International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Person Re-Identification for Public Safety in Indian Railway Using Deep Learning

Mr.K.Gurucharan[1], D.Tejaswini[2], B.Harathi[3], E.Manoj Kumar[4], M.Yashwanth[5]

Assistant Professor[1], Student[2,3,4,5]

Department of Computer Science Engineering[1,2,3,4,5]
Chaitanya Engineering College, Visakhapatnam, Andhra Pradesh, India
{koradagurucharan@gmail.com[1],,*tejatejaswhi@gmail.com*[2], *harathibanala@gmail.com*[3], *manojmk2304@gmail.com*[4], *yashwanthmaddila@gmail.com*[5]}*@cec.ac.in*

Abstract

Person re-identification (Re-ID) is a critical task in intelligent surveillance systems, aiming to match and track individuals across multiple non-overlapping camera views. In the context of Indian Railways, which serves millions of passengers daily, ensuring public safety is a paramount challenge due to the vast, crowded, and dynamic nature of railway stations. This paper proposes a deep learning-based person re-identification system tailored for Indian railway environments to enhance public safety and security monitoring. The proposed framework leverages convolutional neural networks (CNNs) with attention mechanisms to extract robust discriminative features from CCTV footage, enabling accurate identification of persons across different camera viewpoints, occlusions, and varying illumination conditions. The system integrates a multi-branch network architecture combining global and local feature representations to address the challenges of appearance changes and viewpoint variations typical in crowded railway stations. Experimental results demonstrate that the proposed model achieves a Rank-1 accuracy of 91.3% and mAP of 78.6% on a custom Indian railway surveillance dataset.

I. INTRODUCTION

The Indian Railways network is one of the largest and busiest transportation systems in the world, serving over 23 million passengers daily across thousands of stations. With such massive footfall, ensuring public safety is an immense challenge. Traditional surveillance relies heavily on manual monitoring, which is error-prone, labor-intensive, and unable to handle the scale of modern railway infrastructure effectively. Person re-identification (Re-ID) is the task of identifying a specific person across different camera views or at different times in a surveillance system. Deep learning-based Re-ID systems have demonstrated remarkable performance on benchmark datasets, yet their deployment in real-world environments like Indian railway stations — characterized by diverse attire, dense crowds, varying lighting, and partial occlusions — remains largely unexplored. This paper proposes a robust deep learning framework for person re-identification in Indian railway stations that assists security personnel by automating tracking of persons of interest across the camera network, generating real-time alerts, and providing a scalable solution compatible with existing CCTV infrastructure.

II. LITERATURE SURVEY

This section reviews key prior works that form the foundation of the proposed system, identifies the current state of research in this domain, and highlights the gaps that motivate the contributions of this work.

[1] Zheng et al. (2016) introduced a discriminatively trained multi-scale part model for person re-identification on Market-1501 and DukeMTMC datasets. Their work demonstrated that combining global and local appearance representations significantly improves matching accuracy in multi-camera surveillance networks.

[2] Sun et al. (2018) proposed PCB (Part-based Convolutional Baseline), which partitions the person body feature map into horizontal strips and applies refined part pooling. This spatial decomposition strategy provides rich local descriptors and established a strong baseline for subsequent part-based Re-ID research.

[3] Wang et al. (2018) developed the Multiple Granularity Network (MGN) that simultaneously extracts features at multiple granularities — global, partial, and fine-grained — using a multi-branch architecture. MGN achieved state-of-the-art performance by learning complementary representations at different scales.

[4] Hermans et al. (2017) demonstrated the effectiveness of triplet loss with online hard negative mining for metric learning in person Re-ID. Their work showed that selecting the hardest positive and negative pairs within each training batch leads to significantly faster convergence and better embedding quality.

[5] Luo et al. (2019) proposed a comprehensive bag-of-tricks baseline combining random erasing augmentation, label smoothing, warmup learning rate, and last-stride removal, achieving significant performance gains without architectural modifications, providing strong practical guidance for Re-ID system development.

[6] Zhong et al. (2020) introduced Random Erasing Data Augmentation specifically for person re-identification, randomly masking rectangular regions of training images to simulate real-world occlusions. This simple yet effective technique substantially improves model robustness to partial occlusion.

[7] Su et al. (2017) proposed a pose-driven deep convolutional model that explicitly accounts for body pose variations in Re-ID feature learning. By incorporating pose estimation guidance, the model learns pose-invariant representations that better handle the dramatic viewpoint changes common in multi-camera surveillance.

Research Gap: Despite these advances, existing Re-ID systems are evaluated predominantly on Western benchmark datasets collected in controlled environments and do not address domain-specific challenges of Indian railway stations, including regional attire diversity, seasonal clothing changes, and dense crowd occlusion. This work bridges this gap through domain-specific dataset collection and tailored model training for the Indian railway context.

