

Review Article

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

Academic Editor: Juan Vicente Capella Hernandez

Copyright © 2022 Barathi Subramanian et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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

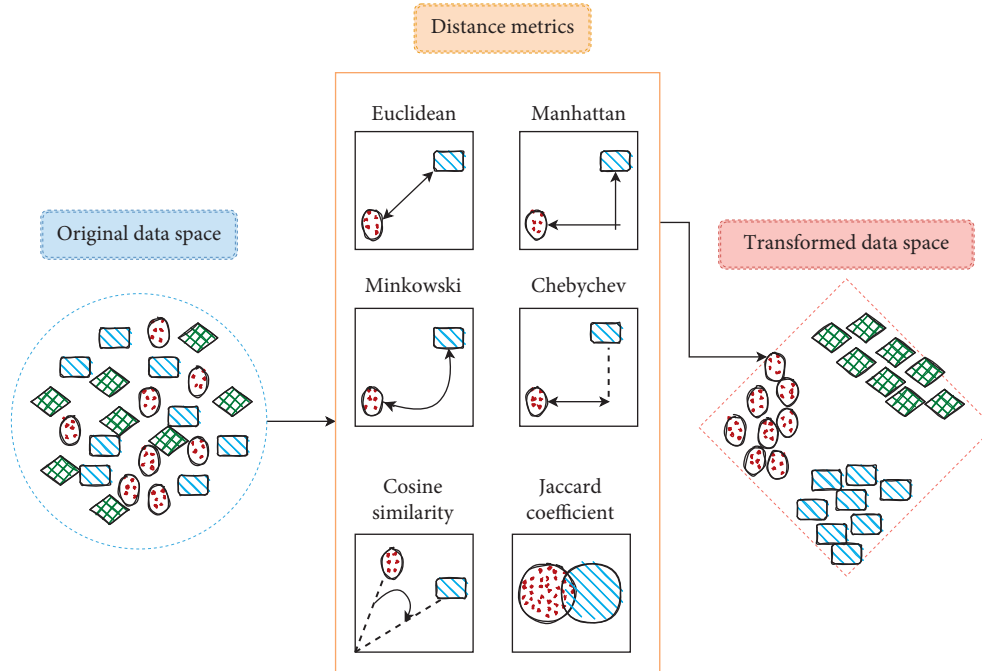


FIGURE 1: General workflow of distance metrics.

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.

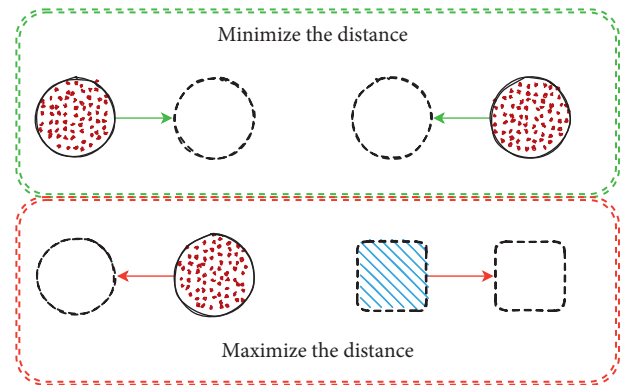


FIGURE 2: Goal of metric learning.

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 $(x_1, x_2, x_3 \dots x_n)$ and $(y_1, y_2, y_3 \dots y_n)$, 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 is equal to y , 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 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:

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

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]:

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

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:

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

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 $p = \text{infinity}$. Figure 3(c) shows MinD measure with MD and ED representation as well.

