

Similarity Learning for Person Re-Identification Using Deep Auto-Encoder

Sevdenur Kutuk¹^a, Rayan Abri¹^b, Sara Abri¹^c and Salih Cetin²^d

¹Mavinci, Ankara, Turkey

²Mavinci, Reading, U.K.

Keywords: Person Re-Identification, Deep Auto-Encoder, YOLOV4, Normalized Cross-Correlation, Cosine Similarity.


Abstract: Person re-identification (ReID) has been one of the most crucial issues in computer vision, particularly for reasons of security and privacy. Person re-identification generally aims to create a unique identity for a person seen in the field of view of a camera and to identify the same person in different frames of the same camera or within the relevant frames of multiple cameras. Due to low resolution and noisy frames, crowded scenes, scenes with occlusion, weather, and light changes, and data sets with insufficient numbers of samples containing different states of the same person for training supervised models, person re-identification remains a challenging and studied problem. In this paper, we propose a hybrid person re-identification model that uses Normalized Cross-Correlation (NCC) and cosine similarity to determine whether extracted features belong to the same person, which we call DAE-ID (Deep Auto-Encoder Identification). The model is built using a pre-trained You Only Look Once Version 4 (YOLOV4) algorithm to detect objects and a convolutional auto-encoder trained on the Motion Analysis and Re-identification Set (Mars) data set for feature extraction. Our method outperforms state-of-the-art methods while outperforming them on the Chinese University of Hong Kong (CUHK03) with 0.966 rank-1 and 0.857 mAP and Duke Multi-Tracking Multi-Camera Re-Identification (DukeMTMC-reID) with 0.956 rank-1 and 0.841 mAP for single-person re-identification.


1 INTRODUCTION


With the rapid development of camera surveillance systems and the rising demand for security, the use of cameras in public and private areas for surveillance and security has increased. Numerous images and videos are captured from various angles and time intervals using cameras in numerous locations, including bus stations, airports, socializing areas, streets, campuses, and government and military facilities. Re-identification (ReID) of individuals is one of the proposed uses of the captured images and videos in this context. ReID has many uses in social security and surveillance systems. As evidenced by the authors' research in (Zheng et al., 2016), it is possible, with the aid of security cameras, to locate criminal suspects, search for a missing child in a large shopping mall, etc., which can be useful.


Person re-identification involves learning the distinguishing characteristics of people in order to determine whether or not they are the same object by tracking the related object or objects on consecutive images. People can appear on multiple cameras in multiple regions in the real world. As the view, exposure, lighting, and resolution of different cameras change, the difficulty of learning to recognize the distinctive features of people increases (Ming et al., 2022). With the development of deep learning, a person's gait, the texture and color of their apparel, etc., can be used to re-identify them. Despite the fact that it has been the subject of numerous studies to date, re-identification remains a significant issue in many areas that are still being investigated.

In this paper, a new method is proposed for human tracking using the convolutional auto-encoder and YOLOV4 (Bochkovskiy et al., 2020). In our proposed method named DAE-ID (Deep Auto-Encoder Identification), real-time object detection is performed using YOLOV4, followed by the selection of the reference object in the desired frame using the mouse. By utilizing the encoder portion of the convolutional

^a <https://orcid.org/0000-0001-7524-8690>

^b <https://orcid.org/0000-0002-2787-2832>

^c <https://orcid.org/0000-0001-6637-9787>

^d <https://orcid.org/0000-0002-9501-7192>

auto-encoder trained on the Mars data set for the selected reference object and other objects in the frame, a code specific to and defining that object is generated. Every 10 frames, the code of the reference object is updated. To continue detecting the reference object in other frames or with different cameras, the reference object's code is compared to the codes of other objects using normalized cross-correlation (NCC) and cosine similarity. The object code with the highest degree of similarity to the reference object's code is considered to belong to the same object. DAE-ID was evaluated using the CUHK03 (Li et al., 2014) and DukeMTMC-reID (Ristani et al., 2016) data sets.