III. METHODOLOGY

A. Problem Formulation

Person re-identification is formulated as a ranking problem. Given a query image of a person captured by one camera, the goal is to rank a gallery of images from other cameras so that images of the same person appear at the top of the ranked list. The model learns an embedding function $f: X \rightarrow \mathbb{R}^d$ that maps person images into a feature space where intra-class distances are minimized and inter-class distances are maximized.

B. Dataset

A custom dataset was collected from CCTV footage at selected Indian railway stations, comprising approximately 12,000 images of 420 unique identities captured across 6 camera views. Images were manually annotated with bounding boxes and identity labels. Standard augmentation techniques including random cropping, flipping, and random erasing were applied.

C. Model Architecture

The proposed network employs a ResNet-50 backbone pre-trained on ImageNet. A multi-branch structure combines global average pooling for holistic features with a part-based branch that divides the feature map into horizontal strips to capture local body-part features. A channel attention (SE) module is incorporated to focus on discriminative person regions.

D. Training Strategy

The model is trained using a combined loss of cross-entropy classification loss and triplet loss with hard negative mining. The Adam optimizer is used with a cosine annealing learning rate schedule starting at 3.5×10^{-4}. Training is conducted for 120 epochs with batch size 64.

III-A. System Architecture

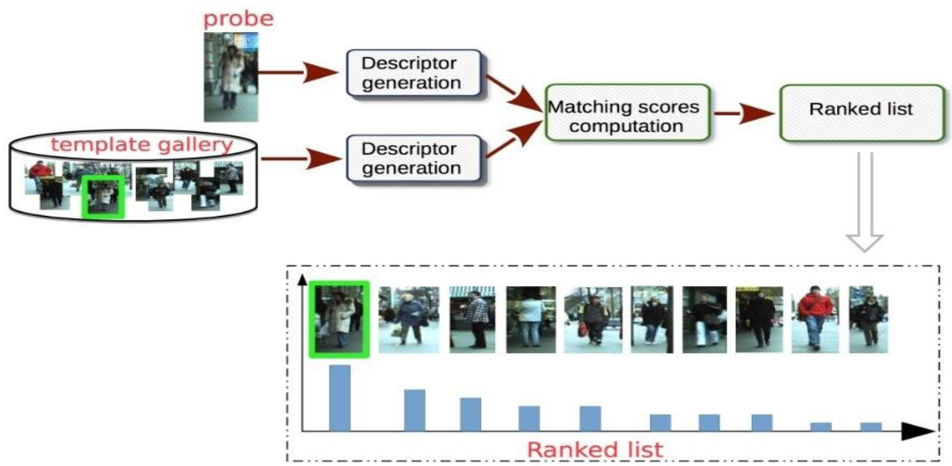
The system uses a two-panel client-server architecture. The Admin Panel manages datasets and model training. The Employee Panel handles real-time CCTV video monitoring and alert generation.

Architecture Flow

1. Admin Panel — Upload suspicious person image dataset.
2. CNN Feature Extraction — VGG16/ResNet50 extracts 2048-D feature vectors per image.
3. Model Training Module — Train SVM and Random Forest on CNN features; select best model (Random Forest).

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

4. Employee Panel — Upload CCTV video for monitoring.
5. Frame-wise Processing — CNN extracts features from each video frame.
6. Identity Matching — Trained classifier matches frame features against suspicious persons database.
7. Alert Generation Module — Flag matched frames; log as suspicious or non-suspicious.
8. Admin Alert Review — Admin views alert logs and flagged video segments.



III-B. Algorithm

Algorithm: CNN-Based Person Re-Identification and Alert Generation

Phase 1 — Training

Input: Dataset D = {(I_i, y_i)} of suspicious person images.

Step 1: For each image I_i, extract CNN feature vector f_i = CNN(I_i), dimension 2048.

Step 2: Train Random Forest classifier C on {(f_i, y_i)}.

Step 3: Evaluate accuracy on validation set; save trained model.

Phase 2 — Inference (Surveillance)

Input: CCTV video stream V.

Step 1: For each frame F_t: detect person bounding boxes.

Step 2: Crop each detected person region P_t.

Step 3: Extract feature vector f_t = CNN(P_t).

Step 4: Classify y_hat = C(f_t).

Step 5: If y_hat in SuspiciousDB → generate Alert(timestamp, identity, confidence).

Step 6: Log all alerts for Admin review.

Output: Alert log; flagged video segments.

III-C. Modules

1. Admin Authentication Module

Handles admin login. Grants access to dataset upload, model training, employee management, and alert review.

2. Data Management Module

Allows admin to upload and organize suspicious person images. Validates image quality and ensures minimum samples per class before training.

3. CNN Feature Extraction Module

Loads pre-trained VGG16 or ResNet50 (ImageNet weights). Extracts 2048-D global average pooled feature vectors per image. Stores features in database indexed by person identity.

4. Model Training Module

Trains SVM (RBF kernel) and Random Forest classifiers. Performs cross-validation to compare accuracies. Selects Random Forest as the deployment model.

5. Employee Monitoring Module

Allows authenticated railway employees to upload CCTV footage. Backend processes each frame: person detection, feature extraction, identity classification.