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

Common values of p are as follows:

- $p = 1$ —MD
- $p = 2$ —ED
- $p = \infty$ —CD

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

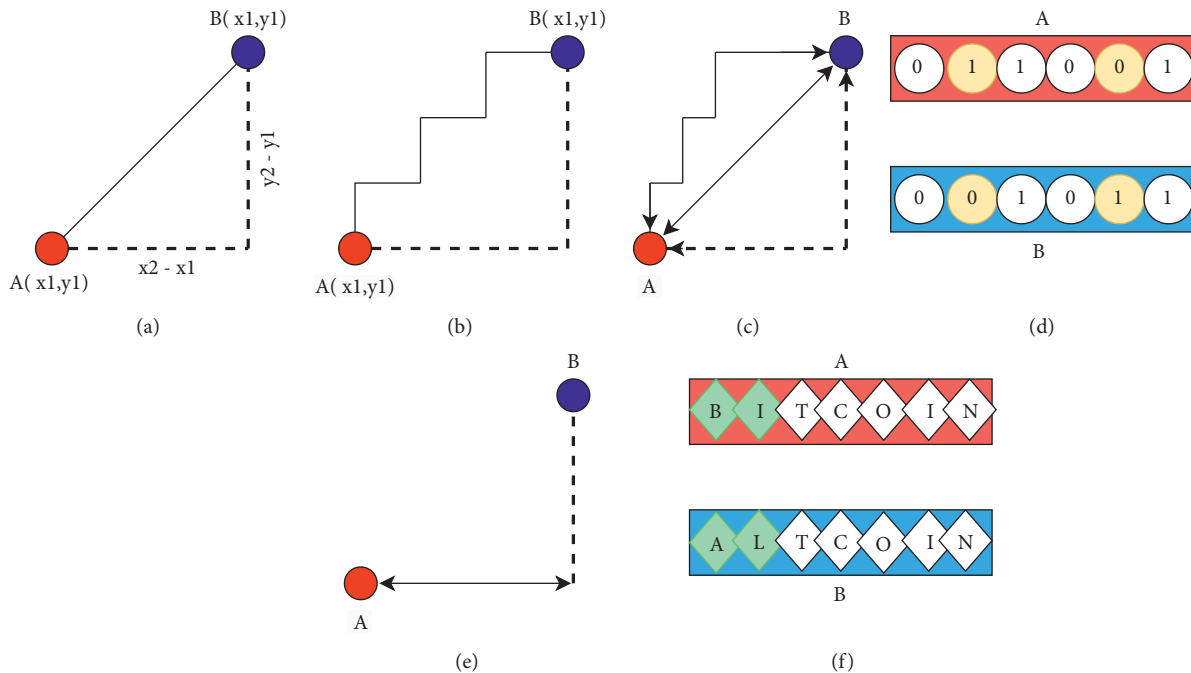


FIGURE 3: Diagrammatic representation of distance metrics. (a) Euclidean distance, (b) Manhattan distance, (c) Minkowski distance, (d) Hamming distance, (e) Chebyshev distance, and (f) Levenshtein distance.

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: $A = \text{"bitcoin"}$ and $B = \text{"Altcoin.}"$ 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, $\text{Levenshtein}(A, B) = 2 * 2 = 4$. 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 (NMI), rank accuracy (R), first tier (FT), second tier (ST), nearest neighbor (NN), discounted cumulated gain (DCG), Emeasure (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

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

	Clustering image (%)		Retrieval of image recall @ R (%)			
	NMI	FI	R=1	R=2	R=4	R=8
[83]	56.2	22.2	46.5	58.1	69.8	80.2
	60.3	27.2	50.9	63.3	74.2	83.3
	61.1	29.4	54.7	66.3	76.0	83.8
	59.2	—	48.1	61.4	71.8	81.8
	—	—	57.1	68.8	78.6	86.5
[84]	55.1	21.7	48.4	61.2	71.9	81.1
	63.8	33.8	71.3	80.0	86.3	91.8
	59.0	32.4	71.1	81.3	87.7	92.0
	—	—	58.2	70.8	80.1	87.6
	—	—	81.3	88.0	92.8	95.8
[85]	87.3	24.7	6 3.1	80.6	91.6	97.8
	88.2	28.1	67.6	83.8	93.0	98.0
	88.7	29.9	80.0	85.0	93.6	98.1
	89.3	—	67.1	83.5	93.3	-
	—	—	84.7	88.4	94.9	98.3
<i>Dataset</i>	<i>Task and results</i>					
	<i>3D shape retrieval</i>					
[86]	FT	ST	NN	DGG	E	MAP
	64.2	72.2	66.0	77.0	35.2	67.2
	72.3	77.4	73.3	82.1	37.0	74.1
	79.7	85.0	76.1	85.4	40.0	81.2
[87]	FT	ST	NN	DGG	E	MAP
	28.2	35.1	26.9	50.0	16.9	29.2
	33.2	39.2	39.9	54.1	19.9	34.1
	46.2	54.2	59.0	67.3	28.0	48.2
	Verification of face accuracy					
[88]	91.46					
	95.12					
	92.99					

