INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 4, April 2026 | ISSN: 2456-4184 | IJNRD.ORG

DEEP METRIC LEARNING FOR USER IDENTIFICATION: A COMPREHENSIVE SURVEY OF METHODS, APPLICATIONS, AND CHALLENGES

[1]Priya Thakkar, [2]Dr. Dheeraj Kumar Singh, [3]Dr. Dinesh Prajapati

[1] PG Student, [2] Associate Professor, [3] Associate Professor
[1]Information Technology,
[1]A.D.Patel Institute of Technology , Anand, Gujarat, India

Abstract : One of the important problems in the field of security systems is reliably verifying the identity of a user or a person. Conventional biometric systems have large intra-class variation and inter-class similarity. Deep Metric Learning (DML) is a suitable method for addressing this problem due to its capability to learn a discriminative embedding space where similar identities are closer to each other. This paper tries to make an exhaustive review of recent advances, central techniques, and applications of DML in user identification. We trace DML's basic contribution to the classic area of face recognition and stretch it to innovative behavioral biometrics- such as identifying users by motion patterns in eXtended Reality (XR) and web browsing histories. We also describe resilient multimodal systems that fuse together such data as facial features plus dynamic signatures. The paper wraps up basic loss functions (e.g., Contrastive, Triplet), advanced variants, network architectures, and some of the most important optimizations such as hard negative mining. Found thereby are problems that shall continue to challenge to extend system scalability and generalization for new users without any form of retraining, not speaking of the deep dimensions toward privacy and security risks with sensitive biometric data. We particularly underscore the crucial and largely neglected issue of demographic bias (e.g., by race and gender) where high overall accuracy can conceal substantial disparities in performance. Key future directions are finally discussed, wherein we place self-supervised learning at the center, anticipate an explosion of Transformer-based models for behavioral data, and highlight Federated Learning as an imperative privacy-preserving solution. This review is meant to guide researchers by synthesizing current trends towards the interdisciplinary need to build systems that shall be accurate, scalable, secure and ethically fair.

IndexTerms **- Deep Metric Learning, User Identification.**

INTRODUCTION

In an increasingly digital world, authentication systems play a very important role in security measures against unauthorized access. There is a critical need for reliable user identification in a wide array of applications ranging from secure biometric authentication and continuous user profiling to personalized services and surveillance[8]. With the advancement of technology, new opportunities are created for biometric applications while at the same time significant new security risks are presented.

In most cases, traditional methods of identification relied on static passwords or simple handcrafted features do not operate optimally within such a high-stake environment. The systems are highly vulnerable to impersonation attacks and mostly cannot manage well with high 'intra-class variability' (the same user appears different over time) and 'inter-class similarity' (different users appear similar). This mostly occurs due to variations in the pose of an individual user, environmental illumination, or different data-capturing devices being used.

1.1 Deep Metric Learning (DML)

To meet such great challenges, biometric systems have embraced with ever-increasing fervor the tools of machine learning, and most particularly the techniques of deep learning (DL). Among these methods, Deep Metric Learning (DML) has come forward as a strong and very efficient paradigm[16]. The core objective for DML is not simple classification but rather learning an embedding space-indeed multi-dimensional-in which datapoints from similar identities (i.e., several motion patterns from one and the same user) are close to each other while those from dissimilar identities are kept apart by a large margin.

This focus on learning a similarity metric makes DML highly potent across a spectrum of user identification tasks. Its success has been recorded in already established fields of face recognition and person re-identification (Re-ID)[18]. Also, DML is evolving into the main enabler for new variants of behavioral biometrics. As of late 2023, research has made the first application of "similarity learning" on user identification motion trace data across other disciplines, including eXtended Reality (XR), which has identified over 50,000 unique virtual reality users based solely on their head and hand motion patterns [3, 9]. The application of this learning similarity method on the trace dataset of sequential behavior patterns has led to user identification from web browsing clickstreams [11]. The "similarity learning" method is being integrated into multimodal biometric systems, which combine different data types to create stronger and more secure user authentication systems that incorporate face and dynamic signature data [7].

