An Efficient Person Re-Identification Model Based on New Regularization Technique

MARWAN ALI ALBAHAR1, ABDULAZIZ ALI ALBAHR2,3, AND MUHAMMAD HATIM BINSAWAD4
1College of Computer Science in Al-Leith, Umm Al Qura University, Mecca 21421, Saudi Arabia
2College of Applied Medical Sciences, King Saud Bin Abdulaziz University for Health Sciences, Al-Ahsa 31982, Saudi Arabia
3King Abdullah International Medical Research Center, Al-Ahsa 36276, Saudi Arabia
4Department of Computer Information Systems, King Abdulaziz University, Jeddah 21589, Saudi Arabia
Corresponding author: Marwan Ali Albahar (marwanalialbahar@gmail.com)

This work was supported by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under Grant D-378-611-1441.

ABSTRACT The aim of person re-identification (ReID) is to recognize the same persons across different scenes. Due to the many demanding applications that utilize large-scale data, more and more attention has been devoted to matching efficiency and accuracy. Many methods that are based on binary coding have been presented to reach efficient ReID. Those methods learn projections to map the high-dimensional features into deep neural networks or compact binary codes through simple insertion of an extra fully connected layer with tanh-like activation. Nevertheless, the former approach needs hand-crafted feature extraction that also wastes a lot of time and complex (discrete) optimizations. In contrast, the latter approach lacks the essential discriminative information to a large extent because of the straightforward activation functions. A ReID framework is proposed in the current work, and it is inspired by the adversarial framework depending on the new regularization approach (ABC-NReg). We embedded the discriminative network into adversarial binary coding (ABC) with our new regularization, which improved the discriminative power combined with the triplet network. ABC-NReg and triplet networks were optimized, and three large-scale benchmark datasets, namely CUHK03, Market-1501, and DukeMTMC-reID datasets, were utilized to test the performance of our proposed model. We further compared the simulation results with the present hashing and non-hashing algorithms. Our model provided better results than other present models using the Market-1501 and DukeMTMC-reID datasets when considering Rank-1. For CUHK03 dataset, the proposed model exceeded the performance of other works when considering Rank 5 and Rank 20.

INDEX TERMS Person re-identification, domain adaptation, novel regularizer.

I. INTRODUCTION With one or more pictures of a pedestrian, person re-identification (ReID) recognizes the person with the same identity across multiple images taken from different angles and viewpoints and in various scenes. ReID assists different potential applications, like long-term cross-view scenario tracking and retrieval of criminals. However, the task is still challenging because of the remarkable differences in poses, viewpoints, and illuminations through various cameras. A lot of ReID methods have been presented [1]–[6]. Most of them acquire high-dimensional (usually thousands or more) features for comprehensively representing persons through different clues (e.g., colors, textures, and spatial-temporal clues). Thus, a higher level of computational complexity is brought to the subsequent similarity measurement (e.g., metric learning). Also, recent large-scale ReID benchmarks include many identities and cameras to create a simulation of real-world scenarios, making the state-of-the-art ReID approaches that currently exist computationally unaffordable [7]. Thus, despite the considerable enhancement in matching accuracy, more challenges have faced the computational and memory requirements. Binary coding (i.e., hashing), utilized by [8], [9], map high-dimensional features into compact binary codes and calculate similarities efficiently in the low-dimensional Hamming space. It is considered a potential solution for efficient ReID. The hashing based ReID methods can be classified into two main categories: The first category learns various projection matrices to map the original features concurrently to a low-dimensional and discriminative
Hamming space. Its objective is mainly a non-convex joint function of multiple subtasks (e.g., similarity-preserving mapping and binary transformation), which needs the explicit design of sophisticated functions and time-consuming non-convex or discrete optimizations. Both memory storage and computational efficiency are major and serious matters, particularly when working with large-scale data. The second category demonstrates a deep neural network-based method, which can process large-scale data in a much more efficient way than the traditional methods by utilizing mini-batch learning algorithms and advanced GPUs. Here, the binary codes are obtained through the insertion of hashing layers at the networks’ final part. The hashing layer is a fully-connected layer, after which a tanh-like activation occurs to transform the outputs into binary form. This simple, uncomplicated scheme barely restricts the outputs under the essential principles of hashing (e.g., balancedness and independence [10]) to generate high-quality binary codes. Additionally, to preserve discriminability, the hashing layers’ outputs tend to lie in the approximately linear part of the tanh-like functions. Thus, direct binarization of the outputs through signification will lead to the loss of the discriminative information. This work presents a new regularization-based adversarial framework combined with a triplet network to address the aforementioned issues. The following part summarizes our main contributions in this paper:

- In this work, we introduce a deep adversarial based binary transformation strategy. The proposed model depends on a generative model based on CNN, which extracts features and a discriminative model to differentiate between the binary and real-valued features. The CNN model is trained to produce binary features in order to confuse the discriminator.
- The discriminative model is integrated with the new regularizer in order to regularize the features into binary codes based on the standards deviation of weight values.
- Adversarial binary coding model is incorporated in an end-to-end new regularization based deep neural network for efficient ReID. The binary transformation and similarity measurement are optimized, and as a result, during feature binarization, the discriminative information is maintained to the maximum extent.
- The framework is tested on three large-scale benchmark datasets (CUHK03 [11], Market-1501 [12], and DukeMTMC-reID [13]) which gives comparatively good results.

