IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 4, August 2025, pp. 2724~2733
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i4.pp2724-2733

Impact of batch size on stability in novel re-identification model

Mossaab Idrissi Alami, Abderrahmane Ez-Zahout, Fouzia Omary
Department of Computer Sciences, Faculty of Sciences, Mohammed V University, Rabat, Morocco

Article Info

Article history:

Received Sep 27, 2024
Revised Jun 15, 2025
Accepted Jul 10, 2025

Keywords:

Batch size
Convolutional neural network
Deep learning
Machine learning
Person re-identification

ABSTRACT

This research introduces ConvReID-Net, a custom convolutional neural network (CNN) developed for person re-identification (Re-ID) focusing on the batch size dynamics and their effect on training stability. The model architecture consists of three convolutional layers, each followed by batch normalization, dropout, and max-pooling layers for regularization and feature extraction. The final layers include flattened and dense layers, optimizing the extracted features for classification. Evaluated over 50 epochs using early stopping, the network was trained on augmented image data to enhance robustness. The study specifically examines the influence of batch size on model performance, with batch size 64 yielding the best balance between validation accuracy (96.68%) and loss (0.1962). Smaller (batch size 32) and larger (batch size 128) configurations resulted in less stable performance, underscoring the importance of selecting an optimal batch size. These findings demonstrate ConvReID-Net's potential for real-world Re-ID applications, especially in video surveillance systems. Future work will focus on further hyperparameter tuning and model improvements to enhance training efficiency and stability.

Corresponding Author:

Mossaab Idrissi Alami
Department of Computer Sciences, Faculty of Sciences, Mohammed V University
B.P. 1014 RP, 4 Avenue Ibn Battouta, Rabat, Morocco
Email: mossaab_idrissialami@um5.ac.ma

1. INTRODUCTION

All video surveillance systems function through four stages: detection, tracking, profile analysis, and re-identification (Re-ID). Detection entails recognizing the existence of humans or things inside a surveilled region, forming the basis for subsequent analysis [1], [2]. Upon detection, the tracking phase guarantees ongoing surveillance of the identified subjects as they traverse various scenes or cameras, ensuring consistent observation throughout time. Profile analysis collects pertinent attributes including appearance, behavior, and biometric information, facilitating comprehensive characterization. Ultimately, Re-ID is essential for correlating persons across several cameras or locations, a vital function for ensuring sustained surveillance continuity. These stages collaborate to improve the efficacy and precision of contemporary video surveillance systems, especially in situations involving densely populated areas, many camera configurations, or intricate security activities [3].

In recent years, the Re-ID problem in surveillance systems has attracted considerable attention due to its wide applications in various fields such as law enforcement, security, and traffic management [4]. Re-ID is the identification of people across different cameras, which is crucial for tracking and monitoring people in large-scale surveillance systems [5]. This task is challenging due to variations in lighting conditions, poses, and occlusions, and therefore. Developing robust and accurate Re-ID models is crucial to overcoming these complexities and improving the performance of surveillance systems.

Traditional Re-ID techniques have depended on manually crafted characteristics, like color histograms and texture descriptors. Nevertheless, these manually crafted characteristics frequently fail to encapsulate intricate patterns and changes within photos. Deep learning methodologies, especially convolutional neural networks (CNNs), have transformed the domain by providing a superior mechanism for feature extraction and recognition [2], [5], [6]. Notwithstanding these developments, attaining consistently precise and resilient Re-ID models continues to pose a considerable challenge, particularly when implemented across heterogeneous datasets and differing monitoring contexts [5].

A significant difficulty that remains inadequately examined is the influence of batch size on training stability and model efficacy in Re-ID tasks [7]. Although batch size is recognized for its impact on convergence speed, model generalization, and training efficiency in deep learning, its function in stabilizing Re-ID models has not been thoroughly examined [8]. In Re-ID systems, where datasets are often big and diverse, training stability becomes a vital aspect in ensuring that models do not overfit or underperform when exposed to new data. Choosing the optimal batch size helps alleviate overfitting, accelerate convergence, and augment the model's generalization ability across various camera perspectives, lighting scenarios, and occlusions.

This study investigates a critical yet underexplored aspect of video surveillance systems: the effect of batch size on training stability and model performance in people (Re-ID) tasks. While prior research has extensively examined key components of video surveillance-such as detection, tracking, and profile analysis limited attention has been given to the specific influence of batch size on Re-ID model convergence and generalization. Previous studies primarily focused on the impact of deep learning techniques like CNNs on feature extraction and recognition. However, the role of batch size in optimizing Re-ID models, especially under diverse and complex monitoring scenarios, remains inadequately addressed. This paper fills this gap by offering a detailed examination of batch size effects, aiming to provide actionable insights for selecting batch sizes that improve model stability and efficiency in real-world surveillance systems.