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 are 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 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], quadruple loss [102], and n-pair loss [108] are types of metric loss functions that allow us to incorporate paired samples, triplet samples, and quadruple samples to increase the data sample size (n). The network training process becomes too time-consuming and memory-intensive when samples are

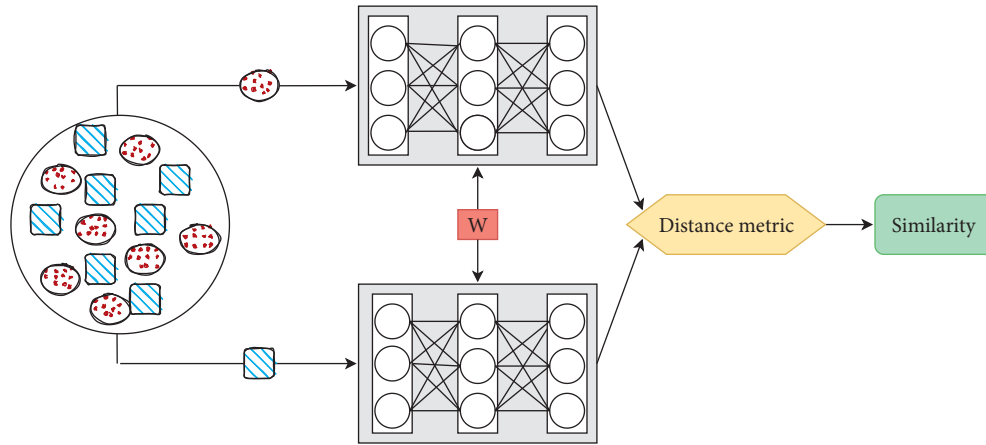


FIGURE 4: An example of a Siamese neural network with DML.

TABLE 2: Literature summary of loss metrics and sample selection.

Sample selection	Metric loss	Task
Easy	Triplet [107]	Recognition of image Recognition of object
Multiple negative	Angular [101]	Clustering of image Retrieval of image Verification of object
Multiple negative	N-pair [108]	Clustering of image Identification of face Verification of face Recognition of object
Easy	Clustering [109]	Clustering of image Retrieval of image Person re- Identification
Easy	Histogram [103]	Recognition of image Retrieval of image Object recognition
Hard negative	Contrastive [89]	Image recognition
Semi-hard negative	Quadruple [102]	Patient similarity
Easy	Part [106]	Person re- Identification
Anchor-neighbor	Hierarchical triplet [100]	Recognition of face Retrieval of image

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

- [1] M. Zongqing, Z. Shuang, W. Xi et al., "Nasopharyngeal Carcinoma Segmentation Based on Enhanced Convolutional Neural Networks Using Multi-Modal Metric Learning," *Physics in Medicine & Biology*, vol. 64, no. 2, p. 64, 2019.
- [2] G. Wei, M. Qiu, K. Zhang et al., "A multi-feature image retrieval scheme for pulmonary nodule diagnosis," *Medicine*, vol. 99, no. 4, 8 pages, Article ID e18724, 2020.