II. RELATED WORK

Traditional approaches normally present certain feature learning algorithms concerning ReID, such as low-level color features [11], [16], [17], local gradients [4], [18], [19], and high-level features [2], [3], [20]. Because of deep neural networks’ outstanding performance, deep learning-based ReID methods [13], [21], [26] have been presented progressively. For example, siamese CNNs [27], [29] and triplet CNNs [6], [30], [31] are extensively utilized for measuring similarity. Lately, multiple binary coding-based approaches [8], [9], [32], [33] have appeared to deal with the high computation and storage costs related to the ReID problems. Moreover, generative adversarial nets (GANs) [11], [12] present a methodology for random mapping variables from a simple distribution to a certain complex one. GANs have been extensively utilized in image generation [12], [34]–[36], style transferring [37], [38] and latent feature learning [39]–[41].

For stabilizing and quantifying the training of GANs, a breakthrough known as Wasserstein GAN (WGAN) was presented in [42] and enhanced in [43]. Lately, GANs are also used for problems concerning image retrieval. In [44], GANs were employed to differentiate between synthetic and real images, to enhance the discriminability of binary codes. GANs were also utilized to improve the intermediate representation of the generator in [45]. Yet, these studies simply utilized tanh-like activations for binarization. Apart from the mentioned works, there are different deep Re-ID models [46]–[48] that present different network architectures. These differences are mainly determined by training losses/objectives. The authors in [49] presented a framework depending on adversarial binary coding (ABC) for efficient person re-identification, which efficiently produces binary and discriminative features from pedestrian images. After that, a trained discriminator network was employed for differentiating between the real-valued features and the binary ones, so that it can lead the feature extractor network to produce features in binary form under the Wasserstein loss.

III. APPROACH

In the current work, a new GAN model is proposed based on the new regularization. The proposed framework employs the existing GANs model combined with the proposed regularizer to compact the high dimensional real-valued data to binary codes. In the first section, we introduce the new regularization model based on the standard deviation of weight values. Next, we shed light on the principles of GANs. Finally, we describe the integration of the new regularizer with the adversarial binary coding (ABC-NReg).

Generalization Error and Regularization: During the training of a predictive model, the learning model sometimes begins to memorize the data, which results in increasing the generalization error. This causes the model to have a good performance on training data, but the performance drastically deteriorates on unseen or test data. To prevent the model from memorization, the term regularization is introduced. Regularization aims at penalizing the learning model to commence generalization through performing well on unseen data. Different types of regularizers set penalty terms on the predictive model. The most prominently utilized regularization methods are L1 and L2 regularizers.

A. L1 OR LASSO REGULARIZER

L1 regularizer penalizes the absolute values of the weight matrix from attaining higher values. Reducing the weight
where $n$ is the number of features in data, and $w$ is the corresponding weight values of each feature.

### B. L2 OR RIDGE REGULARIZER

Sometimes, there is no need for the weight value to be zero for less important features. Thus, the features of corresponding coefficient values are decreased but kept greater than zero. That’s why the square magnitude values are considered from coefficient values. So, if there are very high pair-wise correlations of a group of variables, both regularizers tend to pick only one variable to represent the corresponding group. To improve the classification model’s predictive accuracy and achieve a more concise solution, we need to think beyond the L1 and L2 regularization techniques. The new regularizer considers the dispersion of weight values. It controls the weight matrix’s spread by applying the standard deviation of the weights to the loss function instead of squared magnitude values and absolute values. Thus, it restrains the learning model from getting widespread values from the weight space. The mathematical notation of this regularizer is given in equation 2. The $\lambda$ parameter is used to put an extra penalty on the values of corresponding weight values. Since the $\lambda$ parameter controls the magnitude of the coefficient values, selecting a good $\lambda$ value is a critical task.

$$\lambda \sum_{i=1}^{n} |w_i|$$

(1)

where $n$ is the number of features in data, and $w$ is the coefficient value of each feature.