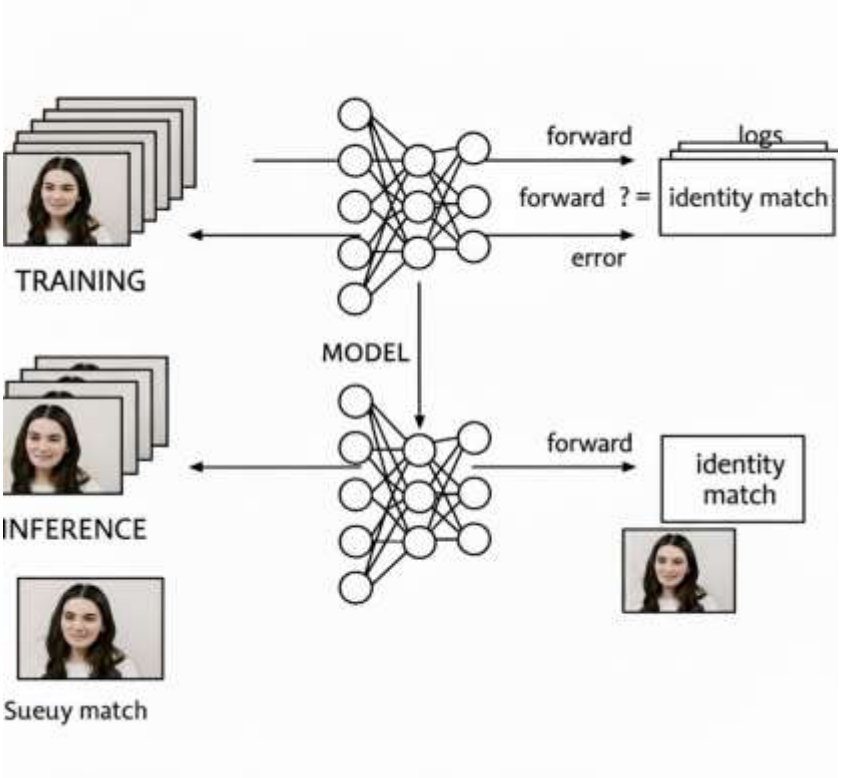

Fig 1.1: deep metric learning

2. Background of DML

Deep Metric Learning (DML) works by creating a high-dimensional embedding space (for instance, capturing details in images or motion data) and learning how to map it to a lower dimensional one. DML aims to achieve an embedding space in which data points positioned close together are semantically related (i.e., close together in meaning) and are, thus, expected to be at a shorter distance from one another than data points that are farther apart.

2.1. Metric Learning

Metric Learning In any application of metric learning, there are two primary components that are important and need to be defined: the metric or distance function which determines how the embeddings in the space are compared and the optimization function which adjusts the space until an optimal configuration is determined[16].

Distance Metrics: The embedding measures and compares dissimilarity of two points which embeddings can be represented as a vector, v_i and v_j. There is a rich literature in DML, however, the most popular embedding metric distance is

Euclidean Distance (L2 Norm): is the shortest distance as the crow flies between two points in the embedding space.

$$D(v_i, v_j) = \|v_i - v_j\|_2 \qquad \text{.....(2.1)}$$

Here n is the dimension of the vectors and v_{ik} and v_{jk} are the k-th components of vectors v_i and v_j respectively.

This is a standard choice for loss functions that aim to pull "similar" points together.

Cosine Similarity: This metric calculates the angular cosine between pairs of vectors, defined as

$$\frac{v_i \cdot v_j}{\|v_i\| \cdot \|v_j\|} \qquad \text{.....(2.2)}$$

The metric is unaffected by the vector. Consequently, it is efficient in tasks such as face verification, wherein the vector's direction is more differentiating than its length.

Loss Functions: The loss function directs the deep learning algorithm on how to arrange the embedding space. Rather than trying to achieve a specific class label (as one would in a classification problem), DML losses seek to minimize distances in a relative manner.

- Contrastive Loss: As you may know, these loss functions work with pairs of samples as either positive pairs (samples from the same identity) or negative pairs (samples from different identities) [16]. The loss will "pull" positive pairs together and "push" apart the negative pairs while maintaining a certain distance from each other, enforcing a margin of separation.

- Triplet Loss: This is the most dominant loss in DML[16]. This loss works with a triplet of samples, an anchor (baseline sample), a positive (another sample belonging to the same identity as the anchor), and a negative (a sample belonging to a different identity from the anchor) [16].The loss function trains the network to push the anchor away from the negative sample more than it does the positive sample which is by a certain margin. This directly works to improve the relative position of the samples in the embedding space and has proven very effective for identification tasks [3]. But also it has issues which we have seen play out in many different versions of the loss function which build upon it for example Meta Prototypical N-Tuple Loss[12] and Triplet Online Instance Matching (TOIM) Loss[15] which are designed to improve its performance and use in identification tasks like person re-identification.

