Linkage Attack on Skeleton-based Motion Visualization

Thomas Carr
University of North Carolina at Charlotte
tcarr23@uncc.edu

Aidong Lu
University of North Carolina at Charlotte
Aidong.Lu@uncc.edu

Depeng Xu
University of North Carolina at Charlotte
dxu7@uncc.edu

ABSTRACT

Skeleton-based motion capture and visualization is an important computer vision task, especially in the virtual reality (VR) environment. It has grown increasingly popular due to the ease of gathering skeleton data and the high demand of virtual socialization. The captured skeleton data seems anonymous but can still be used to extract personal identifiable information (PII). This can lead to an unintended privacy leakage inside a VR meta-verse. We propose a novel linkage attack on skeleton-based motion visualization. It detects if a target and a reference skeleton are the same individual. The proposed model, called Linkage Attack Neural Network (LAN), is based on the principles of a Siamese Network. It incorporates deep neural networks to embed the relevant PII then uses a classifier to match the reference and target skeletons. We also employ classical and deep motion retargeting (MR) to cast the target skeleton onto a dummy skeleton such that the motion sequence is anonymized for privacy protection. Our evaluation shows that the effectiveness of LAN in the linkage attack and the effectiveness of MR in anonymization. The source code is available at https://github.com/Thomasc33/Linkage-Attack

CCS CONCEPTS

• **Security and privacy**; • **Computing methodologies** → **Machine learning**; **Virtual reality**;

KEYWORDS

privacy attack, virtual reality, motion visualization

ACM Reference Format:
Thomas Carr, Aidong Lu, and Depeng Xu. 2023. Linkage Attack on Skeleton-based Motion Visualization. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23), October 21–25, 2023, Birmingham, United Kingdom.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3583780.3615263

1 INTRODUCTION

Visualizing human motion in virtual reality (VR) combines motion capture and virtual reality to create a realistic simulation of a person's movements. This allows for detailed analysis of movements, identification of areas for improvement, and development of training programs. It has numerous applications in fields such as sports, physical therapy, and entertainment [10, 18, 25]. Human motion visualization may reveal sensitive data such as body measurements, movement patterns, and potentially even biometric data [17]. It is important to protect the personal data and identity of individuals who are interacting in virtual reality meta-verse.

Motion capture technologies like Kinect and Perception Neuron output accurate skeleton data for visualization in the VR environment. Such skeleton-based motion visualization contains limited personal information and seem "anonymized". However, several studies prove that personal identification from skeleton data has good performance because of the recent development in deep neural networks. [21] combines gait recognition and neural networks to identify individuals using the skeleton information from Kinect. [26] uses a multi-task Recurrent Neural Networks (RNN) to simultaneously predict person ID and action class. [16] uses Graph Convolutional Networks (GCN) based models to identify individual's gender and identity. The limitations of these identification is that they use ID for supervised learning, where the attacking model tries to re-identify individuals by predicting ID number. It requires the adversarial attacker has access to a large amount of skeleton data from the same person. It also does not work well for individuals who are not included in the adversary's training.

A linkage attack, structured like a Siamese Neural Network [5], allows adversaries to identify sensitive or private information about an individual by linking together anonymized information and publicly available information. Siamese Networks has been used for facial recognition [22, 28], signature verification [4, 7, 29], and object tracking [3, 8, 11, 12] among other tasks. In terms of motion visualization, the target skeleton data is visualized in VR and accessible to the adversary. The adversary can easily extract reference skeleton data from public videos. Then through linkage attacks, the adversary can identify the individuals by matching the target skeleton and the reference skeleton. Unlike the supervised personal ID classification, the linkage attack model applies to any individuals, even the ones it has not seen before.

In this work, we propose a novel framework of Linkage Attack Neural Networks, called LAN, to attack on skeleton-based motion visualization. LAN is inspired by recent development of using deep learning for action recognition on the spatio-temporal skeleton sequences with various network structures, including Recurrent Neural Networks (RNN) [9, 19, 20, 23, 31, 34], Convolutional Neural Networks (CNN) [13, 15, 27, 32], and Graph Convolutional Networks (GCN) [20, 24, 30]. Our proposed LAN model includes two Semantic-Guided Encoders (SGE) and and a matching classifier. The SGE obtain embeddings from the target and reference skeletons separately. SGE consists of a GCN-based joint-level module and a CNN-based frame-level module. The final embedding from SGE encodes personal identifiable information (PII) from both the joint-level and the frame-level. The matching classifier takes the embeddings of the target and reference skeleton motions and predicts whether or not they are from the same individual.