- [3] T. Li, G. Kou, and Y. Peng, "Improving malicious URLs detection via feature engineering: linear and nonlinear space transformation methods," *Information Systems*, vol. 91, Article ID 101494, 2020.
- [4] Y. Luo, H. Hu, Y. Wen, and D. Tao, "Transforming device fingerprinting for wireless security via online multitask metric learning," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 208–219, 2020.
- [5] Y. Liu, D. Pi, and L. Cui, "Metric learning combining with boosting for user distance measure in multiple social networks," *IEEE Access*, vol. 5, pp. 19342–19351, 2017.
- [6] Y. Liu, Z. Gu, T. H. Ko, and J. Liu, "Multi-modal media retrieval via distance metric learning for potential customer discovery," in *Proceedings of the 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, Santiago, Chile, December 2019.
- [7] R. Li, J. Y. Jiang, J. Liu, C. C. Hsieh, and W. Wang, "Automatic speaker recognition with limited data," in *Proceedings of the WSDM 2020 - Proc 13th Int Conf Web Search Data Min*, TX, Houston, USA, February 2020.
- [8] Z. Bai, X. L. Zhang, and J. Chen, "Speaker verification by partial AUC optimization with Mahalanobis distance metric learning," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1533–1548, 2020.
- [9] W. Zheng, B. Zhang, J. Lu, and J. Zhou, "Deep relational metric learning," in *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, Canada, 2021.
- [10] D. López-sánchez, A. González, and J. M. Corchado, "Neurocomputing Visual content-based web page categorization with deep transfer learning and metric learning," vol. 338, pp. 418–431, 2019.
- [11] H. Hu, K. Wang, C. Lv, J. Wu, and Z. Yang, "Semi-supervised metric learning-based anchor graph hashing for large-scale image retrieval," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 739–754, 2019.
- [12] H. Wu, Q. Zhou, R. Nie, and J. Cao, "Effective metric learning with co-occurrence embedding for collaborative recommendations," *Neural Networks*, vol. 124, pp. 308–318, 2020.
- [13] X. Li and Y. Tang, "A Social Recommendation Based on Metric Learning and Network Embedding," in *Proceedings of the 2020 IEEE 5th Int Conf Cloud Comput Big Data Anal ICCCBDA 2020 55–60*, Chengdu, China, April 2020.
- [14] B. Nguyen, B. De Baets, and B. De Baets, *Constraints for Person Re-Identification*, vol. 28, pp. 589–600, 2019.
- [15] C. Zhao, X. Wang, W. Zuo, F. Shen, L. Shao, and D. Miao, "Similarity learning with joint transfer constraints for person re-identification," *Pattern Recognition*, vol. 97, Article ID 107014, 2020.
- [16] J. Liang, Q. Hu, C. Dang, and W. Zuo, "Weighted graph embedding-based metric learning for kinship verification," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1149–1162, 2019.
- [17] F. Dornaika, I. Arganda-Carreras, and O. Serradilla, "Transfer Learning and Feature Fusion for Kinship Verification," *Neural Computing and Applications*, vol. 32, no. 11, p. 7151, 2020.
- [18] D. Wang, Y. Cheng, M. Yu, X. Guo, and T. Zhang, "A hybrid approach with optimization-based and metric-based meta-learner for few-shot learning," *Neurocomputing*, vol. 349, pp. 202–211, 2019.
- [19] C. Wang, G. Peng, and B. De Baets, "Deep feature fusion through adaptive discriminative metric learning for scene recognition," *Information Fusion*, vol. 63, pp. 1–12, 2020.
- [20] X. He, W. Y. Ma, O. King, M. Li, and H. Zhang, "Learning and inferring a semantic space from user's relevance feedback for image retrieval," *Proceedings of the tenth ACM international conference on Multimedia - MULTIMEDIA '02*, vol. 13, pp. 343–346, 2002.
- [21] X. He, W. Y. Ma, and H. J. Zhang, "Learning an image manifold for retrieval," *Proceedings of the 12th annual ACM international conference on Multimedia - MULTIMEDIA '04*, pp. 17–23, NY, New York, USA, October 2004.
- [22] J. He, M. Li, and H. Zhang, "Technical-session-1-content-based-image-retrieval-Manifold-ranking-based-image-retrieval," 2004.
- [23] A. G. Hauptmann and H. I. Search, "Relevance R negative pseudo-relevance feedback in content-based,".
- [24] H. Muller, T. Pun, and D. Squire, "Learning from user behavior in image retrieval: application of Market basket analysis," *International Journal of Computer Vision*, vol. 56, no. 1/2, pp. 65–77, 2004.
- [25] A. Globerson and S. Roweis, "Metric learning by collapsing classes," *Advances in Neural Information Processing Systems*, vol. 18, pp. 451–458, 2005.
- [26] W. Fei and S. Jimeng, "Survey on Distance Metric Learning and Dimensionality reduction in Data Mining," *Data mining and knowledge discovery*, vol. 29, no. 2, p. 564, 2015.
- [27] Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, pp. 207–244, 2009.
- [28] J. Schmidhuber, "Deep Learning in neural networks: an overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.
- [29] Y. LeCun, Y. Bengio, and G. Hinton, *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [30] T. K. Tran and T. T. Phan, "Deep learning application to ensemble learning—the simple, but effective, approach to sentiment classifying," *Applied Sciences*, vol. 9, no. 13, p. 2760, 2019.
- [31] J. Lu, J. Hu, and J. Zhou, "Deep metric learning for visual understanding: an overview of recent advances," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 76–84, 2017.
- [32] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell, "Distance metric learning, with application to clustering with side-information," *Advances in neural information processing systems*, vol. 15, 2002.
- [33] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in *Proceedings of the 24th international conference on Machine learning - ICML '07*, pp. 209–216, Oregon, Corvallis, USA, June 2007.
- [34] H. V. Nguyen and L. Bai, "Cosine similarity metric learning for face," *Journal of Information Science*, vol. 43, pp. 88–102, 2011.
- [35] K. Q. Weinberger, J. Blitzer, and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of machine learning research*, vol. 10, pp. 1473–1480, 2005.
- [36] Y. Duan, J. Lu, J. Feng, and J. Zhou, "Deep localized metric learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 10, pp. 2644–2656, 2018.
- [37] F. Sayeed, M. Hanmandlu, and A. Q. Ansari, "Face recognition using segmental Euclidean distance," *Defence Science Journal*, vol. 61, no. 5, pp. 431–442, 2011.

