Review

Machine Learning-Powered Hand Recognition: Techniques, Evaluation, and Open Problems

Deodato Jiahao , Luisinho Carla * , Valente Carmo

Posted Date: 19 May 2025

doi: 10.20944/preprints202505.1465.v1

Keywords: hand recognition; hand pose estimation; gesture recognition; machine learning; deep learning; convolutional neural networks; graph neural networks; temporal modeling; multimodal fusion; evaluation metrics; self-supervised learning; 3D hand modeling; real-time inference

 doi:10.20944/preprints202505.1465.v1

Article

Machine Learning-Powered Hand Recognition: Techniques, Evaluation, and Open Problems

Deodato Jiahao, Luisinho Carla * and Valente Carmo

Department of Computer Science, University of Lisbon; deodato.jiahao@ulisboa.pt (D.J.); valente.carmo@ulisboa.pt (V.C.)
* Correspondence: luisinho.carla@ulisboa.pt

Abstract: Hand recognition and analysis has emerged as a cornerstone of human-computer interaction, enabling a wide array of applications including gesture-based interfaces, sign language interpretation, virtual and augmented reality, and biometric authentication. Recent advances in machine learning, particularly deep learning, have significantly elevated the capabilities of hand recognition systems by allowing them to learn complex visual and kinematic patterns from large-scale datasets. This survey provides a comprehensive examination of the field, beginning with foundational representations and computational models used for detecting and understanding hand configurations in 2D and 3D. We systematically explore the use of convolutional neural networks (CNNs), graph neural networks (GNNs), recurrent models, and transformer-based architectures, highlighting their effectiveness in various subdomains such as static gesture classification, dynamic gesture recognition, pose estimation, and segmentation. We also present an in-depth discussion of evaluation metrics and experimental protocols commonly used to benchmark performance across different tasks. These include accuracy, precision, recall, mean squared error, percentage of correct keypoints (PCK), intersection-over-union (IoU), and temporal metrics such as edit distance and sequence accuracy. A summary of widely adopted datasets is included, along with a comparative analysis of state-of-the-art results. Despite these advancements, the field continues to face significant challenges related to occlusions, intra- and inter-class variation, domain adaptation, data scarcity, and computational efficiency. We analyze these limitations in detail and review emerging research directions such as self-supervised learning, multimodal fusion, efficient model design for edge deployment, and generative approaches for data synthesis. We further examine the ethical considerations and fairness implications associated with deploying hand recognition technologies in real-world environments. This survey concludes with a synthesis of the current state of the field and a forward-looking perspective on its trajectory. We argue that future progress will require interdisciplinary solutions that combine algorithmic innovation with robust evaluation, ethical deployment, and user-centric design. The insights presented herein aim to inform and inspire future research at the intersection of computer vision, machine learning, and human-centered computing.

Keywords: hand recognition; hand pose estimation; gesture recognition; machine learning; deep learning; convolutional neural networks; graph neural networks; temporal modeling; multimodal fusion; evaluation metrics; self-supervised learning; 3D hand modeling; real-time inference

1. Introduction

The field of hand recognition and analysis has emerged as a critical component in the broader context of human-computer interaction (HCI), biometric authentication, sign language interpretation, virtual and augmented reality (VR/AR), and human activity recognition [1]. Recognizing and interpreting human hand gestures and features in real time is a complex task that requires sophisticated techniques to address challenges such as variability in hand shapes, dynamic poses, occlusions, background clutter, and lighting conditions [2]. With the advent of machine learning, particularly deep learning, substantial progress has been made in enabling systems to automatically learn discriminative

 doi:10.20944/preprints202505.1465.v1

features from large datasets, thereby improving the robustness and accuracy of hand recognition and analysis algorithms. Hand recognition encompasses a variety of tasks including hand detection, hand pose estimation, gesture classification, and hand shape modeling [3]. Traditionally, these tasks were approached using handcrafted features and rule-based heuristics, which often lacked generalizability and failed under real-world variability. The shift toward data-driven approaches has enabled more adaptive and scalable solutions. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and more recently, Transformer-based architectures have shown significant promise in learning spatial and temporal representations of hand movements [4]. Moreover, the integration of 2D and 3D imaging modalities, including depth sensors and multi-camera systems, has enriched the input data used for training machine learning models, resulting in higher fidelity analyses [5]. The relevance of machine learning to hand analysis is further emphasized by its application in real-time systems where performance, speed, and accuracy are critical [6]. For instance, in sign language recognition, models must capture subtle hand articulations and transitions between gestures in a continuous video stream [7]. Similarly, in biometric applications, precise identification of individuals based on unique hand features requires high accuracy and resilience to spoofing attacks [8]. Machine learning methods, particularly those leveraging large annotated datasets, have demonstrated notable success in addressing these demands through transfer learning, data augmentation, and synthetic data generation. In recent years, the research community has also explored the potential of unsupervised and semi-supervised learning techniques to reduce the dependency on labeled data, which is often expensive and labor-intensive to collect [9]. Generative models, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), have been employed to synthesize realistic hand poses and gestures, aiding in data augmentation and simulation. Furthermore, the use of reinforcement learning in gesture-controlled interfaces and robotics has introduced new paradigms for learning hand interactions in dynamic environments [10]. This survey aims to present a structured and detailed examination of the current landscape in hand recognition and analysis using machine learning [11]. It categorizes the literature based on key technical challenges and solutions, evaluates the performance of state-of-the-art systems, and identifies open problems and future directions. The subsequent sections delve into the core components of hand analysis pipelines, from image acquisition and preprocessing to model training, evaluation metrics, and deployment considerations [12]. Through this comprehensive overview, we intend to provide researchers and practitioners with a solid foundation for understanding the capabilities and limitations of machine learning approaches in this domain [13].