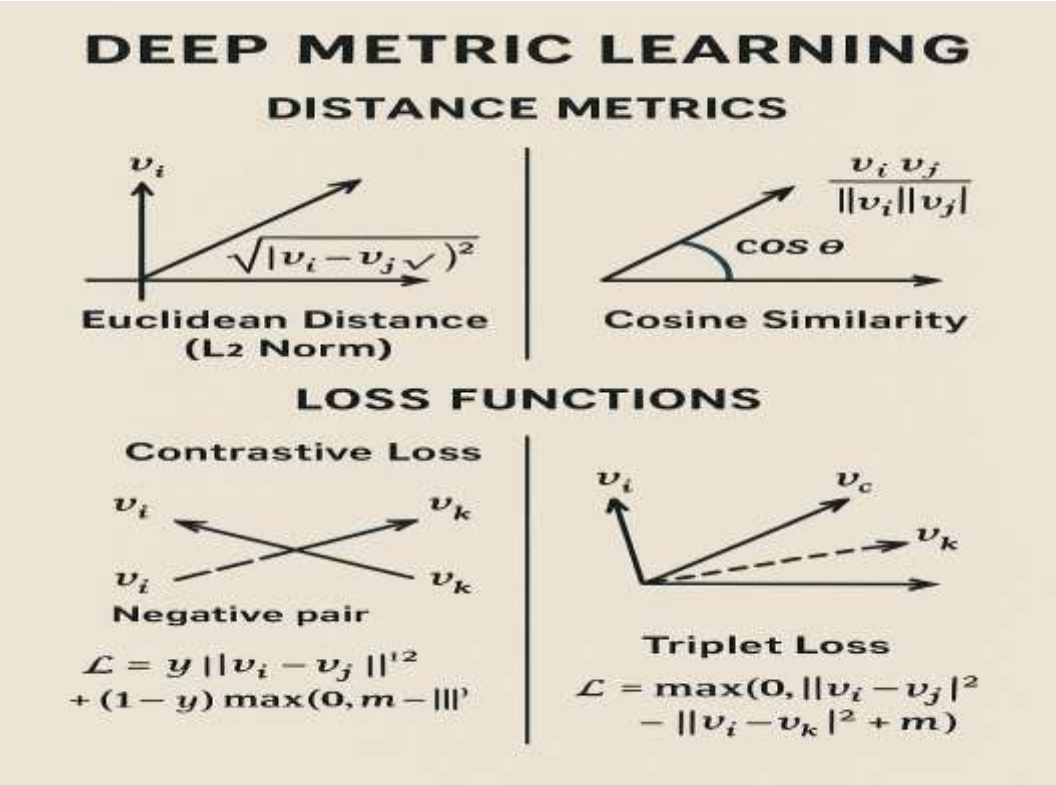

Fig 2.1: deep metric learning: illustrates distance metrics and loss functions for effective embedding space optimization.

2.2. Integration with Deep Learning

Deep Metric Learning's strength is in its use of these loss functions within deep neural networks that serve as very powerful end-to-end feature extractors. The network's base (for example a CNN or Transformer) is trained to work with raw input data and pass it into the embedding space in which the metric learning loss is then determined.

- Convolutional Neural Networks (CNNs): CNNs are a go-to for grid structured data which makes them the workhorse for most image based biometric systems. We see wide use of them in face verification, recognition and image set classification.

- Recurrent Neural Networks (RNNs) and Transformers: As for sequential or temporal data other architectures are better. In this space RNNs like LSTM are used for dynamic biometrics which include online signatures and gait[11]. For example we see RNNs paired with triplet loss which do a great job of learning from sequence data like that of web browsing history for user identification[11]. Also recently we have seen success with Transformer models which use attention mechanisms in bio motion data from head and hand tracking in virtual reality which they use to identify complex temporal patterns.

3. Core Techniques in Deep Metric Learning

No particular individual aspect alone dictates the success of a DML System. It all comes down to how the three components work together: the loss function that determines the DML's objective, the embedding method (network architecture) that features, and the optimization methods that streamline and improve efficiency of the learning.

3.1. Loss Functions

The loss function is the heart of DML. It guides the network in the development of the embedding space which we may think of as a physical representation of the data where in the present context the term "cost" is used to put a value on the present embedding which the loss function then forces into an ideal state that we may define as one of minimal distance between members of the same class (identical identity) and maximal distance between members of different classes (different identity).

- Contrastive Loss: This is a basic, pair based loss. It works on a pair of samples, xi and xj, and their corresponding label y (which is 0 if they are from the same identity and 1 if not). The loss function will push positive pairs (y=0) which are from the same identity together till they touch (zero distance) and at the same time it will push negative pairs (y=1) away until they are at least a distance m apart.

- Triplet Loss: It is probably the most widely used form of DML loss, and it is based on weaknesses of contrastive loss and makes use of three-sample "triplet" which consist of an *anchor* (x_a), a positive during the same identity (x_p), and a negative from a different identity (x_n). The loss works against the network to ensure that the distance between the anchor and the negative is greater than the distance between the anchor and the positive, by at least a margin m[18].This approach has put forth solutions which have been very successful in the domain of behavioral biometrics which include motion patterns [9] and user clickstreams[11]. We see that the shortcoming of the base triplet loss in terms of what it is able to do with respect to selection of proper triplets has been what in large part has driven the development of the more complex loss functions

which extend from it, which includes N-pair loss[16], N-Tuple loss[12] and TOIM loss which are very much at the core of the Re-ID field.

- Proxy-based Losses: In large-scale face recognition problems, when the distinctiveness of the identities increases into the millions, there will come a point where computations will be prohibitive because of the "combinatorial explosion" of possible tuples, a phenomena which is well-known in the literature[16]. Proxy-based techniques avoid this issue by comparing an embedding, not to the other embeddings, but to a learned set of "proxies" or "centers," where each proxy embodies one identity. This changes the computational effort from a difficult sample-to-sample exercise, to a simpler sample-to-proxy exercise.

3.2. Embedding Strategies

The present study reports on what we term a base structure of neural network which is the “backbone” to extract what is differentiating in raw input data. That which designs this backbone is very much at the discretion of the data type.

- Siamese Networks: This is a traditional method of Deep Metric Learning. We use two or more the same (in terms of structure, what they do not) network backbones which at the same time process two or more inputs. Also we apply a contrastive loss to the results which in turn trains the shared backbone to output similar embeddings for which are similar inputs and dissimilar for different inputs[16].

- Triplet Networks: This is a development of the Siamese architecture which we see to use Triplet Loss[16]. We look at three weight sharing backbones that work with the anchor, positive, and negative samples; in this case, the triplet loss is set between the three resultant embedding vectors[11].