In this paper, we present a new CNN model for human Re-ID, designed with convolutional layers, batch normalization, and dropout to optimize feature extraction and improve generalization. Using the Market-1501 dataset, we assess the model's performance across various camera angles. A key focus is the analysis of how different batch sizes impact training stability and model performance, offering clear guidelines for selecting batch sizes to balance convergence speed and generalization. These insights contribute to enhancing the reliability and scalability of Re-ID systems for real-world surveillance.

2. RELATED WORK

Person Re-ID is a significant feature of computer vision, notably in applications connected to surveillance and security. The progression of Re-ID methodologies has transitioned from conventional procedures, which predominantly depended on manually generated features and metric learning, to more sophisticated deep learning strategies. Traditional approaches encountered difficulty in addressing variations in posture, light, and occlusion, generally adopting color histograms and texture descriptors that lacked discriminative strength in complex settings. The emergence of large-scale datasets like Market-1501 has been a turning point, enabling the creation of strong deep learning models capable of learning hierarchical feature representations directly from data. This transition has led in considerable breakthroughs in the discipline, with diverse techniques tackling key concerns such as intra-class variability and inter-class discrimination. The subsequent Table 1 lists noteworthy contributions to the Re-ID area, highlighting their methodology, issues addressed, outcomes attained, and assessment criteria employed.

The research detailed in the Table 1 demonstrates the many ways tried to solve the issues of person Re-ID. While early research focused on handmade features, deep learning models have proven greater performance by automatically learning more robust feature representations. Notable developments, such as the use of triplet loss and part-based models, have considerably improved the management of complicated challenges including spatial misalignment, occlusion, and position fluctuations. Additionally, recent research has studied the impact of batch size on model performance, highlighting the difficult balance between generalization and training efficiency. Collectively, these contributions have laid a strong foundation for future developments in Re-ID systems, notably in surveillance and security applications.

Table 1. Summary of key methods and outcomes in re-ID and related learning techniques

Reference	Method	Issue	Outcome	Evaluation metric (%)
[9]	Cyclical learning rate schedule	Trade-offs in batch size effects	Integrated benefits of small and large batch sizes through adaptive learning rate adjustment	-
[10]	Bag-of-Words Descriptor on Market-1501 dataset	Limited dataset scalability in prior work	Introduced the Market-1501 dataset, establishing a competitive baseline for large-scale ReID	Rank-1: 34.4 mAP: 14.09
[11]	Deep Residual Learning (ResNet)	Degradation problem in deep networks	Allowed the construction of much deeper networks without performance loss	Rank-1: 85.4 mAP: 69.4
[12]	Triplet loss-based learning	Difficulty distinguishing between similar and dissimilar images	Encouraged robust feature learning by using anchor-positive-negative triplet samples	Rank-1: 95.0 mAP: 84.4
[13]	Refined part pooling with a strong convolutional baseline	Pose variation challenge	Improved performance on Market-1501 by focusing on body parts	Rank-1: 92.7 mAP: 80.8
[14]	Part-Aligned Bilinear Representations	Spatial misalignment of body parts in different images	Enhanced feature representation through bilinear pooling	Rank-1: 91.2 mAP: 79.5
[15]	Pose-guided feature alignment for occluded ReID	Occlusion and pose variation challenges	Utilized pose estimation for better handling of occlusions	Rank-1: 85.0 mAP: 64.5
[16]	Quality-aware networks	Low-quality images	Improved ReID performance by assessing and incorporating image quality	Rank-1: 90.6 mAP: 77.6
[17]	Improved algorithm for online signature verification	Challenges in real-time feature extraction	Demonstrated the versatility of deep learning in signature verification	Accuracy: 94.0
[18]	Face recognition in smart home systems using CNNs	Security and automation in smart environments	Practical applications of ReID and face recognition in smart home settings	Accuracy: 93.5
[19]	Analysis of batch size and generalization performance	Large batch sizes lead to poor generalization	Proposed "sharp minima" concept for large batch training	-
[20]	Study of super-convergence with large-batch training	Generalization gap in large-batch training	Found that mixed approaches help reduce generalization gap	-
[21]	Relationship between batch size and learning rate	Necessity for larger learning rates with larger batches	Established the link between batch size and learning rate for effective training	-
[22]	Study on the influence of batch size in metric learning	Need for careful control of learning rates with lower batch sizes	Identified improved generalization with lower batch sizes while managing learning rates	-

3. METHOD

This section presents the proposed CNN model for person Re-ID using the Market-1501 dataset. It outlines the dataset characteristics and the data preprocessing techniques applied. Additionally, it details the model architecture, training procedures, and evaluation metrics used to assess performance.