2. Mathematical Foundations and Modeling of Hand Recognition

The task of hand recognition and analysis can be formally defined as a mapping problem where the goal is to estimate a function $f : \mathcal{X} \to \mathcal{Y}$, such that for an input image or sequence $\mathbf{x} \in \mathcal{X}$, the model predicts a label or structured output $\mathbf{y} \in \mathcal{Y}$. Depending on the specific task, $\mathbf{y}$ may represent a class label (e.g., gesture class), a set of 2D or 3D coordinates corresponding to hand keypoints, or a temporal sequence of poses. This mapping is typically approximated using a parameterized model f_θ, where θ denotes the learnable parameters of the machine learning model. The optimal parameters are obtained by minimizing a loss function $\mathcal{L}(\mathbf{y}, f_\theta(\mathbf{x}))$ over a training set $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N}$ using optimization techniques such as stochastic gradient descent (SGD). For static hand gesture recognition, the input space $\mathcal{X}$ is often composed of single-frame RGB or depth images, while the output space $\mathcal{Y}$ consists of discrete gesture labels. In this case, the problem reduces to a standard classification task, where the cross-entropy loss is widely used:

$$\mathcal{L}_{\text{CE}} = -\sum_{i=1}^{N}\sum_{c=1}^{C} \mathbf{y}_{i,c} \log f_{\theta,c}(\mathbf{x}_i),$$

where C is the number of gesture classes and $\mathbf{y}_{i,c}$ is the one-hot encoded ground truth for class c. For pose estimation, the model predicts a vector of keypoint coordinates $\mathbf{y}_i = [x_1, y_1, \ldots, x_K, y_K]$, and the loss function is typically defined as the Mean Squared Error (MSE):

 doi:10.20944/preprints202505.1465.v1

$$\mathcal{L}_{\text{MSE}} = \frac{1}{NK} \sum_{i=1}^{N} \sum_{k=1}^{K} \|\mathbf{y}_{i,k} - f_{\theta,k}(\mathbf{x}_i)\|^2,$$

where K is the number of keypoints and $\mathbf{y}_{i,k}$ represents the ground truth coordinates for keypoint k [14]. Deep learning-based models such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Transformer encoders have become dominant tools for modeling f_θ. Each architecture introduces different inductive biases and computational characteristics, as summarized in Table 1.

Table 1. Comparison of Common Machine Learning Models for Hand Recognition

Model Type	Spatial Modeling	Temporal Modeling	Computational Cost
CNN (e.g., ResNet, VGG)	Strong	Weak	Moderate
RNN/LSTM	Weak	Strong	High
3D CNN	Strong	Moderate	Very High
Transformer	Strong	Strong	Very High
Hybrid CNN+RNN	Strong	Strong	High

To better understand the information flow and modular structure of a typical hand recognition system, we can abstract the pipeline into the schematic diagram shown in Figure 1 [15]. The pipeline begins with raw input acquisition, which may involve RGB, depth, or multi-modal sensors [16]. This input is passed through a feature extractor $\phi(\cdot)$, often a deep CNN, yielding a compact representation $\mathbf{z} = \phi(\mathbf{x})$ [17]. Depending on the task, this representation is then fed into a classifier $g(\cdot)$ or regressor to produce the final prediction $\hat{\mathbf{y}} = g(\mathbf{z})$ [18].

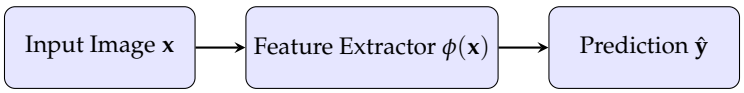

Figure 1. Generalized hand recognition pipeline based on machine learning.

The design of ϕ and g is task-dependent. For gesture classification, g is typically a fully connected softmax layer [19]. For pose estimation, g may be a dense regression head or a heatmap decoder. Moreover, temporal modeling for dynamic gestures is achieved by integrating temporal sequences $\{\mathbf{x}_t\}_{t=1}^{T}$ and leveraging sequential models such as LSTMs or temporal convolutions [20]. The overall loss function in these cases may combine classification and sequence modeling objectives:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{cls}} \mathcal{L}_{\text{CE}} + \lambda_{\text{temp}} \mathcal{L}_{\text{seq}},$$

where λ_{cls} and λ_{temp} control the trade-off between static and temporal components [21]. The flexibility of this formulation allows researchers to tailor their models to specific application domains, whether in static biometric recognition, continuous sign language understanding, or interactive control interfaces [22]. In conclusion, mathematical modeling plays a foundational role in the design and training of hand recognition systems [23]. The choices of input representation, model architecture, and objective functions significantly affect the performance and generalization ability of the resulting systems. Subsequent sections will delve deeper into practical implementations, benchmarking datasets, and performance evaluations of these methods in real-world scenarios [24].

3. Datasets and Benchmarking Protocols

The development and evaluation of hand recognition systems are intrinsically linked to the availability and quality of datasets. Robust benchmarking requires large, diverse, and well-annotated datasets that reflect the real-world variability in hand poses, gestures, backgrounds, occlusions, lighting conditions, and sensor modalities [25]. In this section, we survey widely-used datasets that have

 doi:10.20944/preprints202505.1465.v1

become standard in training and evaluating machine learning models for hand recognition and analysis [26]. We also examine the benchmarking protocols associated with these datasets, including train-test splits, evaluation metrics, and common preprocessing pipelines. Hand gesture datasets can be broadly categorized based on their data modality—RGB, depth, infrared, or multi-modal (e.g., RGB-D)—as well as on the nature of the task they support: static hand pose estimation, dynamic gesture recognition, sign language interpretation, or biometric identification. For example, the American Sign Language Lexicon Video Dataset (ASLLVD) provides multi-view videos of fingerspelling and sign gestures with frame-level annotations, making it highly suitable for temporal models. In contrast, the SHREC dataset emphasizes depth-based hand gestures in a controlled environment and is frequently used to evaluate 3D recognition capabilities. Similarly, datasets such as EgoHands and GTEA utilize egocentric views, offering realistic scenarios for gesture analysis in first-person video streams. Table 2 summarizes several benchmark datasets frequently cited in the literature [27]. For each dataset, we report the type of data modality, number of gesture or pose classes, approximate number of samples, and key applications [28].

