. Liu, L. Wang, and S. Zhao, "Circle-Based Ratio Loss for Person Reidentification," Complexity, vol. 2020, 2020, doi: 10.1155/2020/9860562.

[31] J. SU, X. He, L. Qing, Y. Yu, S. Xu, and Y. Peng, "A New Discriminative Feature Learning for Person Re-Identification Using Additive Angular Margin Softmax Loss," in 2019 UK/ China Emerging Technologies (UCET), 2019, pp. 1–4, doi: 10.1109/UCET.2019.8881838.

[32] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1116–1124.

[33] Z. Zheng, L. Zheng, and Y. 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, pp. 3754–3762.

[34] G. Wang, J. Lai, P. Huang, and X. Xie, "Spatial-Temporal Person Re-Identification," 2018.