Metrics Space and Norm: Taxonomy to Distance Metrics

Barathi Subramanian, Anand Paul, Jeonghong Kim, and K.-W.-A. Chee

Kyungpook National University, Daegu, Republic of Korea

Correspondence should be addressed to K.-W.-A. Chee; aghjuee@knu.ac.kr

Received 30 April 2022; Revised 17 July 2022; Accepted 19 July 2022; Published 6 October 2022

A lot of machine learning algorithms, including clustering methods such as K-nearest neighbor (KNN), highly depend on the distance metrics to understand the data pattern well and to make the right decision based on the data. In recent years, studies show that distance metrics can significantly improve the performance of the machine learning or deep learning model in clustering, classification, data recovery tasks, etc. In this article, we provide a survey on widely used distance metrics and the challenges associated with this field. The most current studies conducted in this area are commonly influenced by Siamese and triplet networks utilized to make associations between samples while employing mutual weights in deep metric learning (DML). They are successful because of their ability to recognize the relationships among samples that show a similarity. Furthermore, the sampling strategy, suitable distance metric, and network structure are complex and difficult factors for researchers to improve network model performance. So, this article is significant because it is the most recent detailed survey in which these components are comprehensively examined and valued as a whole, evidenced by assessing the numerical findings of the techniques.

1. Introduction

Discovering a good distance metric in feature space is vital in the certifiable application. In recent years, distance metric learning has become apparent as a promising field in machine learning, with applications including medicine [1, 2], security [3, 4], social media mining [5, 6], information retrieval [7–9], recommender systems [10, 11], speech recognition [12, 13] and a diversity of computer vision applications, such as person re-identification [14, 15], kinship verification [16, 17], or image classification [18, 19]. Distance measurements are additionally complex in the classification of images [9]. For example, in the KNN classifier, the key is to recognize the set of labelled pictures that are nearest to a given test picture in the space of visual highlights including the assessment of a distance metric. Past work [20–24] has demonstrated the way that distance measurements can fundamentally help KNN grouping precision contrasted with the standard ED. Mahalanobis distance [25–27] in general is directly addressed in currently available studies.

Increasing data volumes provide significant advantages for more accurate classification, in terms of both volume and accuracy. On the other hand, calculations are becoming ever more complex. To meet many computing needs, it is essential to perform operations separately and simultaneously. In this sense, parallel computing allows us to come up with quick, effective machine learning solutions. In conjunction with the rapid progress of GPU technology in current years, deep learning with multilayer structures has become one of the hottest topics in computer science [28]. Deep learning aims at achieving higher abstraction levels in transforming data since it provides a new representation of it over raw data [29, 30]. In the architecture of deep learning, classification forms part of the compact structure.

The notion of DML was introduced in the past few years because of the emergence of deep learning and metric learning [31]. The underlying principle of DML is the concept of sample similarity. An article by Lu et al. [31] presented the concept of DML for tasks involving visual comprehension in 2017. Figure 1 illustrates how the distance metric works. Our study evaluated current methods for image, text, video, and speech tasks. An important
factor in the success of DML is the network structure, loss function, and sample selection, and various aspects of these main factors have been discussed considering recent research. As an additional component, we also presented a quantitative comparison of the methods based on a general framework.

The rest of the article is organized as follows. Section 2 provides some background details about distance metric learning and widely used distance metrics with their recent improvements in DML and follows a discussion about the relationship between deep learning and metric learning. Section 3 explains the existing problems in DML. Section 4 presents some observations about the present and future prospects of DML and finally Section 5 is the conclusion of our study.

2. Metric Learning

2.1. Background of Metric Learning. As far as classification and clustering are concerned, each dataset presents its own set of challenges. Metrics that do not have an adequate learning capability independent of the problem can be viewed as unsuitable for classifying data. It is therefore necessary to obtain positive results from the input data using a good distance metric [32]. Several works utilizing metric learning approaches have been conducted to address this problem [27, 32–35]. Data-driven metric learning approaches can better distinguish between the samples of data if they perform the learning process on the data themselves. A key aim of metric learning is to study a new metric that lessens the gaps among samples of a similar class and raises the distances among samples of distinct classes [36] as shown in Figure 2.