### C. NEW REGULARIZER

Unlike L1 and L2 regularization techniques, the penalty disregards the relationship between features and controls only individual weights among weight matrix entries. Hence, they will not be able to explore data with dependent structures. So, if there are very high pair-wise correlations of a group of variables, both regularizers tend to pick only one variable to represent the corresponding group. To improve the classification model’s predictive accuracy and achieve a more concise solution, we need to think beyond the L1 and L2 regularization techniques. The new regularizer considers the dispersion of weight values. It controls the weight matrix’s spread by applying the standard deviation of the weights to the loss function instead of squared magnitude values and absolute values. Thus, it restrains the learning model from getting widespread values from the weight space. The mathematical formulation of the new regularizer is given in equation 3 and 4.

$$\lambda \sum_{i=1}^{n} \sigma (\theta_i)$$

(3)

where $k$ denotes the number of rows in the weight matrix and $\sigma$ denotes standard deviation of weight values as given below

$$\sigma (\theta_i) = \sqrt{\frac{1}{n} \left( \frac{1}{n} \sum_{d=1}^{n} \theta_{d,i}^2 - \frac{1}{n} \left( \sum_{i=1}^{n} \theta_{d,i} \right)^2 \right)}$$

(4)

### FIGURE 1. Contour plots of L1, L2, and new regularizers.

$\theta$ is squared as it is done in standard deviation, $\theta_i$ is the feature vector $i-th$ while $\theta_{d,i}$ is the $d$-th feature in the $i$-the feature vector. Here the parameter $\lambda$ is used to control the values of weight matrix, and $n$ is the weight vector size. So, $n$ relies on the number of features in the dataset. Particularly, the number of columns in a specific weight matrix.

In figure 1, we have shown the contours for the penalty functions of the new regularizer. Here, the penalty is equal to 1 for all regularizers ($L1$, $L2$, and the new). Further, to maintain the dimensions, we have excluded the sum factor in the new regularizer. So, the spread of the regularizer depends upon the penalty term $\lambda$, the spread increases upon decreasing the $\lambda$ and the reverse for increasing the $\lambda$. We observed that the behavior of the $L2$ norm is circular and incorporates $L1$ regularization. While the new regularization acts like a parabola along with other the regularizations. This allows the new regularizer to take values beyond the $L2$ limits. Therefore, it increases the limit of values (space) to adopt and further squeezes and expands this space based on the penalty term $\lambda$.

1) A BRIEF REVIEW OF GANs

In the GANs model, the following two models are trained simultaneously: the generative model $G$ and the discriminative model $D$. The generative model is trained through mapping a prior distribution $p_z(z)$ to a generative space $G(z; \theta_G)$, while the discriminative model $D(x; \theta_D)$ is trained on input $x$, which yields the probability showing whether the input is coming from the real distribution rather than the $G$. Maximizing the probability of $D$, making a mistake is the task of training model $G$. The loss function for the GANs model is shown as:

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)}[\log D(x)] + E_{z \sim P_z(z)}[\log[1 - D(G(z))]]$$

(5)
WGAN is a variant of GANs presented by Arjovsky et al. [42], [50] to overcome the difficulty in the training process of the GANs model. Similarly, for evaluating the distributions’ similarity, the WGANs model optimizes Wasserstein loss in place of the Jensen-Shannon loss. The Wasserstein loss is given as follows:

$$\max_w E_{x \sim p_{data}(x)}[D(x)] - E_{z \sim p_z(z)}[D(G(z))]$$

In the present study, for adversarial learning, we followed the training process of WGANs, whose gradients are computed based on Wasserstein-1 distance, hence providing stronger stability.

2) ADVERSARIAL BINARY CODING COMBINED WITH THE PROPOSED REGULARIZER

Our binary coding method depends on the method presented by Liu et al. [49], who proposed a method called adversarial binary coding inspired by GANs. Our method is different from their method since we utilized our proposed regularization in the discriminative model. A deep neural network is used in learning the mapping of data from the original distribution. So, the input images to a mapping of binary vectors in a GAN model. Here, we explained the binary mapping of the end-to-end efficient extended-ReID framework.

As given in figure 2, we utilized ResNet-50, a CNN based feature extractor, in our proposed architecture of adversarial binary coding. The ResNet-50 is a trained model that is accessible in the PyTorch library [51]. The output of the feature extractor is to transform the input images to feature vectors. For every bit of the binary vector, we simultaneously executed random sampling through a binary code sampler. Regarding sampling, the Bernoulli distribution was utilized because it satisfies the principle of effective binary coding [10] and the different bits depend on each other. Since the binary vector is considered as the positive sample and the real-valued feature vector as the negative sample, the discriminator is trained based on the new regularization (Eq.4) in order to execute the classification task. Likewise, the task of the extractor is to extract feature vectors based on the Wasserstein loss as given in 6.