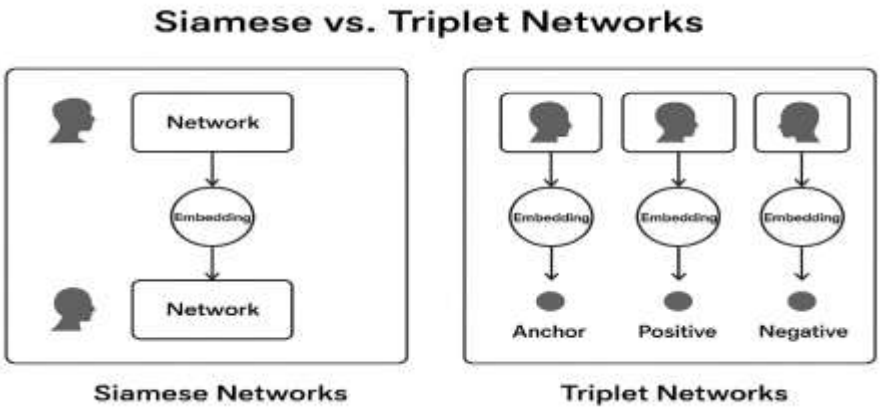

Fig 3.2: specialized backbones : cnn, lstms, spdnets

- Center-based Approaches: it is noted that these designs are made to work with proxy or center based losses. The network does not only train to create a discriminative embedding but also at the same time changes position of what we note as class “centers” in the embedding space which in turn reduces intra-class variation.

- Multitask Learning (MTL): Another key strategy in the field is Multitask Learning (MTL) which has the DML base architecture shared between related tasks[13]. For instance we see models that train for head detection at the same time as they do facial attribute estimation (like age or orientation) which in turn improves overall performance and feature set[13].

- Specialized Backbones: In this, which network to use is a key issue for what data type you have. For image data (for example faces) CNNs are the go to choice. As for temporal data (which may be dynamic signatures or motion data) we see Recurrent Architectures like LSTM and GRU used to identify time dependent trends[11]. Also in very advanced research we see design of custom backbones for certain data types which we may see in SPDNet used for image set classification on SPD manifolds[4].

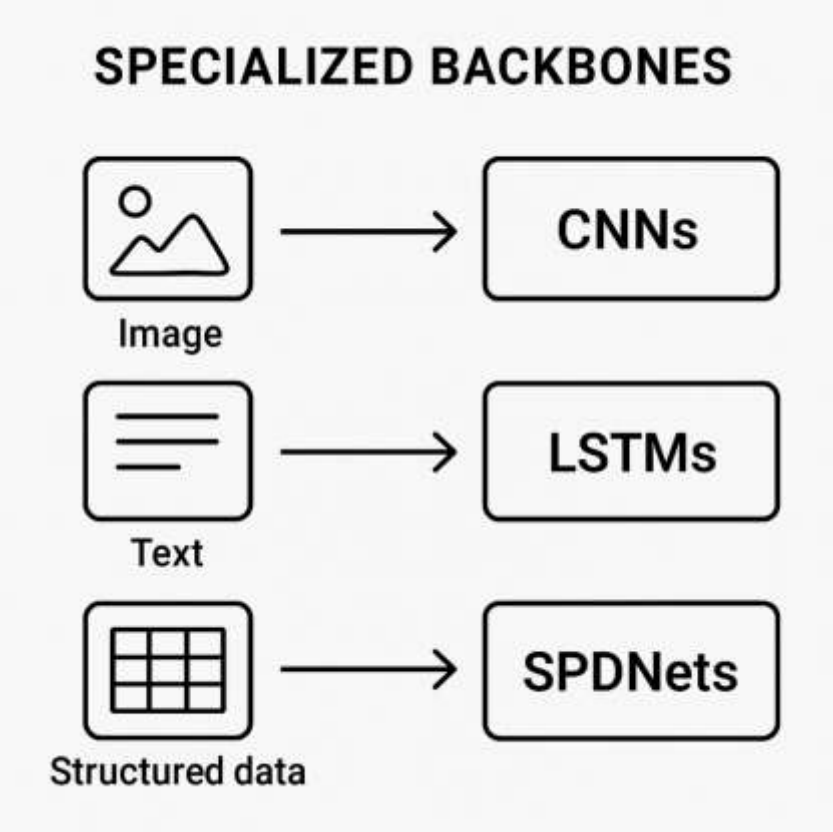

Fig 3.2.1: specialized backbones : cnn, lstms, spdnets

3.3. Optimization Challenges

Training DML models also raises unique challenges that do not appear in regular classification tasks.

- Hard Negative Mining and Sampling: In large datasets, most triplets are deemed 'easy', where the loss surface at that point is zero and hence no training signal can be gained. If the model is only trained on 'easy' triplets, the model will never learn. Hence, the only way to overcome that problem is with 'hard' negative mining, where the sampling method dials in on 'hard' or 'semi-hard' triplets in which the negative sample is incorrectly positioned in close proximity to the anchor[18]. The challenges which come with effective sampling are the driving force in novel loss function designs.[12][15].

- Sampling and Bias: The issue of which sampling method to use is very important in this field which includes issues of performance and also issues of fairness. In a study of face verification systems it was noted that there is a large issue of different performance levels between demographic groups (which may include race, age, and gender) which is a major issue[10]. Also we see that biased sampling during the training process for instance in the case of under-sampling certain groups within the triplet mining may cause or increase these issues of bias.

- Embedding Dimensionality: Choosing at what dimension to output the embedding is a design trade off which is very critical. We see that as we increase the dimension (to 128, 256, or 512) we get better identity separation which in turn gives us better discrimination but at the same time the training and search processes become more compute intensive. As we decrease the dimension we see speed up and also more memory efficiency but at the trade off of possibly not having the "capacity" to fully separate all identities and which in turn will also lower accuracy.

4. Applications in User Identification

Deep Metric Learning (DML) is at the base of what we see in today's identification systems. We have seen a large growth in research related to deep learning in biometrics[2] which reports improved results across many fields. Also it is the case that as a result of DML's success in producing robust features that it is very much at home in the high variable environment of biometric data.