CIKM '23, October 21–25, 2023, Birmingham, United Kingdom

ACM ISBN 979-8-4007-0124-5/23/10…$15.00
https://doi.org/10.1145/3583780.3615263

In addition to the linkage attack model, we propose to use motion retargeting (MR) for privacy protection. Motion retargeting methods transfer the motion from one character to another while maintaining the overall timing and movement patterns. We employ classical and deep motion retargeting methods to project the private target skeleton onto a normalized dummy character to mask the personal identity. Classical motion retargeting, utilizing inverse and forward kinematics, aligns the joints of the dummy character with those of the target character and subsequently maps the motion by casting joint rotations [6]. Deep motion retargeting trains a deep neural network to decompose temporal motion sequences into skeleton-agnostic dynamic motion and static skeleton [2].

To evaluate the effectiveness of our proposed linkage attack model, we experiment on a large widely-used skeleton dataset NTU120 [14]. The experiment results show that LAN is effective in detecting PII leakage in the skeleton data. It generalizes well to unseen characters and action classes. Even if the existing anonymization methods, including motion retargeting, defends the linkage attack, they suffer a big loss in utility.

We summarize our contributions as follows: (1) To the best of our knowledge, this is the first work to perform linkage attacks on the skeleton data. (2) We developed a deep linkage attack model, LAN, which uses semantic guided encoders to encode PII from skeleton sequences and then conducts comparison. (3) We also show that motion retargeting works as a general privacy protection method.

2 METHODOLOGY

2.1 Problem Statement

A 3D skeleton data $\mathbf{s} \in \mathbb{R}^{N\times M\times 3}$ captures the human motion with 3D coordinates $\mathbf{s} = (x_n^m, y_n^m, z_n^m)^{M\times N}$ of M joins over N frames. The skeleton data is visualized in VR so the motion action information can be recognized. The skeleton visualization is anonymous in VR or at least reveals limited PII. The adversary can use public skeleton data to train a linkage attack model. The goal of the linkage attack model is to determine whether two skeleton sequences represent the same person when it is given an anonymous target skeleton $\mathbf{s}_T$ and a reference skeleton $\mathbf{s}_R$ with known identity.

2.2 Linkage Attack Neural Networks

We propose a Linkage Attack Neural Networks, called LAN, for linkage attacks on skeleton-based motion visualizations. Figure 1 shows the overall end-to-end framework. It consists of two Semantic-Guided Encoders E_T, E_R and a matching classifier C. Specifically, SGE E_T and E_R takes in the target skeleton $\mathbf{s}_T$ and the reference skeleton $\mathbf{s}_R$, respectively, to extract low-dimensional embeddings $\mathbf{e}_T, \mathbf{e}_R$, which encodes personal identifiable information from both the static skeleton joint structure and the dynamic dependency. The classifier C takes the embeddings $\mathbf{e}_T, \mathbf{e}_R$ and makes the prediction y on whether the target and reference belong to the same person.

Semantic-Guided Encoder (SGE). The Semantic-Guided Encoder is inspired by a state-of-the-art action recognition model known as the Semantic Guided Neural Network (SGN) [33]. SGE explicitly introduce the high level semantics, joint type and frame index, to improve the representation capability of learned features. SGE first represents the skeleton sequence $\mathbf{s}$ with a dynamics representation (DR). Then the joint-level module (JM) exploit PII from the correlations of joints in the same frame, while the frame-level module (FM) exploits PII from the correlations across frames.

Dynamics Representation (DR). For a given joint $\mathbf{s}_n^m$, we define its dynamics by the position $\mathbf{p}_n^m = (x_n^m, y_n^m, z_n^m)^N \in \mathbb{R}^3$ in the 3D coordinate system, and the velocity $\mathbf{v}_n^m = \mathbf{p}_n^m - \mathbf{p}_{n-1}^m$. Both $\mathbf{p}_n^m$ and $\mathbf{v}_n^m$ go through two fully connected (FC) layers with ReLU activation functions and end up as high-dimensional representations $\tilde{\mathbf{p}}_n^m, \tilde{\mathbf{v}}_n^m$. The final dynamics representation fuses them together by summation as $\mathbf{r}_n^m = \tilde{\mathbf{p}}_n^m + \tilde{\mathbf{v}}_n^m \in \mathbb{R}^{d_1}$, where d_1 is the dimension of the joint representation.