In GANs, the feature extractor can be considered as a function $f(K)$ which perform the task of a generator $G$ under a encoding distribution $q(Z|K)$. Where $K$ is the bath of $n$ images that can be shown as $K = K_1, K_2, \ldots, K_n$ and $O = o_1, o_2, \ldots, o_n$ is the feature vector extracted by the generator. The aim of $q(O|K)$ is to map data from the original distribution $p_k$ to another distribution $q$:

$$q(o_i) = \int q(o_i|K_i)p(K_i)dK_i, \quad i = 1, \ldots, n.$$ (7)

The Wasserstein distance is utilized to regularize the extractor by matching the posterior $q$ to a prior binomial distribution. A binomial distribution is similar to multiple Bernoulli sampling with the same probability. For the feature extractor, the ResNet-50 model is utilized with Rectified Linear Unit (ReLU) as the activation function. ReLU activation function yields values between 0 and 1, inclusive. Thus, using ReLU helps in representing binary codes by 0,1. Furthermore, the weight of the network is initialized with Gaussian distribution that has low values. Our new regularization prevents weights from gaining higher values; thus, it avoids gradient vanishing, and consequently, the weight obtains values near zero. Moreover, unstable optimization may occur because of the absence of normalization. Therefore, we followed the method of [49] to normalize the output feature vectors and sampled binary codes.

D. ReID FRAMEWORK BASED ON TRIPLET LOSS

In order to measure similarity among the binary codes, to transform features to binary form, and to ensure the discriminant behavior of the learned binary code, we integrated
the new ABC-NReg into a triplet network. The mathematical formulation of triplet loss [52] is given below:

$$L(f_i, f_j, g_k, \alpha) = \frac{1}{n} \sum_{i=1}^{n} \max\{d(f_i, f_j) - d(f_i, g_k) + \alpha, 0\}$$  (8)

where \(f_i, f_j\) and \(g_k\) are the input features, and the imposed distance margin is denoted with \(\alpha\) between negative and positive pairs. A similarity between different features is denoted with the term \(d(.).\) Features from the same class are represented as \(f_i\) and \(f_j\), while \(g_k\) shows features from another class. The triplet loss is used because it forces the distances between positive and negative pairs. In a sense, the positive pairs should be smaller than those in negative pairs.

### IV. EXPERIMENTS

#### A. DATASET AND SETTINGS

In the present study, three large-scale ReID benchmark datasets are utilized as follows.

**CUHK03** dataset comprises 14,096 images with 1,467 identities captured through six surveillance cameras. It includes both manually and automatically labeled bounding boxes with different sizes. All the images are resized to 160 × 60 as conducted by [49], and 20% train/test splits criteria are followed as done by [13]. The number of iterations was set to 6,000, and \(\alpha\) values were set to 0.2 and increased to 0.5 with an increment of 0.1 after every 1,000 iterations.

**Market-1501** dataset comprises 32,668 images with 1,501 pedestrians. The number of iterations was set to 8,000, and \(\alpha\) values were set to 0.2 and increased by 0.1 after 2,000 iterations.

**DukeMTMC-reID** dataset comprises 36,411 manually labeled bounding boxes of 1,401 identities captured through 8 cameras having fixed training and testing split. In our simulations, all images are resized to 128 × 64. The number of iterations was set to 8,000, and \(\alpha\) values were set to 0.2 and updated to 0.3 after iteration 2,000 and to 0.4 after iteration 5,000.

Simulations were performed on Amazon EC2 g3.4xlarge instance. The g3.4xlarge instance contains one NVIDIA Tesla M60 GPU with 16 vCPUs and 122GB of RAM. To augment the training images, we have flipped each of the images horizontally. The batch size is initialized to 128, and the learning rate is set to 10^{-3} and decreased to 10^{-4} along the training process. Similarly, for discriminator, the learning rate is set to 10^{-2}.

#### B. EVALUATION

#### C. COMPARISON WITH BINARY CODING BASED METHODS

In this section, our proposed method is compared with the state-of-the-art hashing methods based on ReID, as shown below:

1) Deep Hashing consisting of Deep Semantic Ranking Based Hashing (DSRH) [32] and Deep Regularized Similarity Comparison Hashing (DRSCH) [33].
2) Non-Deep Hashing comprising Cross-Camera Semantic Binary Transformation (CSBT) [9]. As per the CSBT results, it outperformed the previous method. Therefore, this is the main method to which we compare our results.
3) 512-bit ABC+triplet [49]

Results concerning CUHK03 and Market-1501 datasets are displayed in Tables 1 and 2, respectively. Figures 3 to 7 present the triplet and discriminator loss on the three datasets. The progressions of both losses are similar to those presented in [49]. The slight variations are caused by the new regularization in the discrimination model. In this study, our work is similar to the work carried out by [49]. Thus, we mainly compared our performance with their results. For the CUHK03 dataset, when considering Rank 1, our method’s performance is slightly lower compared to that of [49]. Notably, our results are better when considering Rank 5 and Rank 20. Likewise, mAP value is also slightly high. For Market-1501, as displayed in Table 2, our method provides notably better results compared to [49] because the discrimination power of the discriminator is increased as a result of the new regularization method. In other words, the discriminator is well trained to identify the difference