4.1. Person Re-identification (Re-ID)

A key use of Deep Metric Learning is in the domain of Re-ID which is the task of recognizing the same person in nonoverlapping camera views[12][18] which we find to be a very hard metric learning problem which we attribute to differences in pose, lighting and occlusion. Also we see that Triple based losses which include their variations like N-Tuple Loss, TOIM Loss have become the main methods in this field[12][15]. Also a great deal of research is put into unsupervised Re-ID which we see to include use of iterative clustering to develop pseudo labels for fine tuning models on target unsupervised datasets.

4.2. Face Recognition and Attribute Estimation

This is a classic DML application. Landmark DML methods, such as FaceNet and ArcFace, moved the field from simple classification to true similarity-based verification. These models learn a highly discriminative embedding space where all images

of a single individual—regardless of pose, illumination, or expression—are mapped to a tight cluster, separate from all other identities[16].

The issue goes beyond that of just identification which is a function of the model to also include the determination of facial attributes. We see in the multi task learning (MTL) framework which is the case for DML models they at the same time perform face detection and also estimate attributes like age or head pose[13]. Also we note that the success of this approach that uses backbones like Mask-RCNN is in the feature sharing between the detection and estimation tasks[13].

4.3. Behavioral Biometrics

While what you are (physiological biometrics like a face) is the traditional identifier, what you do (behavioral biometrics) puts forth a very powerful and invasive option which is also more continuous in nature [8]. DML does an excellent job at identifying the distinct and very fine grain patterns in human behavior. This includes:

- Keystroke and Signature Dynamics: Identifying a person through their distinctive typing pattern which includes speed and pressure of their dynamic signature [7].
- Voice Identification: Learning out of speech patterns which individual is which (speaker verification).
- Web Browsing and Motion Patterns: We use Deep Machine Learning to develop a unique profile of a user's motion. This has become the main element of Continuous Authentication Systems (CAS)[8]. Also we have seen RNN based DML models used in web browsing histories which we train to learn a "sequence similarity" thus to identify users by their clickstream patterns[11].

4.4. Motion-Based Identification in XR

A new and very recent application of DML is in the field of eXtended Reality (XR) and Virtual Reality (VR) which we are seeing today[3]. In these systems instead of cameras we see the use of high frequency position and rotation data from head mounted displays and hand controllers.

- Unique Identification: Research is reporting that DML models which often use triplet loss are able to develop very powerful similarity metrics from this motion data alone. This was first made known by Miller et al. (2020), who identified 95% of a pool of 511 users based on their typical watching motion[14]. Since then it has grown to identify over 50,000 users[9].
- Versatile Identification: Pretrained models which have been trained for similarity learning are now used to develop what may be called "versatile" identification systems [3] which do away with the need for expensive retraining of each new user and also do well across different sessions, devices and user tasks.
- Behavior Change Measurement: DML based similarity learning is used also to determine changes in a person's motion patterns over time which in turn is a proxy for behavior change[1].

4.5. Multimodal and Cross-Domain Identification

To incept better secure and robust authentication systems we see use of DML which is in fact to combine many biometric modalities. Through the creation of a joint embedding space or application of score level fusion we see these systems' issues related to any one stand alone biometric put to rest[12].

Another obstacle to overcome is User Identity Linkage, which is concerned with connecting user profiles on various platforms (e.g., various social networks)[17].Unsupervised frameworks like CoLink employ DML principles and co-training to align users according to both attribute and relational data[17].This method can approach attribute alignment (e.g., linking "SDE" to "Software Development Engineer") as a sequence-to-sequence translation problem, showcasing DML's versatility beyond basic vector comparison.

5. Datasets and Benchmarks

The issue of Deep Metric Learning models' development and evaluation in user identification is that they are dependent on high quality large scale and diverse public data sets. These benchmarks also which in turn bring about progress and also serve as a common platform to compare the performance and fair play of different algorithms.

5.1. Datasets

- Face Datasets: This is the most mature field in biometrics for DML. We see standards in this area to be:
    - Labeled Faces in the Wild (LFW): A basic dataset for face verification, containing 13,000 images of over 5,000 specific.
    - MS-Celeb-1M & VGGFace2: vast datasets with millions of images across 10.000 (MS-Celeb-1M) and 1000 (VGGFace2) of identities, respectively. The deep-backbone feature extractors used in contemporary face recognition depend on these for training.
    - Multimodal Datasets: Research is increasingly shifting towards multimodal face datasets to overcome the limitations of standard 2D (RGB) images [5]. Recent surveys indicate that these datasets integrate additional modalities, including infrared (IR), depth, and thermal data, to enhance resilience against lighting variations and presentation (spoofing) attacks [6].

- Face Attribute Datasets: Common benchmarks for head pose and attribute estimation include the Prima, BIWI, and UTKFace datasets[13].

- Person Re-Identification (Re-ID) Datasets: The datasets facilitate the matching of individuals across distinct, non-overlapping camera views, a process that significantly depends on DML. Standard benchmarks consist of Market-1501, DukeMTMC-reID, and CUHK03[19].

- Behavioral Datasets: Behavioral biometrics datasets exhibit greater variability and are frequently tailored to specific tasks.