3.1. Dataset

The Market-1501 dataset is used for training and evaluating the proposed Re-ID model. This dataset contains 32,668 annotated bounding boxes of 1,501 identities captured from six different cameras. The dataset is divided into a training set with 12,936 images of 751 identities and a testing set with 19,732 images of 750 identities [2].

3.2. Model architecture

The suggested ConvReID-Net (CNN) architecture for individual Re-ID is constructed with many convolutional and pooling layers, augmented by batch normalization and dropout, to guarantee reliable and effective feature extraction and classification. The model architecture is engineered to utilize extensive feature learning and generalization, mitigating overfitting by implementing dropout layers, while ensuring training stability via batch normalization. Table 2 and Figure 1 comprehensively describe the architecture.

The model initiates with an input layer that accommodates images of dimensions (224, 224, 3). The initial convolutional layer utilizes 32 filters of dimensions (3, 3) with a stride of 1 and 'same' padding. This converts the input into a feature map with dimensions (222, 222, 32). The output dimensions can be determined using the given (1):

$$\theta = \left(\frac{I-K+2P}{S}\right) + 1 \tag{1}$$

where I is the input size, K is the kernel size, P is the padding, and S is the stride. Here, O=(222, 222, 32).

Table 2. Architecture of the proposed CNN model (ConvReID-Net)

Layer	Kernel size	Activation	Output image shape	Param #
Input image	-	-	(224, 224, 3)	-
Conv2D layer 1	(3,3)	ReLU	(222, 222, 32)	896
BatchNormalization	-	-	(222, 222, 32)	128
MaxPooling2D layer 1	(2,2)	-	(111, 111, 32)	0
Dropout	-	-	(111, 111, 32)	0
Conv2D layer 2	(3,3)	ReLU	(109, 109, 64)	18,496
BatchNormalization	-	-	(109, 109, 64)	256
MaxPooling2D layer 2	(2,2)	-	(54, 54, 64)	0
Dropout	-	-	(54, 54, 64)	0
Conv2D layer 3	(3,3)	ReLU	(52, 52, 128)	73,856
BatchNormalization	-	-	(52, 52, 128)	512
MaxPooling2D layer 3	(2,2)	-	(26, 26, 128)	0
Dropout	-	-	(26, 26, 128)	0
Flatten layer	-	-	(86528)	0
Dense (1st fully connected layer)	-	ReLU	(128)	11,075,712
BatchNormalization	-	-	(128)	512
Dropout layer	-	-	(128)	0
Dense (Output layer)	-	SoftMax	(8)	1,032
Total parameters	-	-	-	11,171,400

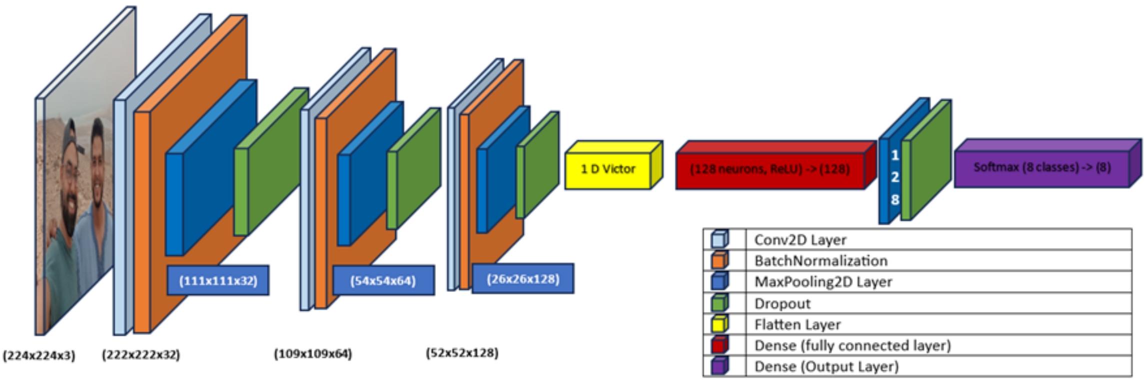


Figure 1. An overview of ConvReID-Net model

A BatchNormalization layer is included after the initial convolutional layer to normalize the activations and enhance stability during training. Following this, a MaxPooling2D layer with a pool size of (2, 2) decreases the spatial dimensions by half, providing an output shape of (111, 111, 32). A dropout layer with a rate of 0.2 is introduced to reduce overfitting by randomly deactivating particular neurons. The second convolutional layer applies 64 filters of size (3, 3), resulting in an output shape of (109, 109, 64). Similar to the preceding block, batch normalization is applied, followed by a MaxPooling2D layer that reduces the output size to (54, 54, 64). Another dropout layer is added with a rate of 0.2 to maintain generalization. The third convolutional layer utilizes 128 filters of size (3, 3), changing the input into an output of shape (52, 52, 128). After batch normalizing and MaxPooling2D, the spatial dimensions are decreased to (26, 26, 128). Again, a dropout layer with a rate of 0.3 is added.