2.2. Definition of Distance Metric Learning. The distance metric is a function that specifies the distance among elements of a set as a non-negative real number and a distance of zero indicates that both elements are equal by that metric. Elements need not be numbers but can instead be vectors, matrices, or arbitrary objects. In the state-space model, a state space is the Euclidean space, but in modern mathematics, the space has the Euclidean plane (a two-dimensional space) in which the variables on $x$ and $y$ (axes) are the state variables. If we consider $x$ and $y$ as members of two sets $(x_1, x_2, x_3, \ldots x_n)$ and $(y_1, y_2, y_3, \ldots y_n)$, then the idea of the distance between two members of this set is termed as a metric. Thus, a metric space has the following four properties to satisfy:

(i) The identity of indiscernible: The distance from $x$ to $y$ is zero if and only if $x$ and $y$ are the same.

(ii) Non-negativity: The distance between two distinct points is positive.
(iii) Symmetry: The distance from \( x \) to \( y \) is the same as the distance from \( y \) to \( x \).

(iv) Triangle inequality: The distance from \( x \) to \( y \) is less than or equal to the distance from \( x \) to \( y \) via any third point \( Z \).

If we relax the identity of an indiscernible condition to \( x \), and the distance from \( x \) to \( y \) is zero, then the distance is called a pseudometric.

2.3. Types of Distance Metrics. Measurements of the distance depend on the situation in which they are performed. The Manhattan Euclidean distance, for example, is useful for computing the distance in certain situations. For other applications, such as the cosine distance, a more refined approach is required. As there exists a wide variety of applications, such as the cosine distance, a more refined approach is required. As there exists a wide variety of distance measures, in the following list we present some of the most widely used distance metrics to compute distances between two points of data. They are as follows:

(i) Euclidean distance (ED)
(ii) Hamming distance (HD)
(iii) Manhattan distance (MD)
(iv) Chebyshev distance (CD)
(v) Levenshtein distance (LD)
(vi) Minkowski distance (MinD)

2.3.1. Euclidean Distance (ED). ED is calculated using the “Pythagoras theorem,” which states that the square of the hypotenuse side in a right-angle triangle is equal to the sum of squares of the other two sides:

\[
\text{Euclidean}(A, B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}. \tag{1}
\]

The ED between two points \( A (x_1, y_1) \) and \( B (x_2, y_2) \) as given in equation (1) is shown in Figure 3(a). Let \( A \) and \( B \) be two observations from our dataset, with \( x_1 \) and \( y_1 \) representing the two aspects of observation \( A \), and \( x_2 \) and \( y_2 \) representing the two features of observation \( B \). The ED should be used whenever we are comparing data that have continuous, numeric properties, such as heights, weights, or wages. A ED correlation-based approach is proposed to recognize 2D human face images [37].

2.3.2. Manhattan Distance (MD). The MD computes the sum of the absolute values of the variation of the coordinates of the two sites as shown in equation (2) rather than squaring the coordinate offset values and then calculating the square root of the sum of the squares. The MD determines how many squares are on a grid, representing the shortest path a car could take between two intersections from point \( A \) to point \( B \) [38]:

\[
\text{Manhattan}(A, B) = |x_1 - x_2| + |y_1 - y_2|. \tag{2}
\]

Figure 3(b) shows MD and ED in tandem. When the features of our observations are whole numbers (1, 2, 3, 4,...) with no decimal place, it becomes logical to apply the MD. A positive integer is always returned by the MD. In [39], the Manhattan tangent distance in outdoor fingerprint localization is proposed and lower computation complexity is achieved using an approximate Manhattan tangent distance.

2.3.3. Chebyshev Distance (CD). CD refers to the measurement of distance between two vectors when their variations are the greatest adjacent to any coordinate dimension. It is also commonly known as chessboard distance. This is because the minimum number of moves required by a king from one square to the next on a chessboard equals the CD between the centers of the squares, if the squares have a side length of 1, as represented in two-dimensional spatial coordinates with axes aligned to the edges of the board. An example of CD is shown in Figure 3(e). In two dimensions, if the points \( A \) and \( B \) have Cartesian coordinates \((x_1, y_1)\) and \((x_2, y_2)\), their CD is calculated as given in the following equation:

\[
\text{Chebyshev}(A, B) = \max(|x_2 - x_1|, |y_2 - y_1|). \tag{3}
\]

2.3.4. Minkowski Distance (MinD). The MinD is essentially a combination of both the ED and MD as shown in equation (4). The MD is obtained by multiplying the MinD by \( p = 1 \), and the ED is obtained by multiplying the MinD by \( p = 2 \). The CD is also given by \( p = \infty \). Figure 3(c) shows MinD measure with MD and ED representation as well.