### TABLE 1. Matching rates, mAP, feature extraction time (F.time), average query time (Q. Time) and memory usage (mega bytes) for storing gallery data on CUHK03 labelled dataset.

| Method      | Rank 1 | Rank 5 | Rank 20 | mAP   | F. Time (s) | Q. Time (s) | Mem. (MB) |
|-------------|--------|--------|---------|-------|-------------|-------------|-----------|
| DRSCH [33]  | 22.0   | 48.4   | 81.0    | -     | -           | -           | -         |
| DSRH [32]   | 14.4   | 43.4   | 79.2    | -     | -           | -           | -         |
| CSBT [9]    | 55.5   | 84.3   | 98.0    | 7.5   | 8.07e-04    | 1.01e-02    |           |
| ABC+triplet [49] | 58.2 | 85.7   | 98.2    | 59.7  | 5.72e-01    | 8.07e-04    | 1.01e-02  |
| Our Method  | 57.9   | 85.9   | 98.3    | 59.8  | 5.72e-01    | 8.07e-04    | 1.01e-02  |

### TABLE 2. Matching rates, mAP, feature extraction time (F.time), average query time (Q. Time) and memory usage (mega bytes) for storing gallery data on Market-1501 dataset.

| Method      | Rank 1 | mAP   | F. Time (s) | Q. Time (s) | Mem. (MB) |
|-------------|--------|-------|-------------|-------------|-----------|
| CSBT [9]    | 42.9   | 20.3  | 2.96e+02    | 2.54e-02    | 1.52e+00  |
| ABC+triplet [49] | 59.3 | 43.8  | 6.0e+00     | 2.54e-02    | 1.52e+00  |
| Our Method  | 60.0   | 44.0  | 6.3e+00     | 2.54e-02    | 1.52e+00  |
TABLE 3. Matching rates, mAP, feature extraction time (s), average query time (s) and memory usage (mega bytes) for storing gallery data, by comparing with state-of-the art non-hashing methods on CUHK03 dataset.

| Method            | Rank 1 | Rank 5 | Rank 20 | mAP  | F. Time(s) | Q. Time (s) | Mem. (MB) |
|-------------------|--------|--------|---------|------|------------|-------------|-----------|
| DeepReID [13]     | 19.9   | 49.8   | 78.2    | -    | -          | -           | -         |
| Improved Deep [21]| 44.9   | 76.4   | 93.6    | -    | -          | -           | -         |
| NSL [58]          | 54.7   | 84.8   | 95.2    | -    | -          | -           | 2.11e+01  |
| Gated CNN [29]    | 61.8   | 80.9   | 51.3    | -    | -          | -           | -         |
| EDM [28]          | 51.9   | 83.6   | -       | -    | -          | 2.85e-02    | 4.90e-01  |
| SIR+CIR [6]       | 52.2   | -      | -       | -    | -          | 1.26e-02    | 5.16e-01  |
| Part-aligned [53] | 81.6   | 97.3   | 99.5    | -    | -          | -           | -         |
| MuDeep [54]       | 75.6   | 94.4   | -       | -    | -          | -           | -         |
| PDC [55]          | 78.3   | 94.8   | 98.4    | -    | -          | -           | -         |
| ABC+triplet [49]  | 68.1   | 90.3   | 98.3    | 61.6 | 5.77e-01   | 1.85e-03    | 1.22e-01  |
| Our Method        | 69.0   | 91.4   | 98.0    | 62.0 | 5.77e-01   | 1.85e-03    | 1.22e-01  |

TABLE 4. Matching rates, mAP, feature extraction time (s), average query time (s) and memory usage (mega bytes) for storing gallery data, by comparing with state-of-the art non-hashing methods on Market-1501 dataset.

| Method         | Rank 1 | mAP  | F. Time(s) | Q. Time (s) | Mem. (MB) |
|----------------|--------|------|------------|-------------|-----------|
| SIDAF [1]      | 20.5   | 8.2  | -          | 2.33e+02    | 9.31e+01  |
| KISSME [16]    | 40.5   | 19.0 | -          | -           | -         |
| sSDC [59]      | 33.5   | 13.5 | 1.58e+04   | 7.47e+02    | 1.45e+03  |
| BoW(best) [14] | 44.4   | 21.9 | -          | 2.08e+00    | 1.5e+01   |
| LOMO+NSL [58]  | 55.4   | 29.9 | 2.96e+02   | -           | 4.06e+03  |
| Gated CNN [29] | 65.9   | 39.6 | -          | -           | 4.32e+01  |
| SpindleNet [56]| 76.9   | -    | -          | -           | -         |
| Part-aligned [53]| 81.0 | 63.4 | -          | -           | -         |
| PDC [55]       | 84.1   | 63.4 | -          | -           | -         |
| ABC+triplet [49]| 73.5   | 52.9 | 6.49e+00   | 7.32e-02    | 4.82e+00  |
| Our Method     | 74.0   | 53.0 | 6.49e+00   | 7.32e-02    | 4.82e+00  |

FIGURE 3. Triplet loss on CUHK03 dataset.