After the last convolution layer, the model flattens the output into a 1D vector of size 86,528 so that the information can be passed on to fully connected layers. The first dense layer has 128 units, where the transformation can be stated by (2):

$$y = W.x + b \tag{2}$$

Here, W and b are the learned weights and biases, and x is the flattened input vector. A BatchNormalization layer and a dropout layer (with a rate of 0.3) follow, helping the network retain generalization and prevent overfitting.

The final output layer consists of a fully connected (dense) layer with the number of units equal to the number of identities in the dataset (8 classes for Market-1501). The SoftMax activation function (3) is employed in this layer to turn the logits into a probability distribution, calculated as (3) [17]:

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}} \tag{3}$$

The model is created using the Nadam optimizer with a learning rate of 0.0001 and categorical cross entropy as the loss function. The training procedure uses accuracy as an evaluation parameter so that performance can be tracked in both the training and validation phases. This design ensures successful feature extraction and classification while limiting the hazards of overfitting, benefiting from dropout and batch normalization at every stage.

In this study, we leveraged a suite of powerful deep learning and machine learning libraries to build and fine-tune the ConvReID-Net model. The architecture was constructed using Keras, where we employed layers such as Conv2D, MaxPooling2D, flatten, dense, dropout, and BatchNormalization to enhance feature extraction and improve generalization. The model was optimized with the Nadam optimizer, which combines the benefits of both Adam and root mean square propagation (RMSProp) optimizers. Data augmentation was handled using ImageDataGenerator, allowing the model to train on varied transformations of the input images. To prevent overfitting and improve learning, callbacks such as EarlyStopping, ReduceLROnPlateau, and LearningRateScheduler were applied. Additionally, l2 regularization and Max-Norm constraints were integrated to further stabilize training. To address class imbalances, class weights were computed using compute_class_weight from Scikit-learn. NumPy and os modules were used for data management, while Matplotlib facilitated the visualization of training and validation metrics throughout the experiments. This comprehensive toolset enabled the successful analysis of batch size impact on the model's performance.

3.3. Convolutional neural network

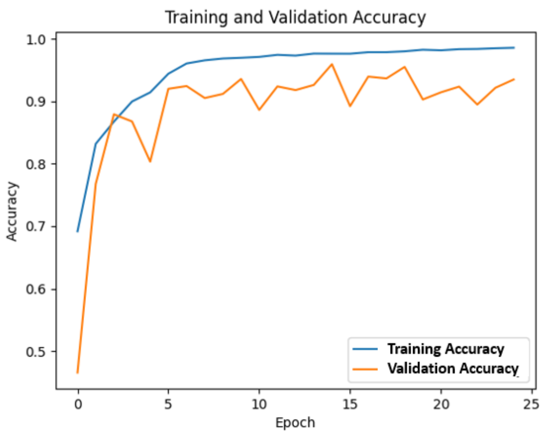
CNNs are among the most popular and successful deep learning systems, particularly for computer vision problems [9]. It’s a specialized type of neural network with a unique topology inspired by biological research [23]. First introduced by Fukushima in 1998, CNNs have a wide range of applications, including activity recognition, phrase classification, biometrics, text recognition, object detection and localization, and analysis of scanned documents. CNNs consist of neurons, each of which has an adaptive weight that adjusts based on the data being processed [24]. The network consists of a single input layer, a single output layer and several hidden layers, which include convolutional layers, pooling layers, fully connected layers and various normalization layers [25]. Our CNN model built using the Keras library and the Python programming language [23].