The main contributions of this work can be summarized as follows:

- 1) The DAE-ID system has demonstrated effectiveness in surmounting several challenges, including adverse weather conditions, fluctuations in lighting, and physical barriers. Despite the fact that the chosen reference may temporarily deviate from the established framework and then reenter, it remains possible to monitor and trace the said reference.

- 2) The real-time functionality of the proposed lightweight model allows for seamless operation when a single individual is chosen as the reference object.

Therefore, our study holds significant importance for activities that necessitate confidentiality and privacy, surpassing alternative ReID models in terms of performance outcomes.

2 RELATED WORK

This section provides a comprehensive overview of previous and modern ReID methods, encompassing both classic approaches and those based on deep learning techniques.

2.1 Traditional Methods

In old methods of solving human ReID problems, information such as color, texture, spatial structure, edges, and contours that can be derived based on the general appearance of the individual may contain lower-level representations, and the received representations are then used together. Utilizing conventional similarity measurement techniques and matching algorithms, such as robust similarity measures (Kostinger et al., 2012; Liao et al., 2015), to redefine the resulting representations. Several methods based on histograms and keypoint detection were devised in this study. To locate the person with the same ID in the other image for the person's ReID, the similar-

ity between the person's facial features must be determined. Traditional measurement techniques include the Euclidean distance (Dokmanic et al., 2015), the cosine distance, and the Mahalanobis distance (Lee et al., 2018).

Although these traditional hand-crafted features do not necessitate complex destruction, they are readily affected by complex backgrounds such as changes in light and occlusion, which causes features to become obscured and individuals to be indistinguishable. In addition, the implementation of traditional methods in the actual world is difficult and costly. Due to this, traditional ReID system characteristics cannot be universally supported and further developed for a variety of reasons.

2.2 Deep Learning-Based Methods

Recent years have seen an explosion in the use of deep learning techniques in computer vision, particularly for object recognition and detection. With the widespread use of deep learning, significant progress has been made in the problem of person ReID that we focused on in this study, and successful results have been obtained in many studies by removing the limitations of traditional methods mentioned in the previous section (Wu et al., 2019). In this section, we provide an overview and introduction to these methods.

Tian et al. (Tian et al., 2018), observing that current deep learning models are biased by capturing an excessive amount of relevance between background views of person images, they devised a series of experiments using newly created data sets to validate the effect of background information. To solve the problem of background bias, they proposed a person-region-guided deep neural network based on human decomposition maps to learn more distinguishable person-part characteristics and augmented the training data with person images with random backgrounds. Chen et al. (Chen et al., 2019), proposed an Attentive Yet Diverse Network (ABD-Net) that integrates attention modules and diversity regulation throughout the entire network in order to learn more representative, robust, and distinctive characteristics. In this study, the authors introduced a pair of complementary attentional modules focusing on channel coupling and position awareness, as well as a new efficient orthogonality constraint to enforce orthogonality on both latent activations and weights. Xia et al. (Xia et al., 2019), proposed a novel mechanism for attention that directly models long-term relationships using quadratic trait statistics. Zheng et al. (Zheng et al., 2019), propose a framework for collaborative learning that integrates end-to-end iden-

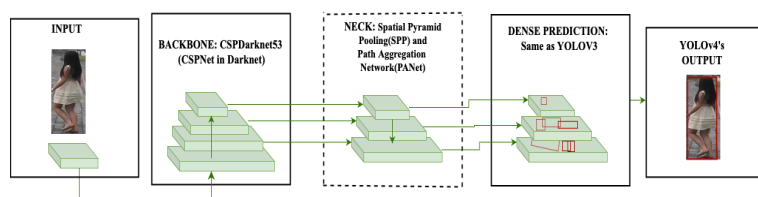


Figure 1: Overview of the YOLOV4 architecture used for human detection.