FIGURE 4. Triplet loss on Market-1501 dataset.

D. COMPARISON WITH THE STATE-OF-THE-ART METHODS

Aside from the comparison with the hashing methods, our proposed method was also compared based on the new regularization with the current non-hashing ReID methods. These methods consist of the following algorithms:

1) Deep Learning based methods These methods mainly include ABC triplet [49], DeepReID [13], SIR+CIR [6], Improved Deep [21], EDM [28], Gated CNN [29], Deeply-learned Part-Aligned Representation [53], Multi-scale Deep Learning [54], Pose-Driven Deep Convolutional Model [55], Spindle Net [56] and SVDNet [57].
2) Metric learning based methods include KISSME [16], NSL [58] and XQDA [2].
3) Local patch matching based methods consist of SDC [59] and BoW [14]

All the comparison results for CUHK03 (detected), Market-1501, and DukeMTMC-reID datasets are displayed...
TABLE 5. Matching rates, mAP, feature extraction time (s), average query time (s), and memory usage (mega bytes) for storing gallery data, by comparing with state-of-the-art non-hashing methods on DukeMTMC-reID dataset.

| Method         | Rank 1 | mAP   | F. Time(s) | Q. Time (s) | Mem. (MB) |
|----------------|--------|-------|------------|-------------|-----------|
| Bow+KISSME [14]| 25.1   | 12.1  | -          | -           | -         |
| LOMO+XQDA [2]  | 30.8   | 17.0  | 2.65e+02   | -           | 3.62e+03  |
| Basel+LSRO [15]| 67.7   | 47.1  | -          | -           | -         |
| SVDNet [57]    | 76.7   | 56.8  | -          | -           | -         |
| ABC+triplet [49]| 77.6   | 47.9  | 5.33e+00   | 6.38e-02    | 4.31e+00  |
| Our Method     | 76.7   | 48.0  | 5.33e+00   | 6.38e-02    | 4.31e+00  |

FIGURE 5. Triplet loss on DukeMTMC dataset.

FIGURE 6. Discriminator Loss on CUHK03 dataset.

FIGURE 7. Discriminator loss on Market-1501 dataset.

FIGURE 8. Discriminator Loss on DukeMTMC-reID dataset.

in Tables 3, 4 and 5, respectively. Among all the state-of-the-art methods, our methods, adopting high-dimensional real-valued features, exhibited good performance, and produced good matching accuracy. Our framework exceeds the present non-hashing algorithms in performance as well as in terms of matching accuracy. It is worth noting now that the query time is the same as that of the method presented by [49], which is quite faster than the non-hashing methods on the existing datasets. Likewise, the memory requirements were also decreased by representing images as 2048-bit binary codes rather than representing 26960-dim real-valued features as adopted by Local Maximal Occurrence Representation (LOMO).

V. CONCLUSION

In the present work, a framework is proposed for efficient person re-identification depending on the recent advances in adversarial learning, particularly those inspired by the study of Liu et al. [49]. We integrate the discriminator in the ABC framework with the new regularization depending on the weight matrix’s standard deviation. Furthermore, for improving the discrimination power, the deep triplet network
is embedded in ABC framework. CUHK03, Market-1501, and DukeMTMC-reID datasets were utilized in testing the performance of our proposed model. We further compared the simulation results with the present hashing and non-hashing algorithms. Our model provided better results than other present models using the Market-1501 and DukeMTMC-reID datasets when considering Rank-1. For the CUHK03 dataset, the proposed model exceeded other works’ performance when considering Rank 5 and Rank 20.

ACKNOWLEDGMENT

The authors, therefore, gratefully acknowledge DSR technical and financial support.

REFERENCES

[1] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, “Person re-identification by symmetry-driven accumulation of local features,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Jun. 2010, pp. 2360–2367.

[2] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, “Person re-identification by local maximal occurrence representation and metric learning,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 2197–2206.

[3] T. Matsukawa, T. Okabe, E. Suzukile, and Y. Sato, “Hierarchical Gaussian descriptor for person re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 1363–1372.

[4] K. Liu, B. Ma, W. Zhang, and R. Huang, “A spatio-temporal appearance representation for vico-based pedestrian re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 1288–1296.