Joint-level Module (JM). The Joint-level Module (JM) adopts GCNs to explore the correlations for the structural skeleton data. After DR, we get $\mathbf{s} = (\mathbf{r}_n^m)^{M\times N}$. The joint type m is converted to a representation with a dimension of d_1, and then concatenated with $\mathbf{r}_n^m$. The semantic representations for joint types are shared for both E_T and E_R. Thus, the joint representation of joint type m at frame n with both the dynamics and semantics of joint type becomes $\tilde{\mathbf{r}}_n^m \in \mathbb{R}^{2d_1}$. All the joints at frame n is represented by $R_n \in \mathbb{R}^{M\times 2d_1}$. The edge weight from the joint i to joint j in the same frame n is modeled by their affinity in the embedded space as $a_n(i,j) = \theta(\tilde{\mathbf{r}}_n^i)^T \phi(\tilde{\mathbf{r}}_n^j)$, where θ and ϕ denote two transformation functions, each implemented by an FC layer. The adjacency matrix A_n is obtained by computing the affinities of all the joint pairs at frame n and then normalization with Softmax. After the residual graph convolution layer, the final output of JM at frame n is $R'_n = A_n R_n W_1 + R_n W_2 \in \mathbb{R}^{d_2}$, where W_1 and W_2 are transformation matrices. The weights are shared for different temporal frames.

Frame-level Module (FM). The Frame-level Module (FM) adopts CNNs to explore the correlations across frames. After JM, we get $\mathbf{s} = (r'^m_n)^{M\times N}$. The frame index n is converted to a representation with a dimension of d_2, and then fused by summation with $\mathbf{r}_n^m$. The semantic representations for frame indices are shared for both E_T and E_R. Thus, the joint representation of joint type m at frame n with both the learned feature and semantics of frame index become $\tilde{\mathbf{r}'}^m_n \in \mathbb{R}^{d_2}$. A Spatial MaxPooling (SMP) layer is applied to aggregate the information across the joints to a dimension of $N \times 1 \times d_2$. A temporal CNN layer is applied to model the dependencies of frames. Then another CNN layer maps it to a high dimensional space of d_3 with a kernel size of 1. In the end, a Temporal MaxPooling (TMP) layer is applied to aggregate the information of all frames. The final output of FM at sequence level is $\mathbf{e} \in \mathbb{R}^{d_3}$. It encodes PII from the correlations of joints and the dependencies of frames.

Matching Classification. The matching classifier is a neural network to compare the extracted PII in the embeddings $\mathbf{e}_T$ and $\mathbf{e}_R$. If the PII belongs to the same person, it predicts matching $y = 0$; otherwise, $y = 1$. The classifier C consists of a 1D convolution layer, two batch normalization layers, and three fully connected layers. The last FC layer uses Sigmoid activation function for classification.

Model Training. To train the LAN model, we construct paired training data $(\mathbf{s}_T, \mathbf{s}_R, y)^{|S|}$ through positive and negative sampling. Positive sampling selects a target skeleton sequence $\mathbf{s}_T$ of an individual, then randomly selects another skeleton sequence from the same person as the reference skeleton $\mathbf{s}_R$. The pair is assigned a matching label $y = 1$. Negative sampling selects a target skeleton sequence $\mathbf{s}_T$ of an individual, then randomly selects a skeleton sequence from any other person as the reference skeleton $\mathbf{s}_R$. The pair