tity relearning with data generation. It consists of a generative module that encodes each individual into a view-code and a build-code, as well as a distinguishing module that shares the viewcoder with the generative module. Without using generated data, the proposed framework is a significant improvement over the baseline. Ren et al. (Ren et al., 2021), which suppresses in-class variations with a new extensible, lightweight, trained without additional supervision, Jensen-Shannon triple loss for comparative distribution learning in ReID. He suggests the auto-encoder HAVANA. The authors claim that HAVANA is the first VAE-based framework for self-reported ReID. Bilakeri et al. (Bilakeri and Kotegar, 2022), use an auto-encoder module to generate a single image at three different scales to increase sample size and overcome the issue of scale variation in person redefinition, and with initial and auto-encoder to demonstrate the impact of auto-encoder as a data increment for person redefinition. To combat noise and occlusion, Sezavar et al. (Sezavar et al., 2023), created a system based on noise-sensitive and deep convolutional neural networks by training an auto-encoder on artificially damaged frames. Although many of these methods based on deep learning have been more successful than traditional methods and have overcome difficulties in general, they still have not been adequate for solving the ReID issue. In addition, many of these studies are difficult to integrate, cumbersome, and do not operate in real time.

3 PROPOSED METHOD

In this paper, a new method is proposed for human tracking using the convolutional auto-encoder and YOLOV4, which we call DAE-ID. In the DAE-ID, the reference object is selected by users to be followed from among the individuals identified by the YOLOV4 algorithm. Then, using the encoder portion of the auto-encoder model trained on the Mars data set, the code of the reference object is extracted and saved following this selection. In certain frames, the code of the reference object is updated. Using this model, the codes of all objects in the new frame are extracted. The object with the highest degree of simi-

larity to the registered reference object is accepted as the same object, and id assignment is performed. Figure 2 depicts the overall structure of the model. This section elaborates on the YOLOV4 algorithm, reference object selection, Auto-Encoder model, similarity algorithms, and model integration.

3.1 Object Detection and Selection of Reference Image with YOLOV4

Using the YOLOV4 model, objects (only people) were identified in each frame. Figure 1 depicts the YOLOV4 model's architecture. This project is essentially one of the subproblems of a larger study that addresses object detection (Abri et al., 2020b; Abri et al., 2020a; Mansoub et al., 2019; Abri et al., 2021), tracking, motion detection and background learning (Ince et al., 2022). Since YOLOV4 is used for other subproblems in this overall project, it is a C++-based project, and YOLOV4 has attained success in object detection, there has been no comparison with other new YOLO models. After using YOLOV4 to detect people, the user is prompted to select a reference object by clicking on the object to be followed among the detected objects. Consequently, the object selected by the user becomes the reference object. The reference object is clipped from the relevant frame and sent to the auto-encoder model for decoding based on the limits determined by the YOLOV4 algorithm for the selected reference object. This process is performed every 10 frames in order to keep the reference object current, and every 10 frames, the current status of the user-selected reference image is updated. The selection of 10 update frequencies is intended to strike a balance between the project's pace and accuracy. Other objects within the frame are clipped in the same manner and sent to the auto-encoder model for comparison with the reference object's code. Figure 3 depicts selecting and updating the reference object.

3.2 Deep Auto-Encoder Structure

An auto-encoder (Hinton and Salakhutdinov, 2006), is a specialized form of an advanced artificial neu-

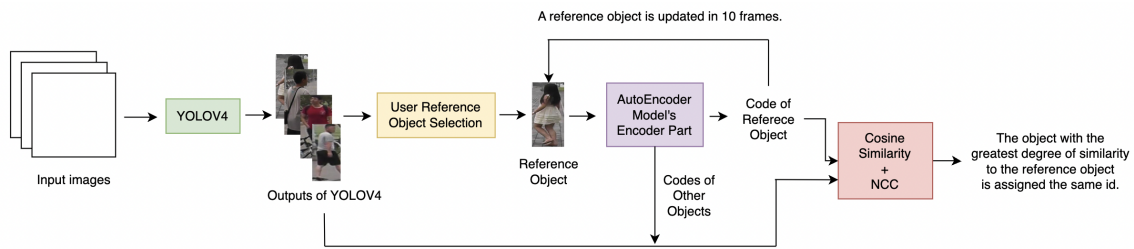


Figure 2: The overall structure of the model.