[5] L. Zheng, Y. Yang, and A. G. Hauptmann, “Person re-identification: Past, present and future,” 2016, arXiv:1610.02984. [Online]. Available: http://arxiv.org/abs/1610.02984

[6] F. Zheng and L. Shao, “Learning cross-view binary identities for fast person re-identification,” in Proc. IJCAI, 2016, pp. 2399–2406.

[7] J. Chen, Y. Wang, J. Qin, L. Liu, and L. Shao, “Fusion person re-identification via cross-camera semantic binary transformation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 3873–3882.

[8] Y. Weiss, A. Torralba, and R. Fergus, “Spectral hashing,” in Proc. NIPS, 2009, pp. 1753–1760.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 2672–2680.

[10] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” 2015, arXiv:1511.06434. [Online]. Available: http://arxiv.org/abs/1511.06434

[11] W. Li, R. Zhao, T. Xiao, and X. Wang, “DeepReID: Deep filter pairing neural network for person re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jan. 2014, pp. 152–159.

[12] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, “Scalable person re-identification: A benchmark,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 1116–1124.

[13] Z. Zheng, L. Zheng, and Y. Yang, “Unlabeled samples generated by GAN improve the person re-identification baseline in vitro,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 3754–3762.

[14] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, “Large scale metric learning from equivalence constraints,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2012, pp. 2288–2295.

[15] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, “Local Fisher discriminant analysis for pedestrian re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2013, pp. 3318–3325.

[16] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang, “Person re-identification by support vector ranking,” in Proc. Brit. Mach. Vis. Conf., vol. 2, 2010, p. 6.

[17] G. Lisanti, I. Masi, A. D. Bagdanov, and A. D. Bimbo, “Person re-identification by iterative re-weighted sparse ranking,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 8, pp. 1629–1642, Aug. 2015.

[18] R. Lan, Y. Zhou, and Y. Tang, “Quaternionic local ranking binary pattern: A local descriptor of color images,” IEEE Trans. Image Process., vol. 25, no. 2, pp. 566–579, Feb. 2016.

[19] E. Ahmed, M. Jones, and T. K. Marks, “An improved deep learning architecture for person re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 3908–3916.

[20] T. Xiao, H. Li, W. Ouyang, and X. Wang, “Learning deep feature representations with domain guided dropout for person re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 1249–1258.

[21] S. Zhou, J. Wang, J. Wang, Y. Gong, and N. Zheng, “Point to set similarity based deep feature learning for person re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 3741–3750.

[22] J. Lin, L. Ren, J. Lu, J. Feng, and J. Zhou, “Consistent-aware deep learning for person re-identification in a camera network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 7054–7063.

[23] W. Chen, X. Chen, J. Zhang, and K. Huang, “Beyond triplet loss: A deep quadruplet network for person re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 403–412.

[24] D. Yi, Z. Lei, S. Liao, and S. Z. Li, “Deep metric learning for person re-identification,” in Proc. 22nd Int. Conf. Pattern Recognit., Aug. 2014, pp. 34–39.

[25] H. Shi, Y. Yang, X. Zhu, S. Liao, Z. Lei, W. Zheng, and S. Z. Li, “Embedding deep metric for person re-identification: A study against large variations,” in Proc. ECCV, 2016, pp. 732–748.

[26] R. R. Varior, M. Haloi, and G. Wang, “Gated siamese convolutional neural network architecture for human re-identification,” in Proc. ECCV, 2016, pp. 791–808.

[27] S.-Z. Chen, C.-G. Guo, and J.-H. Lai, “Deep ranking for person re-identification via joint representation learning,” IEEE Trans. Image Process., vol. 25, no. 5, pp. 2353–2367, May 2016.

[28] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng, “Person re-identification by multi-channel parts-based CNN with improved triplet loss function,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 1335–1344.

[29] F. Zhao, Y. Huang, L. Wang, and T. Tan, “Deep semantic ranking based hashing for multi-label image retrieval,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 1556–1564.

[30] R. Zhang, L. Lin, R. Zhang, W. Zhang, and L. Zhang, “Bit-scalable deep hashing with regularization for similarity learning for image retrieval and person re-identification,” IEEE Trans. Image Process., vol. 24, no. 12, pp. 4766–4779, Dec. 2015.

[31] X. Huang, Y. Li, O. Poursaeed, J. Hopcroft, and S. Belongie, “Stacked generative adversarial networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 5077–5086.

[32] R. Huang, S. Zhang, T. Li, and R. He, “Beyond face rotation: Global and local perception GAN for photorealistic and identity preserving frontal view synthesis,” 2017, arXiv:1704.04086. [Online]. Available: http://arxiv.org/abs/1704.04086

[33] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, “Least squares generative adversarial networks,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 2794–2802.