\[
\text{Minkowski}(A, B) = \left( \sum_{i=1}^{n} |f_{a_i} - f_{b_i}|^p \right)^{1/p}. \tag{4}
\]

Common values of \( p \) are as follows:

\[
\begin{align*}
p & = 1 — \text{MD} \\
p & = 2 — \text{ED} \\
p & = \infty — \text{CD}
\end{align*}
\]

In the event of a decimal number between 1 and 2 (like 1.5), \( p \) can also be given intermediate values between 1 and 2 that provides a balance between ED and MD. If we are developing a distance metric method and are not sure which one to use, experimenting with the MinD with a few various values of \( p \) and seeing which one provides the best result is a good way to optimize one’s models. MinD used in Ref. [40] along with improved fuzzy possibilistic c-means algorithm was proven to be efficient for convex data and \( p \)-dimensional datasets.

2.3.5. Hamming Distance (HD). The HD is essentially a metric for comparing binary strings. The HD is probably the best way to determine the similarity between two data points if we have a dataset with “dummy” Boolean attributes. An example of HD is shown in Figure 3(d). Only if the two observations are from the same data collection can this measure be calculated. We cannot compute distance metrics across observations with different numbers of features, and it
is pointless to do so if the number of features is the same but the actual features are different. Adaptive HD [41] was used in Iris code matching, thereby improving the performance of Iris code matching.

2.3.6. Levenshtein Distance (LD). The LD is an alignment method for pairs of strings. When calculating the LD between two strings, the minimum number of changes in one string to transform into another are considered. As shown in Figure 3(f), consider two strings: \( A = \) “bitcoin” and \( B = \) “Altcoin.” To change the letter from “s” to “t,” two substitutions of the letters are needed, that is, “B” and “I” by “A” and “L.” Thus, Levenshtein \((A, B)=2 \times 2 = 4\). LD is applicable in many fields, including computational linguistics, computer science, natural language processing, and bioinformatics.

2.4. Recent Improvements in DML. The learning performance can be improved by linear metric learning methods, which support more flexible data constraints and flexible constraints in the transformed data space. In addition to having convex formulations, these approaches tend to be robust to overfitting [42]. Other than learning a good metric, it is also likely to develop a better representation of the data using linear approaches. To understand the data better, it is important to understand the nature of the data. Due to their poor ability to capture nonlinear features, linear transformations have a controlled ability to attain optimal execution over new data point representations. To overcome this issue, kernel-based methods are used in metric learning to carry the problem into a nonlinear space [27]. Despite their practicality for solving nonlinear problems, these nonlinear approaches may also negatively affect overfitting. As DML has become more popular, it is conceivable to suggest a solution to overcome the problems of both approaches in a more compact way. Currently, by leveraging neural networks with DML, computer vision applications have produced remarkable results. However, the current methods aim at a single deep distance metric based on pairs or triplets of samples. It is hard for it to handle heterogeneous data and avoid overfitting. To solve this, a boosting-based learning method of multiple deep distance metrics was introduced where the model produces the final distance metric through iterative training of weak distance metrics [43].

3. Deep Metric Learning (DML)

The DML method effectively measures similarities between two samples by mapping images to an embedding space based on ED. To accomplish this, a variety of methods have been proposed for embedding images with discriminative constraints [44–48]. Distances of Matusita and Akaike [49], Euclidean, Mahalanobis, Kullback–Leibler [50], and Bhattacharyya [51] are generally used for data classification as basic similarity metrics. However, these metrics have restricted applicability only to data classification. A Mahalanobis metric-based method was therefore proposed to address this problem by transforming the data into conventional metric learning. With this method, the data are reshaped into a new feature space with a greater distinction power. In most cases, metric learning relies on a linear transformation of data not including any kernel function. Unfortunately, these methods are ineffective in revealing the nonlinear information that is needed to overcome this
problem since they do not provide any apparent success due to issues such as scaling [52–54]. Conventional methods of metric learning solve this issue using linear activation functions, but deep learning uses nonlinear activation functions. Most deep learning approaches use the deep architectural background as the foundation rather than calculating distance metrics in a new representation space of the data. As a result, distance-based methods are one of the most fascinating areas of deep learning [36, 55–60], while DML decreases the distance between dissimilar samples.