  - Motion & Gait: Analysis is an important emerging field, particularly in the context of motion-based identification utilizing XR/VR data [3]. A significant public dataset in this domain is "Who Is Alyx?", which includes head and hand motion data from users engaged in a VR game [9]. Miller et al. (2020) gathered motion data from 511 participants, showing significant identifiability[14].

  - Keystroke & Signature: Continuous authentication systems make use of numerous keystroke dynamics and dynamic signature datasets [8]. Sometimes researchers have to make their own, like Salturk et al. 's dataset from 2024, which combines dynamic signature data from a standard camera with facial data [7].

  - Sequential Data: Data is frequently sourced from various real-world social networks, such as LinkedIn or internal enterprise networks, for tasks like linking user identities [17]. Large-scale clickstream panel data, like that from Comscore, is used by researchers for web browsing analysis [11].

5.2. Evaluation Metrics

A DML system's assessment is contingent upon its particular use (e.g., verification, identification, or retrieval).

- Rank-1 (or Rank-k) Accuracy: utilized in retrieval and identification tasks (such as Person Re-ID). It calculates the proportion of queries where the right identity appears as the top-ranked result (or within the top-k results) [19].

- mean Average Precision (mAP): mAP, which is also frequently used in retrieval, offers a more thorough metric than Rank-1 accuracy since it assesses the entire ranked list and rewards models that rank all correct matches highly in the results [19].

- Receiver Operating Characteristic (ROC) Curve: This is the standard metric for 1:1 verification tasks, such as face verification. It plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various decision thresholds. The Area Under the Curve (AUC) is a popular performance summary metric.

- Fairness and Bias Metrics: The emphasis has shifted to making sure DML systems are fair as they have become more accurate [10]. Significant performance differences between demographic groups can be concealed by standard accuracy metrics. In order to quantify the variations in error rates (such as False Positive or False Negative rates) for groups based on race, gender, or age, specific metrics are employed, such as Disparate Impact and Disparate Mistreatment.

6. Results

Domain /Application	Representative Work	Dataset(s)	Method/Loss	Parameters (Architecture/Training)	Key Results
Behavior Change in XR	Merz et al. (2025)	Fruit collection in VR(22 participants)	Transformer + Contrastive Metric Learning	GRU + TransformerEncoder; 600-frame sequences; 18 features/frame; >500 hyperparameter sweeps	Cross-condition IER: 54–76%
Multimodal Attention Estimation	Daza et al. (2023, DeepFace-Attention)	mEBAL2	CNN fusion (eye-blink + expression + head pose)	RetinaFace + SAN + CNN modules; SVM + NN fusion; 120s temporal windows	85.9% accuracy; +5–7% over baselines
Person Re-Identification	Zhang et al. (2022, MPN-Tuple)	Market-1501, DukeMTMC, MSMT17, CUHK03	Meta Prototypical N-Tuple Loss	ResNet-50 backbone; Adam optimizer ($lr=8\times10^{-4}$); batch P=16, K=4; 600 epochs	+1–4% mAP gain; SOTA across datasets
Sequential User Identification	Vamosi et al. (2022, TL-RNN)	Comscore clickstream	LSTM + Triplet Loss	LSTM embeddings (512D); α tuned via grid search; L1 distance; 1.5M parameters	98.9% accuracy (200 visits); fast inference

Cross-Domain Multitask Biometrics	Mirzaee Bafti et al. (2022)	BIWI, Prima, UTKFace	Mask R-CNN multitask (pose + age)	ResNet-50 backbone; LR=0.001; batch=1000 ROIs; 10 epochs × 1000 iterations	Pose MAE ~6–8°; Age MAE ~5.3 yrs
Behavioral Biometrics (Mobile)	Wang et al. (2020)	McGill, IDNet, ZJU, Osaka	Siamese + Contrastive + Cross-Entropy	LeNet4/VGG8/MobileNetV2; RMSprop optimizer; margin m=1.5, γ=0.1; 100 epochs	>95% accuracy; 99% brute-force defense

7. Challenges and Open Problems

Deep Metric Learning (DML) for user identification has made great strides, but there are still a number of important obstacles and unsolved research issues. To create systems that are scalable, dependable, secure, and equitable, these issues must be resolved.

7.1. Scalability

The computational and logistical demands of DML systems present a major challenge.

- Computational Complexity: Many DML loss functions, such as those based on triplets or N-pairs, require sophisticated and computationally expensive sampling strategies (e.g., hard negative mining) to function[18]. As the number of identities in a dataset grows, the number of potential pairs or triplets explodes, making training time and resource-intensive[18].

- Logistical Scalability: While many systems show promise, their reliance on specialized or expensive hardware (such as 3D sensors or dedicated scanners) can be "prohibitively expensive and logistically challenging" for widespread online deployment. A key challenge is developing DML models that can perform robustly using common, low-cost sensors like standard computer cameras [7].

7.2. Generalization

Generalization, or "versatility," remains a significant hurdle for the real-world application of DML models [3].

- Cross-Domain Generalization: Models trained in one environment often fail when deployed in another. This "cross-domain generalization problem" [3] is a major hurdle. Much research is focused on Unsupervised Person Re-identification, which attempts to adapt a model trained on one labeled dataset to a new, unlabeled one[19]. This is often done by iterative clustering to create pseudo-labels for fine-tuning, a form of self-paced or unsupervised learning[19]. Similarly, unsupervised frameworks are needed for linking identities across different social networks (UIL), which often have completely different attribute schemas and require co-training or translation-based models[17].