3.4. Data processing

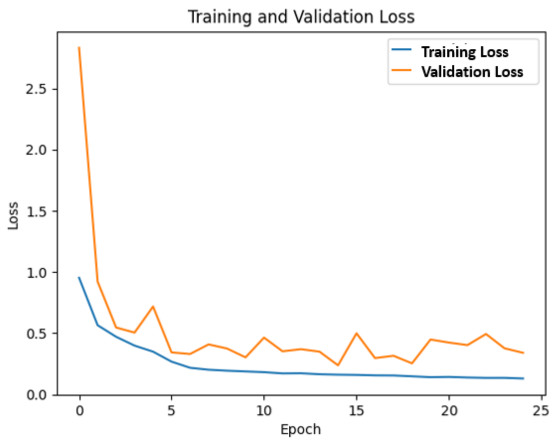
Efficient data preparation is critical for optimizing the performance of our CNN-based person Re-ID model, which employs the Market-1501 dataset. The preprocessing processes include scaling photos to 224×224 pixels to maintain consistency across inputs and permit effective batch processing. Normalization is subsequently applied, scaling pixel values from 0 to 1 by dividing by 255, enhancing the model's stability and convergence during training. To promote generalization and prevent overfitting, data augmentation techniques such as horizontal flip, random crop, rotation, and zoom are used to artificially enlarge the dataset by making variations of the original images. Additionally, mixing the dataset before each epoch guarantees that the model does not learn undesired patterns from the data order, enhancing gradient descent optimization. The preset training set is further separated into training and validation subsets to allow hyperparameter adjustment and early stopping, ensuring the model generalizes well without overfitting. Finally, label encoding transforms person IDs into one-hot vectors, aligning with the SoftMax activation function in the CNN's output layer, which outputs probabilities for each class. This complete preprocessing pipeline ensures the data is well-prepared for training, enhancing both model performance and robustness to fluctuations in real-world circumstances.

3.5. Training procedure

The training approach for the proposed model is carefully developed to provide optimal performance and generalization. Figure 2 shows the learning behavior with batch size=32. Data preparation is handled using Keras' `ImageDataGenerator` Figures 2(a) and 2(b) class, which permits real-time data augmentation during training. Techniques like as horizontal flipping, rotation, and zooming are utilized to boost the model's capacity to generalize by providing variability and preventing overfitting. Training photos are loaded from the `bounding_box_train` directory with these augmentations, while the validation images from the `bounding_box_test` directory remain unaugmented, giving a consistent benchmark for evaluation.



(a)



(b)

Figure 2. Learning behavior with batch size=32 of (a) training and validation accuracy across epochs and (b) loss curve progression

The model is trained over 50 epochs, with the batch size varied progressively throughout the training process to study its impact on model performance and stability. Initially, a batch size of 32 is used, followed by raising the batch size to 64 and subsequently to 128 in later phases. Each batch size divides the dataset into smaller sections for training, allowing the model to process the data efficiently. The number of steps per epoch is dynamically determined by dividing the total number of training samples by the current batch size, ensuring that each epoch processes the whole dataset.

After each epoch, the model's accuracy and loss are evaluated on the validation set to monitor generalization to unseen data. This evaluation directs future tweaks to hyperparameters and helps minimize overfitting. By continuously increasing the batch size, we intend to increase training stability and convergence speed. Early stopping is performed to cease training if the model's performance plateaus on the validation set, providing robust generalization and preventing overfitting during the training process.

3.6. Evaluation metrics

The performance of the proposed model is tested using two key metrics: accuracy and loss. Accuracy shows the proportion of properly identified individuals in the person Re-ID challenge, acting as a direct indicator of the model's capacity to discern between distinct identities. Loss, specifically triplet loss in this context, measures the model's success in minimizing the distance between embeddings of the same identity while maximizing the distance between those of different identities. Throughout the training process, both training and validation accuracy and loss are tracked and charted throughout each epoch, providing a clear overview of the model's learning progression and generalization capabilities. Upon completion of training, the final model is preserved for future usage in real-world Re-ID tasks, confirming its preparedness for deployment in practical applications.