- [38] X.-S. Yang, "Data mining techniques," *Introduction to Algorithms for Data Mining and Machine Learning*, pp. 109–128, 2019.
- [39] Z. Li, X. Zhong, J. Wei, and H. Shi, "The application of manhattan tangent distance in outdoor fingerprint localization," in *Proceedings of the 2018 IEEE Glob Commun Conf GLOBECOM 2018 - Proc*, Abu Dhabi, United Arab Emirates, December 2018.
- [40] H. Chouikhi, M. F. Saad, and A. M. Alimi, "Improved fuzzy possibilistic C-means (IFPCM) algorithms using Minkowski distance," in *Proceedings of the 2017 Int Conf Control Autom Diagnosis, ICCAD*, Hammamet, Tunisia, January 2017.
- [41] A. B. Dehkordi and S. A. R. Abu-Bakar, "Iris code matching using adaptive Hamming distance," in *Proceedings of the IEEE 2015 Int Conf Signal Image Process Appl ICSIPA 2015 - Proc 404–408*, Kuala Lumpur, Malaysia, October 2016.
- [42] A. Bellet, A. Habrard, and M. Sebban, "A Survey on Metric Learning for Feature Vectors and Structured Data," 2013, <https://arxiv.org/abs/1306.6709>.
- [43] Z. Li, "A boosting-based deep distance metric learning method," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 2665843, 9 pages, 2022.
- [44] F. Cakir, K. He, X. Xia, B. Kulis, and S. Sclaroff, "Deep metric learning to rank," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, June 2019.
- [45] I. Elezi, S. Vascon, A. Torcinovich, M. Pelillo, and L. Leal-Taixé, "The group loss for deep metric Learning.pdf," *ECCV*, vol. 12352, 2020.
- [46] Y. Sun, C. Cheng, Y. Zhang et al., *A Unified Perspective of Pair Similarity Optimization*, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, p. 6407p. 6407, 2020.
- [47] X. Wang, X. Han, W. Huang, D. Dong, and M. R. Scott, "Multi-Similarity Loss with General Pair Weighting for Deep Metric Learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5022–5030, 2019.
- [48] B. Yu and D. Tao, "Deep metric learning with tuplet margin loss," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.
- [49] K. Matusita and H. Akaike, "Decision rules, based on the distance, for the problems of independence, invariance and two samples," *Annals of the Institute of Statistical Mathematics*, vol. 7, no. 2, pp. 67–80, 1955.
- [50] A. Elgammal, R. Duraiswami, and L. S. Davis, "Probabilistic tracking in joint feature-spatial spaces," *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings*.vol. 1, 2003.
- [51] F. J. Aherne, N. A. Thacker, and P. I. Rockett, "The Bhattacharyya metric as an absolute similarity measure for frequency coded data," *Kybernetika*, vol. 34, pp. 363–368, 1998.
- [52] J. Yu, X. Yang, F. Gao, and D. Tao, "Deep multimodal distance metric learning using click constraints for image ranking," *IEEE Transactions on Cybernetics*, vol. 47, no. 12, pp. 4014–4024, 2017.
- [53] X. Cai, C. Wang, B. Xiao, and Y. Shao, "Nonlinear metric learning with deep independent subspace analysis network for face verification," *IEICE - Transactions on Info and Systems*, vol. E96.D, no. 12, pp. 2830–2838, 2013.