- Extensibility: A major challenge is creating systems that are "extensible to new users without expensive retraining" [3]. Traditional classification models must be fully retrained to add a new user, whereas DML-based similarity learning offers a path toward enrolling new users simply by generating a new embedding, though ensuring this works robustly is an open problem.

7.3. Privacy and Security

The use of biometric data introduces profound privacy and security concerns.

- Data Sensitivity: Biometric data is privacy-sensitive [9]. Unlike a password, biometric templates cannot be revoked or reissued if compromised. Authentication systems must be fortified against exploitation, as they remain "susceptible to exploitation" from impersonation attacks [7].

- Novel Threats: The emergence of new, highly discriminative biometric identifiers, such as head and hand motion, poses "unique security and privacy threats" [9]. As Miller et al. (2020) argue, this data is so fundamentally identifying that traditional privacy-enhancing tools, such as a "private browsing mode," are "in principle impossible"[14]. Furthermore, standard de-identification practices (e.g., removing a user's name) are insufficient, as the motion data itself acts as a "motion signature" that can be used to re-link data across sessions[14].

7.4. Fairness

The unfairness of DML-based identification systems is possibly the most urgent unresolved issue [10].

- Demographic Bias: It has been "found to exhibit significant biases related to race, age, and gender" in deep learning-based systems [10]. The datasets themselves, which may include unequal representations of these demographic groups, are frequently the source of these biases.

- Masked Disparities: The dependence of contemporary methodologies on accuracy as the primary evaluation metric frequently obscures notable demographic disparities in performance. A system may achieve an overall accuracy of 99%, yet exhibit significant deficiencies in performance for particular demographic subgroups. This represents a significant issue as systems are evaluated on increasingly diverse, non-university samples encompassing a broader spectrum of ages and backgrounds[14].

- Intersectional Bias: An area that remains underexplored is intersectional bias, which refers to performance disparities arising from combinations of demographic factors, such as varying age groups within a specific race or gender. The development of DML models and evaluation metrics that ensure fairness across various intersections represents a significant area of research advancement [10].

8. Future Directions

Deep Metric Learning for user identification is an evolving field, characterised by emerging trends aimed at tackling challenges related to scalability, privacy, and generalisation. Future directions indicate the development of systems that exhibit increased robustness, security, and versatility.

8.1. Cross-Modal and Multimodal Fusion

Multimodal biometric systems are present; however, the advanced integration of various data types through DML represents a significant future trend [5][7]. The objective is to develop a cohesive embedding space that effectively integrates physiological and behavioural signals. This encompasses:

- Integrating disparate data: Research is progressing from straightforward image-based fusion (e.g., RGB + IR) to more intricate pairings, such as combining dynamic behavioural signals like dynamic signatures with facial data [7].

- Joint embedding learning: Future DML models will learn a "joint embedding" that maps several modalities (such as face, voice, and motion) into a single, highly discriminative vector [8], a crucial area of research for Continuous Authentication Systems (CAS), rather than simply combining scores at the end. This research is made possible by the creation of extensive multimodal datasets [5].

8.2. Transformer-Based Architectures

Transformer-based models are an obvious future direction, even though CNNs are the norm for images and RNNs (LSTMs/GRUs) are typical for sequential data.

- Attention Mechanisms: Sequential biometric data is increasingly being processed using the "Attention Is All You Need" (Transformer) architecture, which has transformed natural language processing.

- Modeling long-range dependencies: Because transformers are adept at capturing complex, long-range dependencies in temporal data, they are ideal for modeling behavioral biometrics such as head and hand motion or gait patterns over extended periods of time [1]. This is a clear evolution from RNN/LSTM-based models, which have already shown success in modeling sequential data, like web browsing histories [11].

8.3. Self-Supervised and Pretrained Models

To address the data bottleneck and enhance generalization, DML is transitioning from exclusively supervised learning [2].

- Leveraging Unlabeled Data: Self-supervised learning (SSL) enables models to derive insights from extensive volumes of unlabeled biometric data (e.g., hours of motion data) to generate robust feature representations without the need for costly manual annotation.

- Pretrained Similarity Models: Developing "versatile" models that are "pretrained" on sizable, varied datasets (such as "Who Is Alyx?" [9]) and then refined for particular tasks or users [3] is the way of the future. By enabling the enrollment of new users "without expensive retraining," this method solves the "extensibility" problem [3]. In the person Re-ID community, where unsupervised clustering and fine-tuning are essential methods for adapting models to new domains, this is already a major focus [19]. This also holds true for unsupervised user identity linking between various networks [17].

8.4. Privacy-Preserving Metric Learning (Federated Learning)

Given the "unique security and privacy threats" [9] and the "privacy-sensitive" nature of biometric data [9][11], techniques that train models without centralizing user data are essential.

- Federated Learning (FL): FL is a future-critical technique where the DML model is trained on the user's local device (e.g., their phone or VR headset). A central server receives only the model updates, not the unprocessed biometric data. This directly addresses the fundamental privacy issue by enabling the cooperative training of a strong global model without any user's personal data ever leaving their device.

9. Conclusion

The rise of Deep Metric Learning (DML) as a key technology for contemporary user identification has been mapped out in this survey [3][9]. DML has driven a paradigm shift, moving the field beyond traditional handcrafted features to the automated learning of highly discriminative embedding spaces. This approach has not only achieved state-of-the-art accuracy in established domains like face verification [6] and person re-identification[18] but has also unlocked entirely new biometric modalities, such as the identification of users from their unique head and hand motion patterns in virtual reality [9][11] and even sequential web browsing data[11].