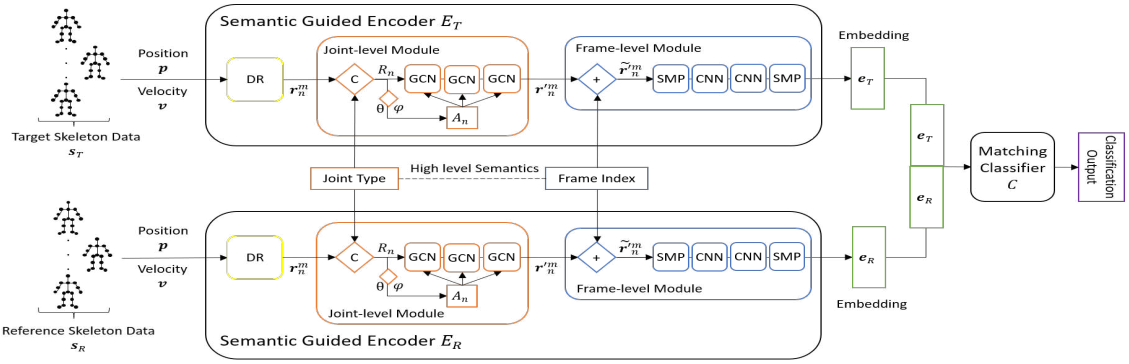

Figure 1: The framework of LAN. It consists of two Semantic-Guided Encoders and a Matching Classifier.

is assigned a matching label $y = 0$. We maintain balanced sampling rates to maximize the classifier performance of the classifier.

The Semantic Guided Encoders (SGEs) are pre-trained for identity classification. The SGE alone is followed by a fully connected layer with Softmax to predict the personal ID. The learned embeddings from the pre-trained SGEs can already capture the PII correlations of joints within and across frames. During the LAN training, the two SGEs E_T, E_R are guided to extract PII related correlations for the purpose of linkage attack. The output embeddings $\mathbf{e}_T$ and $\mathbf{e}_R$ are then concatenated and passed to the classifier. The loss function of LAN is a binary cross entropy loss $CE\,(y, f(\mathbf{s}_T, \mathbf{s}_R))$, where $f(\mathbf{s}_T, \mathbf{s}_R) = C(E_T(\mathbf{s}_T), E_R(\mathbf{s}_R))$.

2.3 Anonymization through Motion Retargeting

To defend against a linkage attack, visualizing the raw skeleton data without personal ID is not enough. To truly anonymize the skeleton, the indirect information in the skeleton motion sequence related to PII should be removed. Previous study [16] creates an adversarial training-based anonymization framework for skeleton action recognition. It modifies the skeleton data to confuse a personal ID classifier and a gender classifier while maintaining the performance of an action recognition model. The limitation of such adversarial training-based defense includes: (1) personal ID classification only works on identities seen by the model, but linkage attack works on skeletons of unseen identities and unseen action class. (2) Adversarial training is confined to the seen actions, individuals, and attackers. (3) The anonymizer only preserves the performance of the included action recognition model. Instead of adversarial training, we propose to use motion retargeting for anonymization. Motion retargeting is not restricted to specific characters, actions, or models that have been previously encountered. It can be easily generalized to any skeleton data, making it a versatile and complementary defense against linkage attacks.

To use motion retargeting for skeleton anonymization, we cast all the raw skeleton data to a "dummy" character. Then we only use the transformed new skeleton for motion visualization in VR. The spatial structure of the skeleton is transformed, which effectively mitigates the indirect PII related to unique spatial attributes. At the same time, the essence of the motion pattern remains largely intact, ensuring that the anonymized data is still valuable for downstream applications. We employ both classical and deep motion retargeting in this work. Classical MR approach is grounded in the principle of Inverse and Forward Kinematics [6, Chapter 4-5], which allows for the calculation of each joint's XYZ position, given a new joint length and XYZ orientation. By preserving the joint rotations it retains a majority of the temporal data in the skeleton's movement while casting the motion sequence from one character to a new character. The benefits of classic approach includes the lack of formal training required and a relatively lower computational cost during evaluation. Recently, deep learning based motion retargeting is developed [1, 2]. It trains a deep neural network to extract a high-level latent motion representation, which is invariant to the skeleton geometry. It decompose temporal motion sequences into explicit latent representations of dynamic motion and static skeleton. Then it re-combines the motion with novel skeletons, and decodes a retargeted temporal sequence. Due to limited space, please check the references for implementation details of the MR approaches.

3 EXPERIMENTS

3.1 Experiment Setup

Dataset We use the NTU RGB+D 60+120 dataset [14], which is a large-scale dataset of human motions captured with the Microsoft Kinect v2 sensor. The dataset was created in two parts, NTU60 (40 actors and 60 actions) and NTU120 (66 new actors and 60 new actions). The skeleton contains position and rotation information for 25 joints. Only the position information is used for our experiments.