- [54] Y. Sun, Y. Zhu, Y. Zhang et al., "Dynamic metric learning: towards a scalable metric space to accommodate multiple semantic scales," in *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [55] G. Dai, J. Xie, F. Zhu, and Y. Fang, "Deep correlated metric learning for sketch-based 3D shape retrieval," in *Proceedings of the 31st AAAI Conf Artif Intell AAAI 2017*, 2017.
- [56] Z. Li and J. Tang, "Weakly supervised deep metric learning for community-contributed image retrieval," *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 1989–1999, 2015.
- [57] B. Harwood, G. Vkb, G. Carneiro, I. Reid, and T. Drummond, "Smart Mining for Deep Metric Learning," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2821–2829, 2017.
- [58] E. Gundogdu, B. Solmaz, A. Koc, V. Yücesoy, and A. A. Alatan, "Deep Distance Metric Learning for Maritime Vessel Identification," in *Proceedings of the 2017 25th Signal Processing and Communications Applications Conference*, Antalya, Turkey, May 2017.
- [59] E. Costa, I. Papatsouma, and A. Markos, "Benchmarking distance-based partitioning methods for mixed-type data," 2022, <https://arxiv.org/abs/2203.16287>.
- [60] C. O. F. Blup, "Distance Based Regression Models," 2021.
- [61] J. Liu, Y. Deng, T. Bai, Z. Wei, and C. Huang, "Targeting ultimate accuracy: face recognition via deep embedding," 2015, <https://arxiv.org/abs/1506.07310>.
- [62] E. Hoffer and N. Ailon, "Semi-supervised deep learning by metric em-bedding," pp. 14–26, 2018, <https://arxiv.org/abs/1611.01449>.
- [63] M. Kaya and H. Ş Bilge, "Deep metric learning: a survey," *Symmetry*, vol. 11, no. 9, p. 1066, 2019.
- [64] I. Lim, A. Gehre, and L. Kobbelt, "Identifying style of 3D shapes using deep metric learning," *Computer Graphics Forum*, vol. 35, no. 5, pp. 207–215, 2016.
- [65] J. Hu, J. Lu, and Y. P. Tan, "Discriminative deep metric learning for face verification in the wild," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, N W Washington, 2014.
- [66] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: a unified embedding for face recognition and clustering," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [67] J. Bromley, J. W. Bentz, L. Bottou et al., "Signature verification using a "siamese" time delay neural network," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 07, no. 04, pp. 669–688, 1993.
- [68] J. Lee, B. Varadarajan, S. Abu-El-Haija, and A. P. Natsev, "Collaborative deep metric learning for video understanding," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 481–490, London, 2018.
- [69] W. Huang, H. Ding, and G. Chen, "A novel deep multi-channel residual networks-based metric learning method for moving human localization in video surveillance," *Signal Processing*, vol. 142, pp. 104–113, 2018.
- [70] J. Hu, J. Lu, and Y. P. Tan, "Deep metric learning for visual tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 11, pp. 2056–2068, 2016.
- [71] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person Re-identification Past, Present and Future," 2016, <https://arxiv.org/abs/1610.02984>.
- [72] M. Chen, Y. Ge, X. Feng, C. Xu, and D. Yang, "Person re-identification by pose invariant deep metric learning with improved triplet loss," *IEEE Access*, vol. 6, pp. 68089–68095, 2018.