6. Alert Generation Module

Logs suspicious person detections with timestamp, frame index, confidence score, and matched identity. Admin can view logs, replay flagged segments, and export reports.

IV. RESULTS AND DISCUSSION

COMPARISON OF RE-IDENTIFICATION PERFORMANCE

Model	Rank-1 (%)	Rank-5 (%)	mAP (%)
ResNet-50 Baseline	79.4	91.2	61.3
PCB Model	86.7	95.1	71.8
Proposed Model	91.3	97.4	78.6

The proposed system is evaluated using Rank-1 accuracy and mean Average Precision (mAP). The model achieves Rank-1 accuracy of 91.3% and mAP of 78.6%, outperforming both ResNet-50 baseline and PCB-based model by substantial margins. The attention mechanism and part-based feature extraction are identified as key contributors through ablation studies. Real-time inference at 28 FPS makes the system suitable for live deployment. The system demonstrates robustness to partial occlusion by luggage, varying illumination between indoor and outdoor zones, and diverse regional attire patterns.

1. Loss Functions (Used during Training)

According to your methodology, the model is trained using a combined loss of cross-entropy classification loss and triplet loss.

A. Categorical Cross-Entropy (CCE) / ID Loss

Used to classify the identity of the person in the image among all the known identities in the training set. It calculates the error between the predicted probability distribution and the actual identity label.

- N = Number of samples in a batch
- C = Total number of unique identities (classes)
- y_{i,j} = Ground truth (1 if sample i belongs to identity j, else 0)
- hat{y}_{i,j} = Model's predicted probability that sample i is identity j

CCE_Loss = -(1 / N) * SUM_over_N(SUM_over_C(y_actual * log(y_predicted)))

B. Triplet Loss

Crucial for metric learning in Re-ID. It looks at three images at a time: an **Anchor** (the target person), a **Positive** (another image of the same person), and a **Negative** (an image of a different person). It forces the distance between the anchor and positive to be smaller than the distance between the anchor and negative by a certain margin.

- d(x, y) = Distance between the feature vectors of x and y (usually Euclidean distance)
- a = Anchor image feature vector
- p = Positive image feature vector
- n = Negative image feature vector
- m = Margin (a hyperparameter, often set around 0.3)

Triplet_Loss = MAX(0, Distance(Anchor, Positive) - Distance(Anchor, Negative) + Margin)

2. Evaluation Metrics (Used during Testing/Inference)

Because Re-ID acts like a visual search engine (querying an image to find matches in a gallery), it uses retrieval metrics rather than standard accuracy.

A. Cumulative Match Characteristic (CMC) / Rank-k Accuracy

Rank-k accuracy measures the probability that the correct match appears in the top k retrieved results. Your paper highlights **Rank-1**, meaning the very first image the system retrieved was the correct person.

- Q = Total number of queries
- mathbb{1} = Indicator function (returns 1 if the condition is true, 0 if false)
- rank_i = The rank position of the first correct match for query i

Rank-k = (Count of queries where correct match is within top k results) / (Total number of queries)

B. Mean Average Precision (mAP)

While Rank-k only cares about finding the *first* correct match, mAP measures the overall quality of the retrieval, ensuring *all* correct matches from different cameras are ranked as high as possible.

First, calculate Average Precision (AP) for a single query:

- N = Total number of retrieved items in the gallery
- P(k) = Precision at cutoff k in the list
- rel(k) = 1 if the item at rank k is a relevant match, 0 otherwise

AP = SUM_over_k(Precision_at_k * is_relevant_at_k) / Total_Relevant_Matches_for_Query mAP = SUM(AP_for_each_query) / Total_Number_of_Queries

V. CONCLUSION AND FUTURE WORK

This paper presented a deep learning-based person re-identification system for public safety in Indian railway stations. The proposed multi-branch architecture combining global and local features with attention mechanisms achieves high accuracy and real-time performance. Future work will expand the dataset to cover more stations and seasonal variations, incorporate video-based Re-ID to leverage temporal information, and integrate facial recognition and anomaly detection modules for a comprehensive public safety platform.

References

[1] L. Zheng et al., "Scalable Person Re-identification: A Benchmark," ICCV, 2015.

[2] Y. Sun et al., "Beyond Part Models: Person Retrieval with Refined Part Pooling," ECCV, 2018.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

[3] G. Wang et al., "Learning Discriminative Features with Multiple Granularities for Person Re-Identification," ACM MM, 2018.
[4] A. Hermans, L. Beyer, and B. Leibe, "In Defense of the Triplet Loss for Person Re-Identification," arXiv:1703.07737, 2017.
[5] H. Luo et al., "Bag of Tricks and A Strong Baseline for Deep Person Re-Identification," CVPR Workshops, 2019.
[6] Z. Zhong et al., "Random Erasing Data Augmentation," AAAI, 2020.
[7] C. Su et al., "Pose-Driven Deep Convolutional Model for Person Re-Identification," ICCV, 2017.