ral network used to transfer the input to the output, i.e., reduce or compress the data's size and resolve it. Typically, auto-encoder architecture consists of completely interconnected layers. As the processes in the study involved images, the convolutional auto-encoder, a type of auto-encoder commonly used in image processing applications, was employed. The convolutional auto-encoder, unlike the auto-encoder, performs the feature extraction process using convolutional layers and codes the input images' properties more effectively.

Figure 4 depict, respectively, the model of a convolutional auto-encoder used in this study. The utilized model is a convolutional auto-encoder model with eight layers, including the input layer, encoder layers, and output layer. In this model, the input images are regarded as the first layer and are then utilized on the encoder side of the three convolutional layers and the three max pooling layers. The entry images of the study's progressive auto-encoder model are the Mars training set converted to 32 x 32 size and gray scale. Input images are converted to image tensors. The first convolution has 64 filters, while the other two only have 32. The pooling layers of 2 x 2 masks minimize the output of the evolutionary layer and allow the characteristics to be moved deeper. Evolutionary layers enable the detection of various characteristics by sliding the filters across the image. This process permits data compression by creating a representation of the input data with fewer dimensions. The pooling layer, the final layer on the encoder side, represents the compressed data. This representation is the output of the encoder network and is a compressed image tensor.

The compressed representation is then restored to its original dimensions using three deconvolutional layers and three unpooling layers on the decoder side. Unlike coder layers, decoder layers use deconvolutional and up-sampling (unpooling) layers to reconstitute the image. The deconvolution layers help restore the original size of the summarized data, while the upsampling layers increase the output size of the summarizer layers. In the final layer, the final image is generated. The output of the decoder layers is a

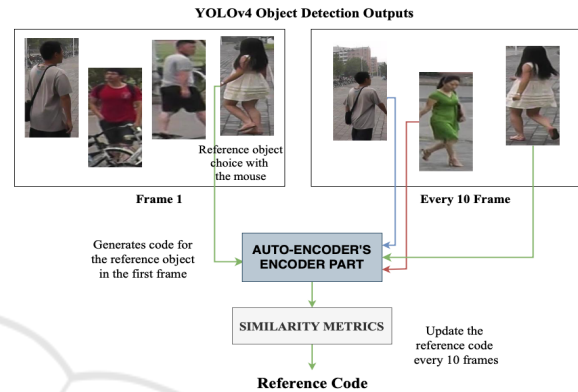


Figure 3: Using YOLOV4 to select an object and updating the reference image.

reconstruction of the original data input. This layer's architecture typically consists of a convolutional layer and can be viewed as an image tensor with the original dimensions of the reconstructed image.

In all auto-encoder structures, Leaky ReLu (Maas et al., 2013), served as the activation function. Leaky ReLU also offers a smooth derivative for negative inputs, thereby preventing situations in which the ReLU function is over-reset (the dead ReLU problem), and is a well-known activation function that produces superior results, particularly in large data sets and complex models; thus, it was chosen for this study.

The model's loss was computed using the mean squared error (MSE). The MSE is an error metric that measures the disparity between a machine learning model's predictions and the actual values and is the average of the squared differences between the actual and predicted values. The lower the MSE, the closer the model's predictions are to the actual values, and the more effective the model is deemed to be. MSE is frequently preferred in auto-encoder models because the input and output data are the same size (typically image data), making error calculation simple.

As a means of regularization, an L1 Regularization ($10e-5$) (Tibshirani, 1996), was used. The L1 regularization restricts the weights' magnitudes to the sum of their absolute values. This can result in some weights being close to or equal to zero, simplify-

ing the model's representation, and preventing overfitting. Additionally, the L1 regularization can be used for feature selection, i.e., it can simplify the model's representation by bringing the weights of unimportant features close to zero. L1 regularization reduces the number of parameters used in training the network; as a result, it can enhance the network's generalization.