The power of DML is demonstrated by the developments detailed in this review, which range from robust multimodal fusion [7] to adaptable pretrained models [3] and innovative loss functions [12][15]. However, this advancement necessitates managing a crucial and intricate balance. Scalability has frequently been sacrificed in the quest for accuracy, with many high-performing systems depending on "prohibitively expensive" specialized hardware or computationally costly models [3].

Additionally, security and privacy are directly at odds with this power [9]. DML models create "unique security and privacy threats" for highly "privacy-sensitive" data as they get better at identifying individuals from new, subtle data types [9][11]. Technical success cannot be the sole criterion, as demonstrated by the problem of demographic bias [10], where high accuracy conceals "significant demographic disparities" in performance.

In the end, no discipline will have an issue with user identification in the future. The results of this survey demonstrate that interdisciplinary approaches are necessary for future advancement. The future of this field is at the nexus of computer vision (to create the models), security (to safeguard the data), and ethics (to guarantee the systems are just, open, and used responsibly).

References

[1] C. Merz, L. Schach, M. L. Fiedler, J. L. Lugrin, C. Wienrich, and M. E. Latoschik, "Unobtrusive in-situ measurement of behavior change by deep metric similarity learning of motion patterns," *arXiv preprint arXiv:2509.04174v1*, 2025.
[2] A. Alduhailan, N. H. Kamarudin, S. N. H. S. Abdullah, and A. Dau, "Deep learning in biometric authentication: Challenges, recent advancements, and future trends," *Journal of Advances in Information Technology*, vol. 16, no. 4, 2025.
[3] C. Rack, K. Kobs, T. Fernando, A. Hotho, and M. E. Latoschik, "Versatile user identification in extended reality using pretrained similarity-learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.
[4] R. Wang, X. J. Wu, Z. Chen, C. Hu, and J. Kittler, "SPD manifold deep metric learning for image set classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 7, pp. 9523–9536, 2024.
[5] K. Al-Mannai, K. Al-Thelaya, J. Schneider, and S. Bakiras, "Multimodal face data sets—A survey of technologies, applications, and contents," *IEEE Access*, vol. 12, pp. 42156–42174, 2024.
[6] R. Daza, L. F. Gomez, J. Fierrez, A. Morales, R. Tolosana, and J. Ortega-Garcia, "DeepFace-attention: Multimodal face biometrics for attention estimation with application to e-learning," *IEEE Access*, vol. 12, pp. 67891–67905, 2024.
[7] S. Salturk and N. Kahraman, "Deep learning-powered multimodal biometric authentication: Integrating dynamic signatures and facial data for enhanced online security," *Neural Computing and Applications*, vol. 36, pp. 11311–11322, 2024.
[8] S. Ayeswarya and K. J. Singh, "A comprehensive review on secure biometric-based continuous authentication and user profiling," *IEEE Access*, vol. 12, pp. 28156–28175, 2024.
[9] V. Nair, W. Guo, J. Mattern, R. Wang, J. F. O'Brien, L. Rosenberg, and D. Song, "Unique identification of 50,000+ virtual reality users from head & hand motion data," in *Proceedings of the 32nd USENIX Security Symposium*, August 2023, pp. 847–863.
[10] I. Sarridis, S. Papadopoulos, C. Koutlis, and C. Diou, "Towards fair face verification: An in-depth analysis of demographic biases," *arXiv preprint arXiv:2307.10011v1*, 2023.
[11] Vamosi, S., Reutterer, T., & Platzer, M. (2022). A deep recurrent neural network approach to learn sequence similarities for user-identification. *Decision Support Systems, 155*, 113718.
[12] Zhang, Z., Lan, C., Zeng, W., Chen, Z., & Chang, S.-F. (2022). Beyond Triplet Loss: Meta Prototypical N-Tuple Loss for Person Re-identification. *IEEE Transactions on Multimedia, 24*, 4158-4169.
[13] Bafti, S. M., Chatzidimitriadis, S., & Sirlantzis, K. (2022). Cross-Domain Multitask Model for Head Detection and Facial Attribute Estimation. *IEEE Access, 10*, 54703-54712.
[14] M. R. Miller, F. Herrera, H. Jun, J. A. Landay, and J. N. Bailenson, "Personal identifiability of user tracking data during observation of 360-degree VR video," *Scientific Reports*, vol. 10, no. 1, p. 17404, 2020.
[15] Li, Y., Yin, G., Liu, C., Yang, X., & Wang, Z. (2020). Triplet Online Instance Matching Loss for Person Re-identification. *IEEE Access* (from arXiv:2002.10560v1).
[16] Kaya, M., & Bilge, H. Ş. (2019). Deep Metric Learning: A Survey. *Symmetry, 11*(9), 1066.
[17] Zhong, Z., Cao, Y., Guo, M., & Nie, Z. (2018). CoLink: An Unsupervised Framework for User Identity Linkage. In *The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)* (pp. 5714-5721).
[18] Hermans, A., Beyer, L., & Leibe, B. (2017). In Defense of the Triplet Loss for Person Re-Identification. *arXiv preprint arXiv:1703.07737v4*.
[19] Fan, H., Zheng, L., & Yang, Y. (2017). Unsupervised Person Re-identification: Clustering and Fine-tuning. *arXiv preprint arXiv:1705.10444v2*.