Table 2. Summary of Key Datasets for Hand Recognition and Analysis

Dataset	Modality	# Classes	# Samples	Applications
ASLLVD	RGB + Skeleton	∼3000 signs	∼10K videos	Sign language recognition
SHREC'17	Depth	14	∼2800 sequences	Dynamic hand gesture recognition
EgoHands	RGB (egocentric)	4	48 video sequences	Hand segmentation, detection
Dexter 1	RGB + Depth	N/A	∼3000 frames	Hand pose estimation
GTEA	RGB	N/A	28 videos	Activity recognition with hands
HandNet	RGB-D + 3D joints	Continuous	∼200K frames	3D hand pose tracking

Each dataset follows its own annotation standard and collection protocol, which poses challenges in developing generalizable models across datasets [29]. For example, some datasets annotate hand joints using 21-point skeletal models (similar to the one used in the MSRA dataset), while others provide bounding boxes, segmentation masks, or raw RGB-D frames without any skeletal markup [30]. This heterogeneity requires dataset-specific preprocessing and model adaptation to align representations [31]. Transfer learning and domain adaptation techniques are increasingly employed to address this mismatch and to leverage information from multiple sources [32]. Benchmarking protocols also vary significantly. Some datasets define strict train/test splits to ensure reproducibility, while others allow user-defined partitions [33]. For instance, the SHREC'17 challenge specifies Leave-One-Subject-Out (LOSO) evaluation to test model generalization across different users [34,35]. In contrast, datasets such as ASLLVD may use a signer-independent split, training on a subset of individuals and testing on unseen signers [36]. Furthermore, certain tasks involve frame-level metrics such as per-joint mean squared error (MSE) for pose estimation, while others rely on classification accuracy, top-k accuracy, or F1-score for gesture recognition [37]. Standardized evaluation is essential for fair comparison across methods [38]. However, inconsistencies in preprocessing (e.g., hand cropping, scaling, background removal), resolution normalization, and temporal alignment often lead to subtle variations in reported results. To mitigate this, several research initiatives have proposed unified frameworks and toolkits that enforce consistent data loading and metric computation. These toolkits are increasingly integrated with deep learning libraries such as PyTorch and TensorFlow to facilitate reproducible experimentation. In summary, datasets are the cornerstone of progress in machine learning-based hand recognition. Despite their variety and richness, challenges remain in aligning their annotations, formats, and evaluation standards. A continued effort toward standardized benchmarks, unified protocols, and open-source pipelines will be critical in advancing this field and enabling fair and meaningful comparisons among state-of-the-art systems [39].

 doi:10.20944/preprints202505.1465.v1

4. Evaluation Metrics and Performance Analysis

Evaluating the performance of hand recognition and analysis systems necessitates the careful selection of appropriate metrics that reflect the nature of the underlying task—whether it is classification, regression, segmentation, or sequence prediction. Accurate and meaningful evaluation enables fair comparison between methods, guides model selection, and provides insights into areas needing improvement [40]. In this section, we systematically present commonly adopted evaluation metrics across various subdomains of hand recognition and analyze their implications in practical settings [41]. For static hand gesture classification, the most widely used metric is overall accuracy, defined as the proportion of correctly predicted labels to the total number of samples:

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(f_\theta(\mathbf{x}_i) = \mathbf{y}_i),$$

where $\mathbb{1}(\cdot)$ is the indicator function, $f_\theta(\mathbf{x}_i)$ is the predicted label, and $\mathbf{y}_i$ is the ground truth label for the i-th sample [42]. While accuracy is intuitive, it can be misleading in imbalanced datasets where some gesture classes dominate. To address this, precision, recall, and the F1-score are employed on a per-class basis:

$$\text{Precision}_c = \frac{\text{TP}_c}{\text{TP}_c + \text{FP}_c}, \quad \text{Recall}_c = \frac{\text{TP}_c}{\text{TP}_c + \text{FN}_c}, \quad \text{F1}_c = \frac{2 \cdot \text{Precision}_c \cdot \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c},$$

where TP_c, FP_c, and FN_c denote true positives, false positives, and false negatives for class c, respectively [43]. These metrics can be aggregated using macro- or micro-averaging to provide a global performance score. For continuous hand pose estimation, particularly in 2D or 3D space, the most prevalent metric is the Mean Squared Error (MSE) over all predicted keypoints [44]. Given predicted joint coordinates $\hat{\mathbf{y}} = [\hat{x}_1, \hat{y}_1, \ldots, \hat{x}_K, \hat{y}_K]$ and corresponding ground truth $\mathbf{y}$, the per-joint MSE is:

$$\text{MSE} = \frac{1}{K}\sum_{k=1}^{K} \|(\hat{x}_k, \hat{y}_k) - (x_k, y_k)\|^2,$$

where K is the number of keypoints. In 3D hand pose estimation, the Euclidean distance is extended to three dimensions [45]. Some benchmarks also report the Percentage of Correct Keypoints (PCK), defined as the proportion of keypoints within a threshold distance δ from the ground truth:

$$\text{PCK} = \frac{1}{N \cdot K}\sum_{i=1}^{N}\sum_{k=1}^{K} \mathbb{1}(\|\hat{\mathbf{y}}_{i,k} - \mathbf{y}_{i,k}\| < \delta).$$

This metric provides interpretable insight into spatial precision, often visualized as PCK curves over varying thresholds. In the context of dynamic hand gestures and sign language recognition, models must not only recognize static configurations but also capture temporal dependencies [46]. Thus, sequence-level evaluation becomes essential [47]. Metrics such as sequence classification accuracy, Levenshtein distance (edit distance), and frame-wise accuracy are commonly reported [48]. Let $\mathbf{Y} = [y_1, y_2, \ldots, y_T]$ and $\hat{\mathbf{Y}} = [\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_T]$ be the ground truth and predicted sequences, respectively [49]. The normalized Levenshtein distance is given by:

$$\text{Edit Distance} = \frac{\text{LD}(\mathbf{Y}, \hat{\mathbf{Y}})}{\max(|\mathbf{Y}|, |\hat{\mathbf{Y}}|)},$$

where $\text{LD}(\cdot, \cdot)$ denotes the minimum number of insertion, deletion, and substitution operations required to transform one sequence into the other [50]. A lower score indicates higher temporal fidelity in gesture prediction [51]. Evaluation for hand segmentation and detection tasks typically uses metrics from object detection and semantic segmentation literature [52]. These include Intersection over Union (IoU), Average Precision (AP), and mean Average Precision (mAP). IoU for a predicted mask $\hat{M}$ and ground truth mask M is defined as:

 doi:10.20944/preprints202505.1465.v1

$$\text{IoU} = \frac{|\hat{M} \cap M|}{|\hat{M} \cup M|},$$

which evaluates the pixel-wise overlap between predicted and actual hand regions [53]. AP scores are calculated at multiple IoU thresholds (e.g., 0.5, 0.75), and mAP is the mean across all thresholds and classes [54]. The selection of metrics is inherently task-specific and should align with the operational goals of the hand recognition system. For instance, in real-time HCI systems, latency and frame-wise throughput are also important and must be evaluated alongside accuracy [55]. Some works report frames per second (FPS) and end-to-end inference time to characterize the deployability of their models [56]. Additionally, ablation studies are often conducted to assess the contribution of individual model components or data augmentation strategies to overall performance. In summary, rigorous evaluation of hand recognition models requires a multi-faceted approach that combines classification, regression, sequence analysis, and real-time efficiency metrics. As models become increasingly sophisticated and are applied to diverse tasks and platforms, a nuanced understanding of these evaluation tools will remain central to driving meaningful progress in the field [57].