DML increases the distance between similar samples, which is directly correlated to the distance between samples [61, 62]. A metric loss function has been utilized in deep learning to perform this task. To illustrate this process, Kaya and Bilge [63] conducted experiments on the MNIST dataset using the Siamese network with contrastive loss and thus proved that the goal of the above method can be used for successful implementation.

### 3.1. Problems in DML

Through deeper, nonlinear subspace learning that acquires embedded feature similarity using deep architectures, DML develops problem-based solutions because of learning from raw data. Its scope ranges from video understanding to virtually re-identifying people, recognizing medical problems, modeling three-dimensional (3D) images [55, 64], verifying facial features [61, 65, 66], and verifying signatures [67]. Understanding videos involves many different problems, such as video annotation, video recommendation, and video search. A metric space can be useful for figuring out solutions to such problems. To demonstrate, Lee et al. [68] initialized their work by extracting audio and visual properties from videos to benefit from a useful content. In addition to feature extraction and embedding algorithms, they showed a triplet embedding model based on deep neural networks, which is also a motivation for future studies. In Ref. [69], the authors prove that deep residual network-based metric learning is an effective approach for learning a moving human localization metric in video surveillance. When compared to popular DML methods, the method surpassed the rest. Visual tasks may not be well served by standard distance metrics since objects differ significantly from one another. Accordingly, Hu et al. [70] used deep learning based on distance metric as a substitute for utilizing a predefined similarity metric to increase distances between positive samples and decrease distances between negative samples for visual tracking.

Re-identification of individuals is another important problem in machine learning. Since deep learning methods have been gaining traction in recent years, the effectiveness of convolutional neural networks has been questioned [71]. An image re-identification task involves identifying the same person in different images taken in various situations. In this way, different distance metrics can be learned to solve these issues [72, 73]. In the context of person re-identification, DML provides us with the opportunity to integrate the input image and changed feature space at end-to-end [74]. Using this approach, a model is constructed based on tiered convolutions and maximum pooling. The proximity differences between inputs are then calculated. Finally, to decide whether the person is the same or different, patch summation attributes, cross-patch attributes, and the softmax function are used. Another study was conducted by Ding et al. [75] to increase the distance between two dissimilar images for triplet loss. However, one image could be incorporated into multiple triplet units, ultimately resulting in more triplet units. Due to this reason, the researchers optimized the gradient descent algorithm, which relies on the number of original images rather than the number of triplets, instead of the number of original images.

The above study categories include deep metric learning studies in diverse disciplines. However, it is likely to identify experiments conducted by researchers from other fields in which some problems regarding similarity in music [76], regression crowdedness [77], search of similar region [78], recognition of volumetric image [79], instance segmentation [80], detection of edge [81], sharpening-pan [82], and so on, were addressed. Due to its high performance in diverse areas, DML can therefore be claimed to make a significant contribution to the literature. Using a similar evaluation protocol for the benchmark datasets, Table 1 illustrates studies that have been published in the top journals and conferences in the past several years. Based on the outcomes presented in Table 1, DML has been productive in many distinct disciplines and each discipline has its evaluation metrics. From Table 1, we can observe that researchers have used different evaluation metrics for different problems. For example, F1 score, normalized mutual information (NMF), rank accuracy (R), first tier (FT), second tier (ST), nearest neighbor (NN), discounted cumulated gain (DCG), E-measure (E), and mean average precision (mAP).

### 3.2. Sample Selection and Loss Functions for DML

Sample selection: There are three main aspects of DML: informational input samples, structural network models, and a metric loss function. The selection of informative samples is arguably as important as the selection of DML models since both deal with metric loss functions and the success of DML depends heavily on the availability of informative samples. Initially, some articles tend to use Siamese networks in embedding learning as an easy sample pair in the beginning [89, 90]. The authors in Ref. [91], however, noted that as the network neared an acceptable performance level, the learning process could be slowed or adversely affected. With hard negative mining [91, 92], more discriminative models were developed to address this problem. Triplet networks use a positive, a negative, and an anchor sample to train a model for classification. A study conducted in Ref. [93] found that some simple triplets were ineffective at updating a model due to their inadequate discriminative power. Therefore, a very convenient and effective way to overcome these problems is to utilize informative sample triplets with more possible train models and an improved sampling strategy rather than just picking random samples [93, 94]. In Ref. [66], semi-hard negative mining was used for the first time to identify negative samples within the margins. But in Ref. [95], it was found that if negative
samples are too close to the anchor, the gradient had a high variance and a low signal-to-noise ratio. To avoid noisy samples, distance-weighted sampling was proposed [95]. In summary, regardless of how well we design mathematical models and architectures, the learning ability of the network is determined by how good the presented samples are presented at discriminating. Thus, the network must be presented with distinct training examples so that the network can gain more representation and learn better. In this way, progress in performance can be attained after choosing informative samples.