Adaptive Moment Estimation (Adam) (Kingma and Ba, 2014), was used as the optimizer. Optimization is the adjustment of the parameters of mathematical models to best optimize a target function. During the training of a machine learning model, it is crucial to optimize the model's parameters for a specific objective. This can improve the model's results and generalizability.

This auto-encoder model's encoder component is utilized. Using the encoder component of the auto-encoder model, the code of each object, which was detected by the YOLOV4 algorithm and transmitted by clipping the borders, was generated.

3.3 Auto-Encoder Training and Implementation

The Mars data set was used for the training of the auto-encoder. After performing 500 epoch training as specified, other stages are started. The encoder part of the trained weight is integrated into the model using CppFlow.

3.4 Similarity Metrics

NCC and cosine similarity were utilized to determine the degree of similarity between two object. In the auto-encoder model, the similarity of the code parts calculated for the detected objects was calculated using both similarity calculation methods and applied in the same manner. NCC and cosine similarity are used together because they produce superior results by compensating for one another's shortcomings.

Shown in Equation 1, NCC is frequently used to locate specific image content in another image and pinpoint the exact location of the image content, and it performs well even in challenging conditions where the brightness of the image changes or there is noise. Indicated by Equation 2, Cosine similarity measures the relationship between two vectors in an inner product space. The cosine of the angle between two vectors reveals whether or not they point in roughly the same direction.

$$R(x, y) = \frac{\sum_{x', y'} (I(x+x', y+y') - \bar{I}(x, y)) \cdot (t(x', y') - \bar{t})}{\sqrt{\sum_{x', y'} (I(x+x', y+y') - \bar{I}(x, y))^2 \cdot \sum_{x', y'} (t(x', y') - \bar{t})^2}} \quad (1)$$

$$\sin(A, B) = \cos(\Theta) = \frac{A \cdot B}{\|A\| \|B\|} \quad (2)$$

3.5 Identification Number Assignment by Similarity Metrics

The reference image is the cropped image of the user-selected object, while the compared images are cropped images of other objects in the current frame. The process of ID assignment consisted of repeating the following steps, shown in Figure 5:

Step 1: Compares the code portion of each image to be compared with the code of the reference image using cosine similarity and NCC.

Step 2: In the second step, each pair's cosine similarity is multiplied by 0.4, and the NCC values are multiplied by 0.6.

Step 3: The pair of images with the highest value when compared to the reference image are considered to be images of the same object.

At this stage, the matching ratios of the reference image and other images on Mars's test data, the CUHK03 and DukeMTMC-reID data sets, calculated using NCC and cosine similarity, are discussed in the 3.4 section.

4 EXPERIMENTS

The DAE-ID was evaluated using the CUHK03 and Duke MTMC-reID data sets. For both sets of data, random samples were drawn from random classes, and their similarity to other classes and samples within those classes was analyzed. Model performance was determined based on whether the class information of the most similar samples was identical. Cumulative matching characteristic (CMC) rank-1 accuracy, and Mean Average Precision (mAP) are used as evaluation criteria.

4.1 Data Sets

This section describes the Motion Analysis and Re-identification Set (Mars), Chinese University of Hong Kong (CUHK03), and Duke Multi-Tracking Multi-Camera Re-Identification (DukeMTMC-reID) data sets. The Mars data set was used to train the auto-encoder, while the CUHK03 and DukeMTMC-reID data sets were utilized to validate the model.

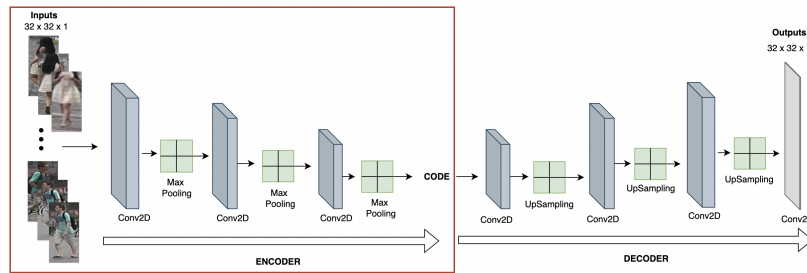


Figure 4: Structure of the convolutional auto-encoder model used for reid.