5. Challenges and Limitations in Hand Recognition Using Machine Learning

Despite the rapid advancement of machine learning techniques, hand recognition and analysis remains a fundamentally challenging problem due to the high degrees of variability, ambiguity, and complexity inherent to human hand motion and appearance [58]. In this section, we examine the primary challenges and limitations faced by current systems, drawing attention to both theoretical constraints and practical bottlenecks that hinder generalization, robustness, and scalability [59]. One of the foremost challenges lies in the variability of hand appearance across individuals [60]. Differences in hand shape, skin tone, size, and texture introduce significant intra-class variation, making it difficult for models to generalize well without large-scale, diverse training datasets. Moreover, the presence of accessories (e.g., rings, watches, sleeves) and occlusions (e.g., overlapping fingers, object interactions) often result in partial visibility, which confounds keypoint detection and gesture classification. Even with advanced convolutional backbones, such variability can lead to brittle predictions when models are tested on unseen users or environments. Another major challenge is the dynamic and articulated nature of the hand [61]. With over 20 degrees of freedom, the human hand is capable of producing an enormous range of poses and gestures, many of which differ subtly [62]. Capturing fine-grained articulations requires high-resolution input and precise annotation, which are often lacking in real-world datasets [63]. Furthermore, similar global hand poses may correspond to entirely different gestures depending on finger configurations or contextual cues, leading to high inter-class ambiguity. This calls for models that can incorporate spatial hierarchies and relational dependencies between joints, a task that remains nontrivial especially under computational constraints [64]. From a temporal perspective, dynamic gesture recognition is subject to additional challenges [65]. Variability in gesture execution speed, non-uniform transitions, and co-articulation effects—where gestures blend into each other—make it difficult to segment and classify gestures reliably in continuous streams [66]. Temporal models like LSTMs and Transformers partially alleviate these issues, but they demand significant computational resources and large quantities of sequentially labeled data, which are costly to obtain and annotate accurately [67]. Another limitation is domain generalization and adaptation. Models trained on a specific dataset often exhibit degraded performance when applied to different domains, such as from controlled lab settings to in-the-wild environments [68]. This domain shift arises from differences in lighting conditions, camera quality, background clutter, and sensor types [69]. While domain adaptation techniques—such as adversarial training and domain-invariant feature learning—offer promising directions, they are not yet universally effective or easy to implement in real-time systems [70]. In addition, the annotation burden is a considerable bottleneck. High-quality annotations for hand keypoints, masks, and gesture sequences are expensive to produce and often require expert supervision. Manual annotation is time-consuming, while automatic or semi-automatic

 doi:10.20944/preprints202505.1465.v1

tools often suffer from accuracy and consistency issues. This scarcity of labeled data restricts the performance ceiling of supervised learning models, particularly in tasks like 3D hand pose estimation, where obtaining ground truth requires complex sensor setups like motion capture systems. Real-time performance and computational efficiency are also persistent challenges [71]. Many state-of-the-art models achieve high accuracy but at the cost of latency and memory consumption, making them unsuitable for embedded systems or edge devices. For instance, 3D CNNs and Transformer-based models tend to be computationally intensive and may not meet the stringent timing requirements of interactive applications. Trade-offs between accuracy and speed must be carefully managed, often requiring lightweight architectures, pruning, or model quantization—all of which can degrade model precision if not executed judiciously [72]. Ethical and privacy considerations also emerge in the context of hand recognition, particularly when deployed in surveillance, biometric authentication, or assistive technologies [73]. The collection and usage of hand data raise concerns regarding user consent, data security, and potential biases encoded in the model. For example, models trained predominantly on data from a limited demographic may fail to perform equitably across diverse populations, leading to biased outcomes or exclusion [74]. In summary, the field of hand recognition using machine learning faces multifaceted challenges, encompassing data limitations, model generalization, real-time processing, and ethical deployment [75]. Overcoming these hurdles requires interdisciplinary collaboration, innovations in data acquisition and labeling, and the development of algorithms that are not only accurate but also interpretable, efficient, and fair [76]. Addressing these challenges holistically is critical to the successful integration of hand recognition systems in real-world applications [77].

6. Future Directions and Emerging Trends

As the field of hand recognition and analysis using machine learning continues to evolve, several emerging directions and technological trends point toward transformative possibilities [78]. These future avenues are driven by both the limitations of current systems and the growing demand for intelligent, responsive, and human-centered computing interfaces. In this section, we outline key research directions likely to define the next generation of hand recognition systems [79]. One promising trend is the development of **multimodal learning frameworks** that combine visual, depth, inertial, and audio data to build more robust and context-aware models. Multimodal fusion allows for the disambiguation of complex hand gestures and enhances resilience against occlusions, poor lighting, and sensor noise [80]. For instance, coupling RGB frames with depth maps or inertial motion unit (IMU) signals provides complementary views of hand dynamics that are especially useful in wearable or egocentric applications. Attention-based fusion models and cross-modal transformers are gaining traction for their ability to selectively integrate features across modalities in a task-adaptive manner. Another future direction lies in **self-supervised and unsupervised learning**, which aim to alleviate the dependence on labor-intensive annotations [81]. Recent advancements in contrastive learning, masked autoencoding, and temporal consistency learning have enabled models to learn semantically rich representations from unlabeled hand data [82]. These techniques not only reduce data labeling costs but also improve generalization, particularly when deployed in new environments or across different user demographics [83]. Pretraining on large-scale, diverse datasets using self-supervised objectives, followed by lightweight finetuning, is expected to become a standard pipeline in this domain [84]. **Generative modeling** is another area of growing interest [85]. Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion models have demonstrated their capacity to synthesize realistic hand poses and gesture sequences, which can be used for data augmentation, simulation, and synthetic dataset generation [86]. By leveraging such models, researchers can overcome data scarcity and produce training samples with controllable attributes such as pose, lighting, and camera angle [87]. In particular, conditional GANs and pose-guided diffusion models show promise for generating photorealistic hand images paired with precise joint annotations. **3D hand modeling and reconstruction** will continue to advance, fueled by improved depth estimation algorithms and parametric hand mesh representations such as MANO [88]. The shift toward