Loss functions: DML models involve loss functions as one of the primary components. To accomplish maximum feature depiction among the various objects, DML uses different loss functions. Studies have found that contrastive loss can benefit a Siamese network [89, 96]. A Siamese network, as illustrated in Figure 4, is an effective model to increase or decrease the distance between objects to enhance classification performance. To obtain a meaningful pattern among images in DML, shared weights are used that positively affect the performance of a neural network, as illustrated in Figure 4. Furthermore, sharing weights has significant advantages in terms of memory and time. Moreover, combining the Siamese network and CNN has many benefits [97], which include learning similarity from direct image pixels, informing color and textures at the same time, and its flexibility. As part of the metric learning model [98], Mahalanobis metrics and Siamese CNN were combined for the re-identification of individuals, whereas Mahalanobis metrics were used for classification. A face recognition algorithm based on softmax and center loss was proposed by the authors of Ref. [99]. Like the contrastive loss, the center loss attempts to find deep features that decrease the distances between their centers, but the softmax loss attempts to increase the distances between classes. Using class-based hierarchical trees, the authors proposed a new metric loss based on triplet loss in Ref. [100]. In a similar vein, Wang et al. [101] conceptualized a novel angular loss to improve DML. The authors of [102] demonstrated that they could achieve a greater degree of closeness between objects by using quadruple samples. Like quadruplet loss, histogram loss [103] utilizes quadruplet samples for training. Unlike other losses, it does not require tuning parameters since its similarity distributions are calculated using histograms. As compared to other losses, it achieves superior results in experimental studies using re-identification datasets, such as CUHK03 [104] and Market-1501 [105]. Using an SVM learning constraint to minimize learning risk in the person a re-identification task was proposed by Yao et al. [106]. The goal of part loss is to target the various parts of the body instead of concentrating on a single point. State-of-the-art loss metrics in the literature are encapsulated in Table 2 in detail.

### 4. Discussion

A prior section of this article discussed how DML can be applied to tasks such as face verification, recognition, and person re-identification. Training samples for single categories are limited for these tasks with many categories. It is possible to complicate a successful training process if there are not enough samples for each category. A DML algorithm can process two, three, or four samples using a network structure, such as the Siamese network, triplet network, or quadruple network. Using these network structures permits significant increases in training data with greater accuracy. This means that even small samples in a single category can improve DML. The authors of Ref. [102] demonstrated that they could achieve a greater degree of closeness between objects by using quadruple samples. Like quadruplet loss, histogram loss [103] utilizes quadruplet samples for training. Unlike other losses, it does not require tuning parameters since its similarity distributions are calculated using histograms. As compared to other losses, it achieves superior results in experimental studies using re-identification datasets, such as CUHK03 [104] and Market-1501 [105]. Using an SVM learning constraint to minimize learning risk in the person a re-identification task was proposed by Yao et al. [106]. The goal of part loss is to target the various parts of the body instead of concentrating on a single point. State-of-the-art loss metrics in the literature are encapsulated in Table 2 in detail.

### Table 1: Comparative result analysis of benchmark datasets for various DML problems.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Task and results</th>
<th>3D shape retrieval</th>
<th>Verification of face accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>FT</td>
<td>ST</td>
<td>NN</td>
<td>DGG</td>
</tr>
<tr>
<td>64.2</td>
<td>72.2</td>
<td>66.0</td>
<td>77.0</td>
</tr>
<tr>
<td>72.3</td>
<td>77.4</td>
<td>82.3</td>
<td>71.2</td>
</tr>
<tr>
<td>79.7</td>
<td>85.0</td>
<td>85.4</td>
<td>75.7</td>
</tr>
<tr>
<td>FT</td>
<td>ST</td>
<td>NN</td>
<td>DGG</td>
</tr>
<tr>
<td>28.2</td>
<td>35.1</td>
<td>26.9</td>
<td>50.0</td>
</tr>
<tr>
<td>33.2</td>
<td>39.2</td>
<td>39.9</td>
<td>54.1</td>
</tr>
<tr>
<td>46.2</td>
<td>54.2</td>
<td>59.0</td>
<td>67.3</td>
</tr>
</tbody>
</table>