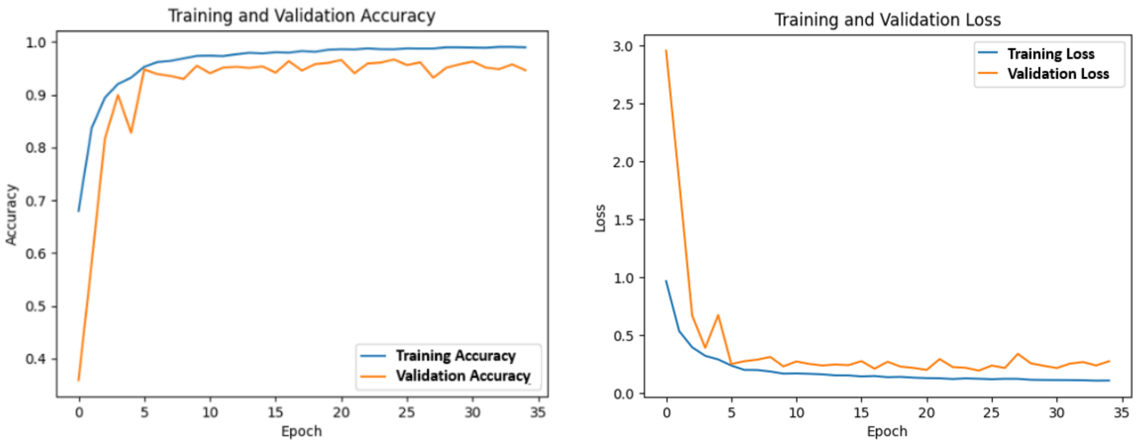
4. RESULTS AND DISCUSSION

4.1. Results

We found that batch size significantly impacts model performance and stability in training the ConvReID-Net. Figure 3 shows the model convergence with batch size=64. As shown in Table 3, batch size 64 correlates with the highest validation accuracy of 96.68% (Figure 3(a)) and the lowest validation loss of 0.1962 (Figure 3(b)), indicating optimal convergence. Batch size 32 also showed strong performance, achieving a validation accuracy of 95.92% (Figure 2(a)) and a validation loss of 0.2371 (Figure 3(b)), demonstrating effective learning but slightly less stability compared to batch size 64. In contrast, Figure 4 shows the learning stability with batch size=128, batch size 128 showed poorer generalization with a higher validation loss of 0.4156 (Figure 4(a)) and a lower validation accuracy of 90.03% (Figure 4(b)). This suggests that batch size 64 allows for more efficient learning and smoother convergence, while larger batch sizes such as 128 tend to result in noisier gradient updates and hinder the model's ability to generalize well to new data.

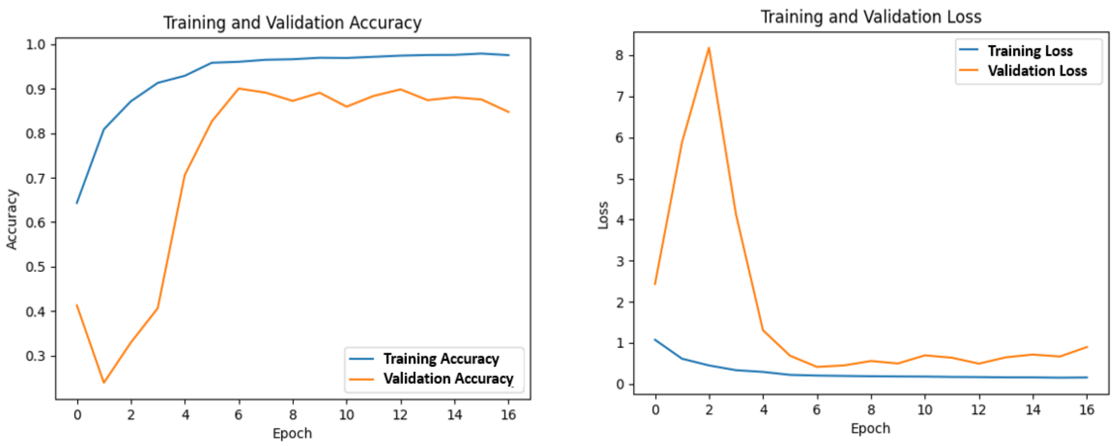
Table 3. Impact of batch size on validation accuracy, loss, and model generalization

Batch size	Validation accuracy	Validation loss	Training stability	Generalization
32	95.92	0.2371	Moderate stability with fluctuations in validation loss	Slight overfitting
64	96.68	0.1962	Stable: consistent learning and convergence	Optimal generalization
128	90.03	0.4156	Unstable: large swings in accuracy and loss	Underfitting



(a) (b)

Figure 3. Model convergence with batch size=64 of (a) accuracy trends during training and validation and (b) loss dynamics per epoch



(a) (b)

Figure 4. Learning stability with batch size=128 of (a) accuracy evolution for training and validation sets and (b) corresponding loss patterns

4.2. Discussion

In our study, we observed that batch size 32, while capable of achieving reasonable performance with a validation accuracy of 95.92% and a loss of 0.2371, had limitations in terms of training stability and convergence speed. These findings align with other studies, such as those by Novack *et al.* [26], which reported that smaller batch sizes tend to introduce more noise in gradient updates, slowing the model's ability to stabilize and leading to higher variations in loss and accuracy. However, as suggested by Lin *et al.* [27], this increased noise can also promote better exploration of the loss surface, potentially preventing overfitting in simpler tasks.

Batch size 64 yielded the best trade-off between convergence speed and generalization, achieving the highest validation accuracy of 96.68% and the lowest loss of 0.1962. This supports previous research indicating that medium-sized batches allow more frequent updates while still capturing diverse patterns in the data, as suggested by Oyedotun *et al*. [28], and it's highest than result's given in Table 1 [10]–[12], [26]. Larger batch sizes, such as 128, though computationally efficient, demonstrated severe instability with a validation accuracy of 90.03% and higher loss (0.4156), corroborating studies by Hoffer *et al*. [22], which found that large batches tend to smooth out gradients excessively, reducing the model's ability to capture subtle patterns and generalize well.