 doi:10.20944/preprints202505.1465.v1

mesh-based understanding, rather than sparse keypoints, facilitates a richer geometric analysis of hand surfaces and allows integration with physical simulation engines for fine-grained interaction modeling [89]. Coupling 3D hand tracking with scene understanding and object interaction models opens the door to full-fledged human-object interaction (HOI) systems, which are essential for applications in robotics, AR/VR, and assistive technology [90]. **Real-time and on-device inference** will also be a focal point of future research, driven by the need to deploy hand recognition systems in mobile, wearable, or resource-constrained settings. Efficient neural architectures such as MobileNet, ShuffleNet, and Transformer-Lite variants are being explored for maintaining performance under strict latency and power constraints. Techniques like knowledge distillation, quantization, pruning, and neural architecture search (NAS) are likely to be more widely adopted to strike an optimal balance between model size and accuracy [91]. On the algorithmic side, **causal and continual learning** approaches are gaining attention to address challenges associated with temporal dynamics and lifelong adaptation. Gesture understanding systems must be capable of learning from streaming data, adapting to new users, and incorporating feedback without catastrophic forgetting [92]. Few-shot and zero-shot learning paradigms, enabled by meta-learning frameworks, will help systems generalize to novel gestures and users with minimal supervision [93]. From an application perspective, the integration of hand recognition into **AR/VR interfaces, sign language translation, remote collaboration, and touchless control systems** will continue to expand [94]. These systems demand a seamless fusion of spatial accuracy, temporal coherence, and intuitive user experience. Advances in haptic feedback and embodied AI will further enrich interaction modalities, allowing users to engage with digital environments in natural and immersive ways. Finally, **fairness, transparency, and accountability** in hand recognition systems will become increasingly important as these models are embedded into sensitive applications such as healthcare, surveillance, and education [95]. Future research must prioritize model interpretability, robust evaluation across diverse populations, and mechanisms for user consent and feedback [96]. Standardized benchmarks that account for fairness, privacy, and reliability will help align technological progress with ethical principles [97]. In conclusion, the future of hand recognition using machine learning is poised to benefit from a convergence of algorithmic innovation, hardware acceleration, and interdisciplinary collaboration [98]. By addressing current limitations and embracing emerging methodologies, the field can advance toward building intelligent, inclusive, and context-aware systems capable of understanding and responding to the full richness of human hand behavior [99].

7. Conclusions

Hand recognition and analysis through machine learning has evolved into a pivotal subfield within computer vision and human-computer interaction, offering a rich spectrum of applications that span from gesture-based interfaces to sign language translation, robotics, augmented reality, and biometric authentication. This survey has comprehensively examined the landscape of hand recognition by reviewing core techniques, datasets, evaluation protocols, challenges, and emerging trends.

From a methodological perspective, machine learning—particularly deep learning—has dramatically enhanced the precision and generalization capabilities of hand recognition systems. Techniques such as convolutional neural networks, recurrent architectures, graph-based models, and more recently, transformers, have been successfully adapted to capture the spatial, temporal, and kinematic characteristics of human hand motion. These advances have been further empowered by the availability of annotated datasets and increasingly powerful computational resources.

Despite the progress, several enduring challenges remain. Issues such as occlusion, intra-class variability, domain generalization, annotation costs, and real-time performance constraints continue to limit the deployment of hand recognition systems in unconstrained environments. In response, emerging approaches such as self-supervised learning, multimodal fusion, 3D mesh modeling, and efficient on-device inference present promising avenues to overcome these limitations and bridge the gap between research prototypes and deployable systems.

 doi:10.20944/preprints202505.1465.v1

Looking forward, the future of hand recognition is expected to be shaped by interdisciplinary collaboration that unites insights from computer vision, signal processing, cognitive science, and human-computer interaction. The integration of machine learning with novel sensing modalities, such as event cameras, bio-signal acquisition devices, and wearable IMUs, will further enhance the capacity to interpret nuanced hand behaviors with contextual awareness. At the same time, researchers must remain attentive to issues of fairness, inclusivity, and transparency, ensuring that hand recognition technologies are accessible, trustworthy, and respectful of user privacy.

In sum, the trajectory of machine learning-based hand recognition reflects a broader movement toward intelligent and adaptive systems capable of interpreting natural human behavior. As the boundaries of perception and interaction technologies continue to expand, hand recognition stands as a key enabler of seamless, intuitive, and human-centered computing. Continued innovation in this domain promises not only technical breakthroughs but also transformative impacts across accessibility, education, communication, and beyond.