Moreover, combining the Siamese network and CNN has many benefits [97], which include learning similarity from direct image pixels, informing color and textures at the same time, and its flexibility. As part of the metric learning model [98], Mahalanobis metrics and Siamese CNN were combined for the re-identification of individuals, whereas Mahalanobis metrics were used for classification. A face recognition algorithm based on softmax and center loss was proposed by the authors of Ref. [99]. Like the contrastive loss, the center loss attempts to find deep features that decrease the distances between their centers, but the softmax loss attempts to increase the distances between classes. Using class-based hierarchical trees, the authors proposed a new metric loss based on triplet loss in Ref. [100]. In a similar vein, Wang et al. [101] conceptualized a novel angular loss to improve DML. The authors of [102] demonstrated that they could achieve a greater degree of closeness between objects by using quadruple samples. Like quadruplet loss, histogram loss [103] utilizes quadruplet samples for training. Unlike other losses, it does not require tuning parameters since its similarity distributions are calculated using histograms. As compared to other losses, it achieves superior results in experimental studies using re-identification datasets, such as CUHK03 [104] and Market-1501 [105]. Using an SVM learning constraint to minimize learning risk in the person a re-identification task was proposed by Yao et al. [106]. The goal of part loss is to target the various parts of the body instead of concentrating on a single point. State-of-the-art loss metrics in the literature are encapsulated in Table 2 in detail.

#### 4. Discussion

A prior section of this article discussed how DML can be applied to tasks such as face verification, recognition, and person re-identification. Training samples for single categories are limited for these tasks with many categories. It is possible to complicate a successful training process if there are not enough samples for each category. A DML algorithm can process two, three, or four samples using a network structure, such as the Siamese network, triplet network, or quadruple network. Using these network structures permits significant increases in training data with greater accuracy. This means that even small samples in a single category can improve the performance of the network. According to Table 1, DML algorithms have demonstrated excellent performance for these tasks, even when there are a lot of categories and few samples per category.

When evaluating DML, which includes metric loss function, sampling strategy, and network structure, all the network components should be considered together. The sample to be presented to the network and its relationship with the metric loss function is determined by the dataset. Losses such as contrastive loss [89], triplet loss [107], quadruplet loss [102], and n-pair loss [108] are types of metric loss functions that allow us to incorporate paired samples, triplet samples, and quadruplet samples to increase the data sample size (n). The network training process becomes too time-consuming and memory-intensive when samples are
paired or tripled. Depending on the situation, training networks become exponentially more difficult.

The hard negative mining method [91, 92] and semi-hard negative mining method [66, 102] provide informative samples for training to overcome these problems. Despite providing the desired results in specific tasks, hard mining and semi-hard mining strategies consume a great deal of time and memory compared to the traditional method. In addition, the GPU memory limit makes it impossible at times when using large batch sizes. This can be overcome by clustering loss [109], which has an excellent metric function that requires no data preparation. The authors in Ref. [66] used a CPU cluster to implement their mining strategy to achieve a huge batch on CPU clusters, while deep metric learning is typically performed on a GPU. It may not be possible for some datasets to achieve fast convergence with the metric loss function. To solve this problem, the weights from pretrained network models may be used to ensure faster convergence and better differentiation in embedding space [108].

5. Conclusion

A field of research that researchers have taken interest in recently is DML based on distance metrics. Several academic papers have contributed immensely to the literature on this topic. This article fills the literature gap by providing a comprehensive look at DML that considers all aspects of the technology and the problems associated with this field. Most current studies conducted in this area are commonly influenced by Siamese and triplet networks in DML and proved their higher efficiency on benchmark datasets and specific tasks. However, studies are limited to a few areas. This could be fascinating for researchers given that there are many aspects of DML that have not yet been explored, such as the shortcomings of existing approaches. Thus, DML is still open for future research and can be improved in the long run.

Conflicts of Interest

The authors declare no conflicts of interest in relation to this article.

Acknowledgments

This research was carried out with the support of the Kyungpook National University Research Fund, 2021.

References


[104] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deep ReID: deep filter pairing neural network for person re-identification," in...