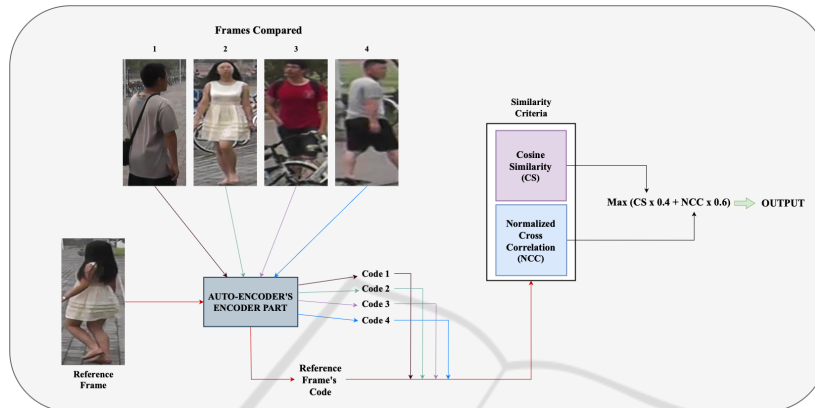


Figure 5: Identification Number Assignment by NCC and Cosine Similarity.

4.1.1 Motion Analysis and Re-Identification Set (Mars)

The Mars data set is a data set used for motion detection, tracking, and re-identification and an extension of the Market-1501 data set (Tsfaye et al., 2017). It consists of approximately 8,000 to 20,000 images of 1,261 unique individuals wearing different clothing, positioned at different angles, and containing different body parts. Poor image quality and variations in the poses, colors, and illuminations of pedestrians make it difficult to achieve high matching accuracy. Figure 6 depicts a selection of images from the Mars data set. The auto-encoder tracking model was trained using this data set. For training, the entire set of training and test data was utilized.

4.1.2 Chinese University of Hong Kong (CUHK03) Data Set

The CUHK03, consists of 14,097 photographs of 1,467 unique individuals; six campus cameras were



Figure 6: Samples from the Mars data set (1, 11, 127, and 459).

installed for image collection, and two campus cameras were used to photograph each individual.

4.1.3 Duke Multi-Tracking Multi-Camera Re-Identification (DukeMTMC-reID) Data Set

Duke Multi-Tracking Multi-Camera Re-Identification (Duke MTMC-reID) is an image-based re-identification data set. It consists of 16,522 training images for 702 identifiers, 2,228 query images for the remaining 702 identifiers, and 17,661 gallery images.

4.2 Experiments Results

The model developed in this study was evaluated using Rank-1 and mAP metrics on the CUHK03 and DukeMTMC-reID data sets. Due to the fact that the calculations using these metrics were validated using random samples from both data sets, the model was executed ten times and the mean was calculated. For both data sets, the outcomes of CS alone, NCC alone, and CS and NCC together were analyzed. 0.218 thresh was chosen when only CS was used, 0.75 thresh was chosen when only NCC was used, and 0.65 thresh was chosen when both NCC and CS were used. As part of a more comprehensive analysis, the effect of using CS alone, NCC alone, or a combination of

Table 1: Comparison of the model with different similarity measures for DukeMTMC-reID and CUHK03.

	DukeMTMC-reID Data Set			CUHK03 Data Set		
	CS	NCC	CS and NCC	CS	NCC	CS and NCC
Rank-1	0.677	0.924	0.956	0.175	0.942	0.966
mAP	0.173	0.724	0.841	0.012	0.757	0.857

Table 2: Comparison of the DAE-ID to other models utilizing the DukeMTMC-reID and CUHK03 data sets.