References

1. Samai, D.; Bensid, K.; Meraoumia, A.; Taleb-Ahmed, A.; Bedda, M. 2d and 3d palmprint recognition using deep learning method. In Proceedings of the 2018 3rd international conference on pattern analysis and intelligent systems (PAIS). IEEE, 2018, pp. 1–6.
2. Xin, Z.D.P.; Xin, P.; Xiaoling, L.; Xiaojing, G. Palmprint recognition based on deep learning. In Proceedings of the International Conference on Wireless, Mobile and Multi-Media (ICWMMN). IET, 2015.
3. Li, B.; Dai, Y.; Cheng, X.; Chen, H.; Lin, Y.; He, M. Skeleton based action recognition using translation-scale invariant image mapping and multi-scale deep CNN. In Proceedings of the 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). IEEE, 2017, pp. 601–604. Visited on 10/03/2022.
4. Zhong, D.; Du, X.; Zhong, K. Decade progress of palmprint recognition: A brief survey. *Neurocomputing* **2019**, *328*, 16–28.
5. Zhang, W.; Lin, Z.; Cheng, J.; Ma, C.; Deng, X.; Wang, H. STA-GCN: two-stream graph convolutional network with spatial–temporal attention for hand gesture recognition. *The Visual Computer* **2020**, *36*, 2433–2444. Visited on 15/03/2022.
6. Yang, Z.; Huangfu, H.; Leng, L.; Zhang, B.; Teoh, A.B.J.; Zhang, Y. Comprehensive competition mechanism in palmprint recognition. *IEEE Transactions on Information Forensics and Security* **2023**.
7. Hamilton, W.; Ying, Z.; Leskovec, J. Inductive representation learning on large graphs. *Advances in neural information processing systems* **2017**, *30*. Visited on 23/03/2022.
8. Fei, L.; Zhao, S.; Jia, W.; et al. Toward efficient palmprint feature extraction by learning a single-layer convolution network. *IEEE Transactions on Neural Networks and Learning Systems* **2022**, *34*, 9783–9794.
9. Garcia-Hernando, G.; Yuan, S.; Baek, S.; Kim, T.K. First-person hand action benchmark with rgb-d videos and 3d hand pose annotations. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 409–419. Visited on 21/03/2022.
10. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv preprint arXiv:1710.10903* **2017**. Visited on 17/03/2022.
11. Bauer, A.; Trapp, S.; Stenger, M.; Leppich, R.; Kounev, S.; Leznik, M.; Chard, K.; Foster, I. Comprehensive exploration of synthetic data generation: A survey. *arXiv preprint arXiv:2401.02524* **2024**.
12. Rai, B.S.; Pakkala, P.G.R.; Neha, G.; et al. Intelligent Framework for Early Prediction of Type–II Diabetes by Accelerating Palm Print Images using Graph Data Science. In Proceedings of the IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER). IEEE, 2023, pp. 207–212.
13. Shi, L.; Zhang, Y.; Cheng, J.; Lu, H. Non-local graph convolutional networks for skeleton-based action recognition. *arXiv preprint arXiv:1805.07694* **2018**, *1*, 3. Visited on 11/03/2022.
14. Caputo, F.M.; Prebianca, P.; Carcangiu, A.; Spano, L.D.; Giachetti, A. Comparing 3D trajectories for simple mid-air gesture recognition. *Computers & Graphics* **2018**, *73*, 17–25. Visited on 03/06/2022.
15. Jia, W.; Huang, D.S.; Zhang, D. Palmprint verification based on robust line orientation code. *Pattern Recognition* **2008**, *41*, 1504–1513.
16. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the IEEE conference on computer vision and pattern recognition. Ieee, 2005, Vol. 1, pp. 886–893.

 doi:10.20944/preprints202505.1465.v1

17. Li, G.; Kim, J. Palmprint recognition with local micro-structure tetra pattern. *Pattern Recognition* **2017**, *61*, 29–46.
18. Alfalahi, H.; Khandoker, A.H.; Chowdhury, N.; et al. Diagnostic accuracy of keystroke dynamics as digital biomarkers for fine motor decline in neuropsychiatric disorders: a systematic review and meta-analysis. *Scientific reports* **2022**, *12*, 7690.
19. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In Proceedings of the International Conference on Learning Representations, 2020.
20. Kong, A.K.; Zhang, D. Competitive coding scheme for palmprint verification. In Proceedings of the Proceedings of the International Conference on Pattern Recognition. IEEE, 2004, Vol. 1, pp. 520–523.
21. Gomez-Barrero, M.; Rathgeb, C.; Galbally, J.; Busch, C.; Fierrez, J. Unlinkable and irreversible biometric template protection based on bloom filters. *Information Sciences* **2016**, *370*, 18–32.
22. Song, Y.; Wang, T.; Cai, P.; Mondal, S.K.; Sahoo, J.P. A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. *ACM Computing Surveys* **2023**, *55*, 1–40.
23. Kong, A.K.; Zhang, D. Competitive coding scheme for palmprint verification. In Proceedings of the Proceedings of the International Conference on Pattern Recognition. IEEE, 2004, Vol. 1, pp. 520–523.
24. Zhang, D.; Guo, Z.; Lu, G.; Zhang, L.; Zuo, W. An online system of multispectral palmprint verification. *IEEE transactions on instrumentation and measurement* **2009**, *59*, 480–490.
25. Liang, X.; Li, Z.; Fan, D.; et al. Innovative contactless palmprint recognition system based on dual-camera alignment. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **2022**, *52*, 6464–6476.
26. Wu, X.; Zhao, Q.; Bu, W. A SIFT-based contactless palmprint verification approach using iterative RANSAC and local palmprint descriptors. *Pattern recognition* **2014**, *47*, 3314–3326.
27. Ungureanu, A.S.; Salahuddin, S.; Corcoran, P. Toward unconstrained palmprint recognition on consumer devices: A literature review. *IEEE Access* **2020**, *8*, 86130–86148.
28. Simonyan, K. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* **2014**.
29. Lu, W.; Tong, Z.; Chu, J. Dynamic hand gesture recognition with leap motion controller. *IEEE Signal Processing Letters* **2016**, *23*, 1188–1192. Visited on 17/07/2022.
30. Fei, L.; Xu, Y.; Zhang, D. Half-orientation extraction of palmprint features. *Pattern Recognition Letters* **2016**, *69*, 35–41.
31. Rivera, G.; Florencia, R.; García, V.; et al. News classification for identifying traffic incident points in a Spanish-speaking country: A real-world case study of class imbalance learning. *Applied Sciences* **2020**, *10*, 6253.
32. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings, 2010, pp. 249–256. Visited on 02/05/2022.
33. De Smedt, Q.; Wannous, H.; Vandeborre, J.P. Heterogeneous hand gesture recognition using 3D dynamic skeletal data. *Computer Vision and Image Understanding* **2019**, *181*, 60–72. Visited on 07/03/2022.
34. Nguyen, K.; Proença, H.; Alonso-Fernandez, F. Deep learning for iris recognition: A survey. *ACM Computing Surveys* **2024**, *56*, 1–35.
35. Pham, V.T.; Tran, T.H.; Vu, H. Detection and tracking hand from FPV: benchmarks and challenges on rehabilitation exercises dataset. In Proceedings of the 2021 RIVF International Conference on Computing and Communication Technologies (RIVF). IEEE, 2021, pp. 1–6.
36. Fei, L.; Zhang, B.; Xu, Y.; Guo, Z.; Wen, J.; Jia, W. Learning discriminant direction binary palmprint descriptor. *IEEE Transactions on Image Processing* **2019**, *28*, 3808–3820.
37. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778. Visited on 17/03/2022.
38. Zou, Y.; Liu, C.; Shao, H.; Zhong, D. Unsupervised palmprint image quality assessment via pseudo-label generation and ranking guidance. *IEEE Transactions on Instrumentation and Measurement* **2023**, *72*, 1–11.
39. Godbole, A.; Grosz, S.A.; Jain, A.K. Contactless Palmprint Recognition for Children. In Proceedings of the 2023 International Conference of the Biometrics Special Interest Group (BIOSIG). IEEE, 2023, pp. 1–7.
40. Fei, L.; Zhang, B.; Jia, W.; Wen, J.; Zhang, D. Feature extraction for 3-D palmprint recognition: A survey. *IEEE Transactions on Instrumentation and Measurement* **2020**, *69*, 645–656.