One limitation of our study is that we only explored batch sizes of 32, 64, and 128. It would be beneficial to examine a broader range of batch sizes, especially smaller ones, as done by Hoffer *et al*. [22], to see if further improvements can be made in generalization. Additionally, the impact of early stopping, while useful in preventing overfitting, may have caused the model to miss out on additional learning opportunities, particularly with smaller batch sizes. Future research could explore the use of dynamic batch sizing, where the size is adjusted during training based on model performance, as well as investigating the effects of other optimizers and learning rate schedules to further improve stability and generalization. This would build on the resilience of batch size 64 seen in our study, potentially offering a more robust method for people Re-ID tasks.

5. CONCLUSION

This study assessed the influence of batch size on the ConvReID-Net model's performance for people Re-ID, confirming that batch size has a crucial impact on training stability, convergence speed, and generalization capacity. We found that a batch size of 64 yielded optimal results, achieving a high validation accuracy of 96.68% and the lowest validation loss of 0.1962, reflecting a balanced model performance. In contrast, smaller (batch size 32) and larger (batch size 128) batches resulted in suboptimal outcomes, with larger batches showing signs of overfitting and limited generalization. These results provide strong evidence that careful tuning of batch size is critical for improving model efficiency and robustness, particularly in complex scenarios like video surveillance systems. Future research should explore batch size adjustments with different datasets and architectures, investigate adaptive batch-sizing methods, and examine model enhancements such as attention mechanisms or triplet loss strategies. Such developments could significantly enhance the practical application of Re-ID models in security systems.

FUNDING INFORMATION

This research received no external funding.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mossaab Idrissi Alami	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Abderrahmane Ez-Zahout				✓	✓			✓	✓	✓	✓	✓		
Fouzia Omary			✓		✓		✓			✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/pengcw1/market-1501/data.

REFERENCES

[1] A. Ezzahoutz, H. M. Youssef, and R. O. H. Thami, "Detection evaluation and testing region incoming people's in a simple camera view," in *Second International Conference on the Innovative Computing Technology (INTECH 2012)*, 2012, pp. 179–183, doi: 10.1109/INTECH.2012.6457804.

[2] M. I. Alami, A. Ez-Zahout, and F. Omary, "Enhanced people re-identification in cctv surveillance using deep learning: a framework for real-world applications," *Informatics and Automation*, vol. 24, no. 2, pp. 583–603, 2025, doi: 10.15622/ia.24.2.8.

[3] A. Ezzahout, Y. Hadi, and R. O. H. Thami, "Performance evaluation of mobile person detection and area entry tests through a one-view camera," *Journal of Information Organization*, vol. 2, no. 3, pp. 135–143, 2012.

[4] M. Bukhari, S. Yasmin, S. Naz, M. Maqsood, J. Rew, and S. Rho, "Language and vision based person re-identification for surveillance systems using deep learning with LIP layers," *Image and Vision Computing*, vol. 132, 2023, doi: 10.1016/j.imavis.2023.104658.

[5] Y. X. Peng, Y. Li, and W. S. Zheng, "Revisiting person re-identification by camera selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 5, pp. 2692–2708, 2024, doi: 10.1109/TPAMI.2023.3324374.

[6] M. I. Alami, A. Ez-Zahout, and F. Omary, "Comparative study of person re-identification techniques based on deep learning models," *Informatics and Automation*, vol. 24, no. 3, pp. 982–1001, 2025, doi: 10.15622/ia.24.3.9.

[7] S. L. Smith, P. J. Kindermans, C. Ying, and Q. V. Le, "Don't decay the learning rate, increase the batch size," in *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 2018, pp. 1–11.

[8] N. S. Keskar, J. Nocedal, P. T. P. Tang, D. Mudigere, and M. Smelyanskiy, "On large-batch training for deep learning: Generalization gap and sharp minima," in *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, 2017, pp. 1–16.

[9] N. M. Vasanth and R. Pandian, "Fast region based convolutional neural network ResNet-50 model for on tree Mango fruit yield estimation," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 2, pp. 1084–1091, 2024, doi: 10.11591/ijeecs.v33.i2.pp1084-1091.

[10] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1116–1124, doi: 10.1109/ICCV.2015.133.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[12] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," *arXiv-Computer Science*, pp. 1–17, 2017.

[13] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline)," in *Computer Vision – ECCV 2018*, Cham, Switzerland: Springer, 2018, pp. 501–518, doi: 10.1007/978-3-030-01225-0_30.