- [73] X. Yang, P. Zhou, and M. Wang, "Person reidentification via structural deep metric learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 10, pp. 2987–2998, 2019.
- [74] A. Hermans, L. Beyer, and B. Leibe, *Defense of the Triplet Loss for Person Re-identification*, <https://arxiv.org/abs/1703.07737>, 2017.
- [75] S. Ding, L. Lin, G. Wang, and H. Chao, "Deep feature learning with relative distance comparison for person re-identification," *Pattern Recognition*, vol. 48, no. 10, pp. 2993–3003, 2015.
- [76] R. Lu, K. Wu, and Z. Duan, "Deep ranking: triplet matchnet for music metric learning state key Lab of Intelligent Technologies and Systems Tsinghua National Laboratory for Information Science and Technology," *IEEE Int Conf Acoust Speech, Signal Process.*, New Orleans, LA, USA, March 2017.
- [77] Q. Wang, J. Wan, and Y. Yuan, "Deep metric learning for crowdedness regression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 10, pp. 2633–2643, 2018.
- [78] Y. Liu, K. Zhao, and G. Cong, "Efficient similar region search with deep metric learning," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, United Kingdom, August 2018.
- [79] X. Wang and M. Liu, "Multi-view deep metric learning for volumetric image recognition," in *Proceedings of the 2018 IEEE International Conference on Multimedia and Expo (ICME)*, San Diego, CA, USA, July 2018.
- [80] A. Fathi, Z. Wojna, V. Rathod et al., "Semantic Instance Segmentation via Deep Metric Learning," 2017, <https://arxiv.org/abs/1703.10277>.
- [81] S. Cai, J. Huang, X. Ding, and D. Zeng, "Semantic edge detection based on deep metric learning," in *Proceedings of the 2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, Xiamen, China, November 2017.
- [82] Y. Xing, M. Wang, S. Yang, and L. Jiao, "Pan-sharpening via deep metric learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, pp. 165–183, 2018.
- [83] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, *The Caltech-UCSD Birds-200-2011 Dataset. Comput. Neural Syst. Tech. Report, CNS-TR-2011-001*, Calif. Inst. Technol, Pasadena, CA, USA, 2011.
- [84] J. Krause, M. Stark, J. Deng, and L. Fei-Fei, "3D object representations for fine-grained categorization," in *Proceedings of the 2013 IEEE International Conference on Computer Vision Workshops*, pp. 554–561, Sydney, NSW, Australia, December 2013.
- [85] H. O. Song, Y. Xiang, S. Jegelka, and S. Savarese, "Deep metric learning via lifted structured feature embedding," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, June 2016.
- [86] B. Li, Y. Lu, A. Godil et al., "A comparison of methods for sketch-based 3D shape retrieval," *Computer Vision and Image Understanding*, vol. 119, pp. 57–80, 2014.
- [87] B. Li, Y. Lu, C. Li et al., "A comparison of 3D shape retrieval methods based on a large-scale benchmark supporting multimodal queries," *Computer Vision and Image Understanding*, vol. 131, pp. 1–27, 2015.
- [88] A. J. Howell and H. Buxton, "Towards unconstrained face recognition from image sequences," in *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, pp. 224–229, Killington, VT, USA, October 1996.
- [89] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)*, pp. 1735–1742, New York, NY, USA, June 2006.
- [90] S. Bell and K. Bala, "Learning visual similarity for product design with convolutional neural networks," *ACM Transactions on Graphics*, vol. 34, no. 4, pp. 1–10, 2015.