 doi:10.20944/preprints202505.1465.v1

41. Mu, M.; Ruan, Q. Mean and standard deviation as features for palmprint recognition based on Gabor filters. *International Journal of Pattern Recognition and Artificial Intelligence* **2011**, *25*, 491–512.
42. Chan, T.H.; Jia, K.; Gao, S.; et al. PCANet: A simple deep learning baseline for image classification? *IEEE transactions on image processing* **2015**, *24*, 5017–5032.
43. Boutros, F.; Struc, V.; Fierrez, J.; Damer, N. Synthetic data for face recognition: Current state and future prospects. *Image and Vision Computing* **2023**, *135*, 104688.
44. Yulin, F.; Kumar, A. Best: Building evidences from scattered templates for accurate contactless palmprint recognition. *Pattern Recognition* **2023**, *138*, 109422.
45. De Smedt, Q.; Wannous, H.; Vandeborre, J.P.; Guerry, J.; Le Saux, B.; Filliat, D. Shrec'17 track: 3d hand gesture recognition using a depth and skeletal dataset. In Proceedings of the 3DOR-10th Eurographics Workshop on 3D Object Retrieval, 2017, pp. 1–6. Visited on 21/03/2022.
46. Wang, J.G.; Yau, W.Y.; Suwandy, A.; et al. Person recognition by fusing palmprint and palm vein images based on "Laplacianpalm" representation. *Pattern Recognition* **2008**, *41*, 1514–1527.
47. Chopra, S.; Hadsell, R.; LeCun, Y. Learning a similarity metric discriminatively, with application to face verification. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2005, Vol. 1, pp. 539–546.
48. Gomez-Barrero, M.; Galbally, J.; Rathgeb, C.; Busch, C. General framework to evaluate unlinkability in biometric template protection systems. *IEEE Transactions on Information Forensics and Security* **2017**, *13*, 1406–1420.
49. Kumar, A.; Shekhar, S. Personal identification using multibiometrics rank-level fusion. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **2010**, *41*, 743–752.
50. Ross, T.Y.; Dollár, G. Focal loss for dense object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2980–2988.
51. Aishwarya, D.; Gowri, M.; Saranya, R. Palm print recognition using liveness detection technique. In Proceedings of the 2016 Second International Conference on Science Technology Engineering and Management (ICONSTEM). IEEE, 2016, pp. 109–114.
52. Liu, H.; Sun, D.; Xiong, K.; Qiu, Z.; et al. A hybrid approach to protect palmprint templates. *The Scientific World Journal* **2014**, *2014*.
53. Sardar, A.; Umer, S.; Rout, R.K.; Khan, M.K. A secure and efficient biometric template protection scheme for palmprint recognition system. *IEEE Transactions on Artificial Intelligence* **2022**.
54. Emporio, M.; Caputo, A.; Giachetti, A. STRONGER: Simple TRajectory-based ONline GEsture Recognizer **2021**. Visited on 15/08/2022.
55. Wu, W.; Zhang, Y.; Li, Y.; Li, C. Fusion recognition of palmprint and palm vein based on modal correlation. *Mathematical Biosciences and Engineering* **2024**, *21*, 3129–3145.
56. Cho, S.; Oh, B.S.; Toh, K.A.; et al. Extraction and cross-matching of palm-vein and palmprint from the RGB and the NIR spectrums for identity verification. *IEEE Access* **2019**, *8*, 4005–4021.
57. Wirth, R.; Hipp, J. CRISP-DM: Towards a standard process model for data mining. In Proceedings of the Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining. Manchester, 2000, Vol. 1, pp. 29–39.
58. Zhang, D.; Guo, Z.; Lu, G.; et al. Online joint palmprint and palmvein verification. *Expert Systems with Applications* **2011**, *38*, 2621–2631.
59. Khan, M.S.; Li, H.; Zhao, C. Deep secure PalmNet: A novel cancelable palmprint template protection scheme with deep attention net and randomized hashing security mechanism. *Computers & Security* **2024**, *142*, 103863.
60. Indian Institute of Technology Delhi. IITD Touchless Palmprint Database (Version1.0). https://www4.comp.polyu.edu.hk/$\sim$csajaykr/IITD/Data\protect\discretionary{\char\hyphenchar\font}{}{}base_Palm.htm, 2006.
61. Sepas-Moghaddam, A.; Etemad, A. Deep gait recognition: A survey. *IEEE transactions on pattern analysis and machine intelligence* **2022**, *45*, 264–284.
62. Zhu, Q.; Zhou, Y.; Fei, L.; et al. Multi-spectral palmprints joint attack and defense with adversarial examples learning. *IEEE Transactions on Information Forensics and Security* **2023**, *18*, 1789–1799.
63. Xing, Y.; Zhu, J. Deep learning-based action recognition with 3D skeleton: a survey, 2021. Visited on 07/03/2022.
64. Zhong, D.; Zhu, J. Centralized large margin cosine loss for open-set deep palmprint recognition. *IEEE Transactions on Circuits and Systems for Video Technology* **2019**, *30*, 1559–1568.