[14] Y. Suh, J. Wang, S. Tang, T. Mei, and K. M. Lee, "Part-aligned bilinear representations for person re-identification," in *Computer Vision – ECCV 2018*, Cham, Switzerland: Springer, 2018, pp. 418–437, doi: 10.1007/978-3-030-01264-9_25.

[15] J. Miao, Y. Wu, P. Liu, Y. Ding, and Y. Yang, "Pose-guided feature alignment for occluded person re-identification," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 542–551, doi: 10.1109/ICCV.2019.00063.

[16] Y. Liu, J. Yan, and W. Ouyang, "Quality aware network for set to set recognition," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4694–4703, doi: 10.1109/CVPR.2017.499.

[17] M. Leghari, S. Memon, L. Das Dhomeja, A. H. Jalbani, and A. A. Chandio, "Online signature verification using deep learning based aggregated convolutional feature representation," *Journal of Intelligent and Fuzzy Systems*, vol. 43, no. 2, pp. 2005–2013, 2022, doi: 10.3233/JIFS-219300.

[18] G. Firmasyah, J. Joniwan, A. M. Widodo, and B. Tjahjono, "Preventing child kidnaping at home using cctv that utilizes face recognition with you only look once (YOLO) algorithm," *Journal of Social Research*, vol. 2, no. 9, pp. 3291–3304, 2023, doi: 10.55324/josr.v2i9.1403.

[19] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.

[20] D. R. Wilson and T. R. Martinez, "The general inefficiency of batch training for gradient descent learning," *Neural Networks*, vol. 16, no. 10, pp. 1429–1451, 2003, doi: 10.1016/S0893-6080(03)00138-2.

[21] L. N. Smith and N. Topin, "Super-convergence: very fast training of neural networks using large learning rates," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, May 2019, p. 36, doi: 10.1117/12.2520589.

[22] E. Hoffer, I. Hubara, and D. Soudry, "Train longer, generalize better: Closing the generalization gap in large batch training of neural networks," *arXiv-Statistics*, pp. 1–15, 2018.

[23] K. Khunratchasana and T. Treenuntharath, "Thai digit handwriting image classification with convolution neuron networks," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 1, pp. 110–117, 2022, doi: 10.11591/ijeecs.v27.i1.pp110-117.

[24] J. Wang, X. Song, L. Sun, W. Huang, and J. Wang, "A novel cubic convolutional neural network for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 4133–4148, 2020, doi: 10.1109/JSTARS.2020.3008949.

[25] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," *Progress in Artificial Intelligence*, vol. 9, no. 2, pp. 85–112, 2020, doi: 10.1007/s13748-019-00203-0.

[26] Z. Novack, S. Kaur, T. Marwah, S. Garg, and Z. Lipton, "Disentangling the mechanisms behind implicit regularization in SGD," *arXiv-Computer Science*, pp. 1–15, 2023.

[27] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *arXiv-Computer Science*, pp. 1–10, 2018.

[28] O. K. Oyedotun, K. Papadopoulos, and D. Aouada, "A new perspective for understanding generalization gap of deep neural networks trained with large batch sizes," *Applied Intelligence*, vol. 53, no. 12, pp. 15621–15637, 2023, doi: 10.1007/s10489-022-04230-8.

BIOGRAPHIES OF AUTHORS

Mossaab Idrissi Alami is Ph.D. Student in Faculty of Sciences, Mohammed V University in Rabat. Holds a Master's degree in Big Data Engineering from the Faculty of Sciences, Mohammed V University, Rabat, Morocco (2019). Obtained a Bachelor's degree in Information Systems Administration from the same faculty in 2013, and a Senior Technician Certificate (BTS) in Computer Engineering from the Technical High School in Fez, Morocco (2010). He can be contacted at email: mossaab_idrissialami@um5.ac.ma.

Prof. Abderrahmane Ez-Zahout is currently an Hability Professor of Computer Science at Department of Computer Science, Faculty of Sciences at Mohammed V University. His researches are in fields of computer sciences, digital systems, big data, and computer vision. Recently, he works on intelligent systems. He has served as invited reviewer for many journals. Besides, he is also involved in NGOs, student associations, and managing non-profit foundation. He can be contacted at email: a.ezzahout@um5r.ac.ma.

Prof. Fouzia Omary is the leader IPSS team research and currently a full Professor of Computer Science at Department of Computer Science, Faculty of Sciences, Mohammed V University. She is a full-time researcher at Mohammed V University. She is the chair of the international conference on cyber-security and blockchain and also the leader of the national network of blockchain and cryptocurrency. She can be contacted at email: f.omary@um5r.ac.ma.