- [91] E. Simo-Serra, E. Trulls, L. Ferraz, I. Kokkinos, P. Fua, and F. Moreno-Noguer, "Discriminative learning of deep convolutional feature point descriptors," in *Proceedings of the 015 IEEE International Conference on Computer Vision (ICCV)*, N W Washington, 2015.
- [92] M. Bucher, H. S. Ephane, and J. F. Eric, "Hard negative mining for Metric learning based zero-shot Classification," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 524–531, Cham, November 2016.
- [93] Y. Cui, F. Zhou, Y. Lin, and S. Belongie, "Fine-grained categorization and dataset bootstrapping," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1153–1162, 2017.
- [94] Y. Movshovitz-Attias, A. Toshev, T. K. Leung, S. Ioffe, and S. Singh, "No fuss distance metric learning using Proxies," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 360–368, 2017.
- [95] C. Wu, U. T. Austin, and A. Amazon, "Supplementary material: sampling matters in deep embedding learning," *Iccv*, 2017.
- [96] Y. Jeong, S. Lee, D. Park, and K. H. Park, "Accurate age estimation using multi-task siamese network-based deep metric learning for front face images," *Symmetry*, vol. 10, no. 9, p. 385, 2018.
- [97] D. Yi, Z. Lei, and S. Liao, "Deep metric learning for person Re-identification and De-identification," in *Proceedings of the 22nd International Conference on Pattern Recognition*, pp. 34–39, Stockholm, August 2014.
- [98] H. Shi, X. Zhu, S. Liao, Z. Lei, Y. Yang, and S. Z. Li, "Constrained deep metric learning for person Re-identification," pp. 34–39, 2015, <https://arxiv.org/abs/1511.07545>.
- [99] Y. Wen, K. Zhang, B. Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Proceedings of the European Conference on Computer Vision*, pp. 499–511, Cham, September 2016.
- [100] W. Ge, W. Huang, D. Dong, and R. Matthew, "Deep Metric Learning with Hierarchical Triplet Loss," in *Proceedings of the European Conference on Computer Vision (ECCV)*, p. 288p. 288, 2018.
- [101] J. Wang, F. Zhou, S. Wen, X. Liu, and Y. Lin, "Deep metric learning with angular loss," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, October 2017.
- [102] J. Ni, J. Liu, C. Zhang, D. Ye, and Z. Ma, "Fine-grained patient similarity measuring using deep metric learning," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 1189–1198, Singapore, Singapore, 2017.
- [103] E. Ustinova and V. Lempitsky, "Learning deep embeddings with histogram loss," *Advances in Neural Information Processing Systems*, pp. 4177–4185, 2016.
- [104] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: deep filter pairing neural network for person re-identification," in

Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, NW Washington, June 2014.

- [105] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable Person Re-identification : A Benchmark University of Texas at San Antonio," in *Proceedings of the IEEE international conference on computer vision*, Montreal, BC, 2015.
- [106] H. Yao, S. Zhang, R. Hong, Y. Zhang, C. Xu, and Q. Tian, "Deep representation learning with Part Loss for person Re-identification," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 2860–2871, 2019.
- [107] E. Hoffer and N. Ailon, "Deep metric learning using triplet network. Pattern Recognition, Image Anal Comput Vision," *Appl Springer*, vol. 9370, pp. 84–92, 2015.
- [108] K. Sohn, "Improved deep metric learning with multi-class N-pair loss objective," *Advances in Neural Information Processing Systems*, vol. 29, pp. 1857–1865, 2016.
- [109] H. O. Song, S. Jegelka, V. Rathod, and K. Murphy, "Deep Metric Learning via Facility Location," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2206–2214, 2017.