[34] I. Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired Image-to-Image translation using cycle-consistent adversarial networks,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 2223–2232.

[35] P. Isola, J.-Y. Zhu, T. Zhao, and R. He, “Beyond face rotation: Global and local perception GAN for photorealistic and identity preserving frontal view synthesis,” 2017, arXiv:1704.04086. [Online]. Available: http://arxiv.org/abs/1704.04086

[36] M. A. Albahar et al.: Efficient Person ReID Model Based on New Regularization Technique

171056 VOLUME 8, 2020
[42] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein GAN,” 2017, arXiv:1701.07875. [Online]. Available: http://arxiv.org/abs/1701.07875

[43] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, “Improved training of wasserstein GANs,” 2017, arXiv:1704.00028. [Online]. Available: http://arxiv.org/abs/1704.00028

[44] Z. Qiu, Y. Pan, T. Yao, and T. Mei, “Deep semantic hashing with generative adversarial networks,” in Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Aug. 2017, pp. 225–234.

[45] J. Song, “Binary generative adversarial networks for image retrieval,” in Proc. AAAI, 2018.

[46] S. Ding, L. Lin, G. Wang, and H. Chao, “Deep feature learning with relative distance comparison for person re-identification,” Pattern Recognit., vol. 48, no. 10, pp. 2993–3003, 2015.

[47] H. Shi, X. Zhu, S. Liao, Z. Lei, Y. Yang, and S. Z. Li, “Constrained learning for person re-identification,” 2015, arXiv:1511.07545. [Online]. Available: https://arxiv.org/abs/1511.07545

[48] E. Ustinova, Y. Ganin, and V. Lempitsky, “Multiregion bilinear convolutional neural networks for person re-identification,” 2015, arXiv:1512.05300. [Online]. Available: https://arxiv.org/abs/1512.05300

[49] Z. Liu, J. Qin, A. Li, Y. Wang, L. Van Gool, “Adversarial binary coding for efficient person re-identification,” in Proc. IEEE Int. Conf. Multimedia Expo (ICME), Jul. 2019, pp. 700–705, doi: 10.1109/ICME.2019.00126.

[50] M. Arjovsky and L. Bottou, “Towards principled methods for training generative adversarial networks,” 2017, arXiv:1701.04862. [Online]. Available: http://arxiv.org/abs/1701.04862

[51] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[52] V. Balntas, E. Riba, D. Ponsa, and K. Mikolajczyk, “Learning local feature descriptors with triplets and shallow convolutional neural networks,” in Proc. Brit. Mach. Vis. Conf., 2016, pp. 1–11.

[53] L. Zhao, X. Li, Y. Zhuang, and J. Wang, “Deeply-learned part-aligned representations for person re-identification,” in Proc. ICCV, Oct. 2017, pp. 3219–3228.

[54] X. Qian, Y. Fu, Y.-G. Jiang, T. Xiang, and X. Xue, “Multi-scale deep learning architectures for person re-identification,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 5399–5408.

[55] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, and Q. Tian, “Pose-driven deep convolutional model for person re-identification,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 3960–3969.

[56] H. Zhao, M. Tian, S. Sun, J. Shao, J. Yan, S. Yi, X. Wang, and X. Tang, “Spindle net: Person re-identification with human body region guided feature decomposition and fusion,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 1077–1085.

[57] Y. Sun, L. Zheng, W. Deng, and S. Wang, “SVDNet for pedestrian retrieval,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 3800–3808.

[58] L. Zhang, T. Xiang, and S. Gong, “Learning a discriminative null space for person re-identification,” in Proc. CVPR, Jun. 2016, pp. 1239–1248.

[59] R. Zhao, W. Ouyang, and X. Wang, “Unsupervised salience learning for person re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2013, pp. 3586–3593.

MARWAN ALI ALBAHAR received the B.S. degree in computer science from King Faisal University, in 2011. the M.Sc. degree (Hons.) in computer science from Frostburg State University, in 2015, and the Ph.D. degree from the University of Eastern Finland, in 2018. His main research interests include computer networks and security, cybersecurity, and artificial intelligence.

ABDULAZIZ ALI ALBAHR is currently an Assistant Professor with the College of Applied Medical Sciences, King Saud Bin Abdulaziz University for Health Sciences. His main research interests include natural language processing and data mining. His research interests include extracting keyphrases, summarizing text, and identifying prerequisite relations from educational data.

MUHAMMAD HATIM BINSAWAD received the master’s degree (applied information technology) and successfully completed his post baccalaureate certificate in information systems management from Towson University, USA, and the Ph.D. degree in information systems from the University of Technology Sydney (UTS), in 2019. He is currently an Assistant Professor with the Department of Computer Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University.