 doi:10.20944/preprints202505.1465.v1

65. Hedegaard, L.; Iosifidis, A. Continual inference: a library for efficient online inference with deep neural networks in pytorch. *arXiv preprint arXiv:2204.03418* **2022**.
66. Chen, X.; Guo, H.; Wang, G.; Zhang, L. Motion feature augmented recurrent neural network for skeleton-based dynamic hand gesture recognition. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017, pp. 2881–2885. Visited on 08/03/2022.
67. Ruan, S.; Li, Y.; Qin, H. LSFM: Light Style and Feature Matching for Efficient Cross-Domain Palmprint Recognition. *IEEE Transactions on Information Forensics and Security* **2024**.
68. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. arXiv 2017. *arXiv preprint arXiv:1710.10196* **2018**, pp. 1–26.
69. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Proceedings of the AAAI conference on artificial intelligence, 2017, Vol. 31.
70. Akremi, M.S.; Slama, R.; Tabia, H. SPD Siamese Neural Network for Skeleton-based Hand Gesture Recognition. In Proceedings of the VISIGRAPP (4: VISAPP), 2022, pp. 394–402. Visited on 04/06/2022.
71. Ohn-Bar, E.; Trivedi, M. Joint angles similarities and HOG2 for action recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2013, pp. 465–470. Visited on 07/03/2022.
72. Wu, X.; Zhao, Q.; Bu, W. A SIFT-based contactless palmprint verification approach using iterative RANSAC and local palmprint descriptors. *Pattern Recognition* **2014**, *47*, 3314–3326.
73. Liu, B.; Feng, J. Palmprint orientation field recovery via attention-based generative adversarial network. *Neurocomputing* **2021**, *438*, 1–13.
74. Hawkins, D.M. The problem of overfitting. *Journal of chemical information and computer sciences* **2004**, *44*, 1–12.
75. Niepert, M.; Ahmed, M.; Kutzkov, K. Learning convolutional neural networks for graphs. In Proceedings of the International conference on machine learning. PMLR, 2016, pp. 2014–2023. Visited on 22/03/2022.
76. Yang, Z.; Xia, W.; Lu, Z.; et al. Hypernetwork-based physics-driven personalized federated learning for CT imaging. *IEEE Transactions on Neural Networks and Learning Systems* **2023**.
77. Matkowski, W.M.; Chai, T.; Kong, A.W.K. Palmprint recognition in uncontrolled and uncooperative environment. *IEEE Transactions on Information Forensics and Security* **2019**, *15*, 1601–1615.
78. Li, Z.; Liang, X.; Fan, D.; et al. BPFNet: A unified framework for bimodal palmprint alignment and fusion. In Proceedings of the Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part VI 28. Springer, 2021, pp. 28–36.
79. Chen, Y.; Zhao, L.; Peng, X.; Yuan, J.; Metaxas, D.N. Construct dynamic graphs for hand gesture recognition via spatial-temporal attention. *arXiv preprint arXiv:1907.08871* **2019**. Visited on 16/03/2022.
80. De Smedt, Q.; Wannous, H.; Vandeborre, J.P. Skeleton-based dynamic hand gesture recognition. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016, pp. 1–9. Visited on 08/03/2022.
81. Siddhad, G.; Khanna, P.; Ojha, A. Cancelable biometric template generation using convolutional autoencoder. In Proceedings of the International Conference on Computer Vision and Image Processing. Springer, 2020, pp. 303–314.
82. Wang, F.; Leng, L.; Teoh, A.B.J.; Chu, J. Palmprint false acceptance attack with a generative adversarial network (GAN). *Applied Sciences* **2020**, *10*, 8547.
83. Lin, C.; Chen, Y.; Zou, X.; Deng, X.; Dai, F.; You, J.; Xiao, J. An unconstrained palmprint region of interest extraction method based on lightweight networks. *Plos one* **2024**, *19*, e0307822.
84. De Smedt, Q.; Wannous, H.; Vandeborre, J.P. Skeleton-based dynamic hand gesture recognition. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016, pp. 1–9. Visited on 07/03/2022.
85. Gu, J.; Wang, Z.; Kuen, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognition* **2018**, *77*, 354–377.
86. Yang, Z.; Kang, M.; Teoh, A.B.J.; Gao, C.; Chen, W.; Zhang, B.; Zhang, Y. A Dual-Level Cancelable Framework for Palmprint Verification and Hack-Proof Data Storage. *IEEE Transactions on Information Forensics and Security* **2024**, *19*, 8587–8599.
87. Yang, W.; Wang, S.; Hu, J.; Tao, X.; Li, Y. Feature extraction and learning approaches for cancellable biometrics: A survey. *CAAI Transactions on Intelligence Technology* **2024**, *9*, 4–25.

 doi:10.20944/preprints202505.1465.v1

88. Jalali, A.; Mallipeddi, R.; Lee, M. Deformation invariant and contactless palmprint recognition using convolutional neural network. In Proceedings of the International Conference on Human-agent Interaction, 2015, pp. 209–212.
89. Wang, Y.; Fei, L.; Zhao, S.; Zhu, Q.; Wen, J.; Jia, W.; Rida, I. Dense Hybrid Attention Network for Palmprint Image Super-Resolution. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **2024**.
90. Xie, Z.; Guo, Z.; Qian, C. Palmprint gender classification by convolutional neural network. *IET Computer Vision* **2018**, *12*, 476–483.
91. Kipf, T.; Fetaya, E.; Wang, K.C.; Welling, M.; Zemel, R. Neural relational inference for interacting systems. In Proceedings of the International Conference on Machine Learning. PMLR, 2018, pp. 2688–2697. Visited on 18/03/2022.
92. Shi, L.; Zhang, Y.; Cheng, J.; Lu, H. Skeleton-based action recognition with multi-stream adaptive graph convolutional networks. *IEEE Transactions on Image Processing* **2020**, *29*, 9532–9545. Visited on 04/06/2022.
93. Zhu, Q.; Xu, N.; Zhang, Z.; et al. Cross-spectral palmprint recognition with low-rank canonical correlation analysis. *Multimedia Tools and Applications* **2020**, *79*, 33771–33792.
94. Svoboda, J.; Masci, J.; Bronstein, M.M. Palmprint recognition via discriminative index learning. In Proceedings of the International Conference on Pattern Recognition (ICPR). IEEE, 2016, pp. 4232–4237.
95. Caffagni, D.; Cocchi, F.; Barsellotti, L.; Moratelli, N.; Sarto, S.; Baraldi, L.; Cornia, M.; Cucchiara, R. The (r) evolution of multimodal large language models: A survey. *arXiv preprint arXiv:2402.12451* **2024**.
96. Chatfield, K. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531* **2014**.
97. Huang, Z.; Van Gool, L. A riemannian network for spd matrix learning. In Proceedings of the Thirty-first AAAI conference on artificial intelligence, 2017. Visited on 03/06/2022.
98. Kumar, A.; Wong, D.C.; Shen, H.C.; et al. Personal verification using palmprint and hand geometry biometric. In Proceedings of the Audio-and Video-Based Biometric Person Authentication: 4th International Conference, AVBPA 2003 Guildford, UK, June 9–11, 2003 Proceedings 4. Springer, 2003, pp. 668–678.
99. Li, S.; Fei, L.; Zhang, B.; Ning, X.; Wu, L. Hand-based multimodal biometric fusion: A review. *Information Fusion* **2024**, p. 102418.