Models	DukeMTMC-reID Data set		CUHK03 Data set	
	Rank-1	mAP	Rank-1	mAP
SONA (Xia et al., 2019)	89.4	78.3	-	-
Abd-Net (Chen et al., 2019)	89.0	78.6	-	-
DG-Net (Zheng et al., 2019)	86.6	74.8	-	-
HAVANA (Ren et al., 2021)	89.4	80.8	-	-
Auto-Encoder for Scale-Invariant (Bilakeri and Kotegar, 2022)	85.5	74.1	-	-
Deep CNN and Auto-Encoders (Sezavar et al., 2023)	-	-	94.4	-
Background-Bias (Tian et al., 2018)	-	-	92.5	-
DAE-ID	95.6	84.1	96.6	85.7

CS and NCC on the DukeMTMC-ReID data set was highlighted. The DAE-ID can operate in real time at 18-21 frames per second. Moreover, the outputs of DAE-ID were compared to those of other models utilizing the same data sets. Also the values associated with the places denoted by "-" have not been calculated.

CS and NCC achieved the best results in the CUHK03 data set with 0.966 rank-1 and 0.857 mAP, and in the DukeMTMC-reID data set with 0.956 rank-1 and 0.841 mAP, as shown in Table 1.

The comparison of DAE-ID to other models is displayed in Table 2. Empty fields in the table indicate that the model was not validated against the specified data set. According to Table 2, the tests we conducted with rank-1 and mAP on both data sets, using NCC and CS in conjunction, produced superior results compared to other models. The base model is the model we propose without consideration for similarity.

5 CONCLUSION

It is a challenging problem in computer vision to produce a unique identifier for the same person in different cameras or different frames of the same camera, as well as to detect the same person again under different lighting, occlusion, re-entering the camera, weather, and resolution conditions. It can be utilized in a variety of contexts, such as the disappearance of anyone, the possibility of guilt, or the pursuit of criminals, and it is essential for maintaining security and public order. Years of research have been devoted to finding a solution to the ReID problem, particularly in the fields of security and defense. Although the stud-

ies carried out so far have provided general success in the face of the various difficulties mentioned, there is still no real-time model that provides sufficient success in detecting the person as the same person when leaving the frame and re-entering.

In this article, we propose a real-time model for single-person re-identification that uses a hybrid YOLOV4 and convolutional auto-encoder model to generate a unique identity for the selected person and uses cosine similarity and normalized cross-correlation to calculate the similarity of the selected person with other people, which we call DAE-ID. We explain why cosine similarity and normalized cross-correlation are employed together for similarity measurement. In single-person re-identification, DAE-ID achieved 18-21 frames per second, a performance that could be utilized in real time. On the CUHK03 and DukeMTMC-reID data sets, we compared DAE-ID to others using rank-1 and mAP metrics and found that DAE-ID performed better.

ACKNOWLEDGEMENTS

This research is supported by Mavinci in UK, an R&D company in information and communication technologies, security and defense areas with the capability of software development, artificial intelligence, and machine learning.

REFERENCES

- Abri, R., Abri, S., Yarici, A., and Çetin, S. (2020a). Multi-thread approach to object detection using yolov3. *2020 Joint 9th International Conference on Informat-*

- ics, *Electronics & Vision (ICIEV) and 2020 4th International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, pages 1–6. IEEE.
- Abri, S., Abri, R., and Çetin, S. (2021). An analytical comparison of approaches to real-time object detection to handle concurrent surveillance video streams. In *2021 6th International Conference on Frontiers of Signal Processing (ICFSP)*, pages 43–47. IEEE.
- Abri, S., Abri, R., Yarıcı, A., and Çetin, S. (2020b). Multi-thread frame tiling model in concurrent real-time object detection for resources optimization in yolov3. In *Proceedings of the 2020 6th International Conference on Computer and Technology Applications*, pages 69–73.
- Bilakeri, S. and Kotegar, K. (2022). Strong baseline with auto-encoder for scale-invariant person re-identification.
- Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.
- Chen, T., Ding, S., Xie, J., Yuan, Y., Chen, W., Yang, Y., Ren, Z., and Wang, Z. (2019). Abd-net: Attentive but diverse person re-identification. volume 2019-October.
- Dokmanic, I., Parhizkar, R., Ranieri, J., and Vetterli, M. (2015). Euclidean distance matrices: essential theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 32(6):12–30.
- Hinton, G. E. and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507.
- Ince, E., Kutuk, S., Abri, R., Abri, S., and Cetin, S. (2022). A light weight approach for real-time background subtraction in camera surveillance systems. In *2022 IEEE 5th International Conference on Image Processing Applications and Systems (IPAS)*, pages 1–6. IEEE.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kostinger, M., Hirzer, M., Wohlhart, P., Roth, P. M., and Bischof, H. (2012). Large scale metric learning from equivalence constraints.
- Lee, K., Lee, K., Lee, H., and Shin, J. (2018). A simple unified framework for detecting out-of-distribution samples and adversarial attacks. *Advances in neural information processing systems*, 31.
- Li, W., Zhao, R., Xiao, T., and Wang, X. (2014). Deep-reid: Deep filter pairing neural network for person re-identification.
- Liao, S., Hu, Y., Zhu, X., and Li, S. Z. (2015). Person re-identification by local maximal occurrence representation and metric learning. volume 07-12-June-2015.
- Maas, A. L., Hannun, A. Y., Ng, A. Y., et al. (2013). Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3. Atlanta, Georgia, USA.
- Mansoub, S. K., Abri, R., and Yarıcı, A. (2019). Concurrent real-time object detection on multiple live streams using optimization cpu and gpu resources in yolov3. *SIGNAL*, pages 23–28.
- Ming, Z., Zhu, M., Wang, X., Zhu, J., Cheng, J., Gao, C., Yang, Y., and Wei, X. (2022). Deep learning-based person re-identification methods: A survey and outlook of recent works. *Image and Vision Computing*, 119:104394.
- Ren, J., Ma, X., Xu, C., Zhao, H., and Yi, S. (2021). Havana: Hierarchical and variation-normalized autoencoder for person re-identification.
- Ristani, E., Solera, F., Zou, R., Cucchiara, R., and Tomasi, C. (2016). Performance measures and a data set for multi-target, multi-camera tracking. In *Computer Vision—ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8–10 and 15–16, 2016, Proceedings, Part II*, pages 17–35. Springer.
- Sezavar, A., Farsi, H., and Mohamadzadeh, S. (2023). A new model for person reidentification using deep cnn and autoencoders. *Iranian (Iranica) Journal of Energy & Environment*, 14(4):314–320.
- Tesfaye, Y. T., Zemene, E., Prati, A., Pelillo, M., and Shah, M. (2017). Multi-target tracking in multiple non-overlapping cameras using constrained dominant sets. *arXiv preprint arXiv:1706.06196*.
- Tian, M., Yi, S., Li, H., Li, S., Zhang, X., Shi, J., Yan, J., and Wang, X. (2018). Eliminating background-bias for robust person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5794–5803.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.
- Wu, D., Zheng, S.-J., Zhang, X.-P., Yuan, C.-A., Cheng, F., Zhao, Y., Lin, Y.-J., Zhao, Z.-Q., Jiang, Y.-L., and Huang, D.-S. (2019). Deep learning-based methods for person re-identification: A comprehensive review. *Neurocomputing*, 337:354–371.
- Xia, B. N., Gong, Y., Zhang, Y., and Poellabauer, C. (2019). Second-order non-local attention networks for person re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3760–3769.
- Zheng, L., Yang, Y., and Hauptmann, A. G. (2016). Person re-identification: Past, present and future. *arXiv preprint arXiv:1610.02984*.
- Zheng, Z., Yang, X., Yu, Z., Zheng, L., Yang, Y., and Kautz, J. (2019). Joint discriminative and generative learning for person re-identification. volume 2019-June.