Linkage Attack. *Implementation details.* To train the Linkage Attack Neural Networks (**LAN**), we utilize the entire NTU60 dataset, which has 40 actors. For testing, we employ the 66 unseen actors from the NTU120 dataset. We use a default sampling size of 400 per target actor for both the positive sampling (featuring the same actor) and negative sampling (featuring different actors), yielding 32,000 training samples and 52,800 testing samples.

Baselines. In Table 1, we compare the performance of the proposed LAN model to two baselines. (1) A frame-wise random forest (**RF**) is trained on 1.2 million samples and tested on 2.07 million samples. Through hyperparameter tuning, we select $n\_estimator = 100$ and $max\_depth = 10$. (2) A multi-layer perceptron model (**MLP**) is trained on 400,000 samples and tested it on 690,000 samples. The MLP model had 4 layers with sizes of [1000, 100, 100, 1]. The input is a flattened sequence of 50 frames.

Anonymization. *Implementation details.* Classical motion retargeting (**CMR**) does not require training. For the dummy skeleton,

Table 1: Linkage attack performance comparison

Attack Model	Precision	Recall	F-1 score
LAN	0.6830	0.8138	0.7427
MLP	0.7059	0.6852	0.6954
RF	0.7346	0.7708	0.6576

we use an average skeleton based on all the actors in the NTU60+120 dataset, which averages on the Euclidean distances along the joint paths. The motion from the original skeleton identity is retargeted to the average dummy skeleton while preserving the overall timing and movement. Deep motion retargeting (**DMR**) is based on [2]. By encoding static and dynamic data, swapping the static data, then decoding we achieve a retargeted skeleton. We use a random actor as the dummy skeleton to cast all of our sequence data to.

Baselines. We compare the MR anonymizer with the **UNet** and **ResNet** anonymizers from [16]. Both are trained on the NTU60 dataset. Consequently, the additional NTU120 data remains unseen to the anonymizer. This allows us to evaluate the performance of the anonymizer on unseen data in terms of actors and actions. This evaluation also aligns with the split of the linkage attack models.

Utility evaluation. Additionally, we compare the proposed MR algorithm to the UNet and ResNet models by evaluating its utility with action recognition. We use the SGN model [33] for action recognition due to its state-of-the-art performance. An effective anonymization method should strike a balance between privacy protection and maintaining the utility of the data. This balance is crucial for real-world applications, as overly aggressive anonymization may render the data unusable for its intended purpose.

3.2 Experimental Results

Linkage Attack. *Comparison against baselines.* As shown in Table 1, the proposed LAN detects a significant leakage of private information, with an F-1 score of 0.7427. The proposed LAN produce an F-1 score that is 4.73% higher than MLP and 8.51% higher than RF. This comparison demonstrates that the semantic guided embeddings by SGE effectively capture the PII encoded in the joints correlations within and across frames. We also evaluated how the end-to-end training of LAN improved the attack performance over one with the SGE layers frozen after pre-training. The end-to-end training yields an average of 2% higher F-1 score on testing.

Scalability analysis. We conduct the scalability analysis of LAN on the availability of the training sample to the adversary. We vary the sampling size per actor when constructing the training dataset via positive and negative sampling. As seen in Figure 2, having more training data available to the adversary yields a higher attacking F-1 score. When the training data are limited, the LAN model still has a relatively high F-1 score. For example, when the sampling size reduce to 100 (25% of the default setting), LAN still achieves 0.7136 F-1 score. But it only takes 20% run time than the default.

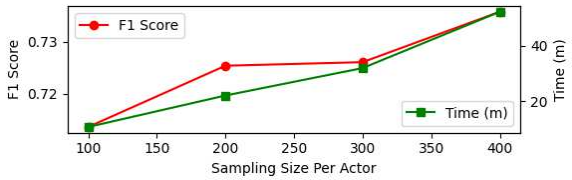

Figure 2: Scalability analysis on sampling size per actor

Table 2: Linkage attack on the anonymized data

Data for Visualization	Precision	Recall	F-1 score
Raw data	0.6830	0.8138	0.7427
UNet	0.5	1.0	0.6667
ResNet	0.5	1.0	0.6667
CMR	0.5004	0.9963	0.6662
DMR	0.5057	0.8977	0.6469

Anonymization. *Defense against linkage attacks.* We use anonymization on skeleton data to defend against the linkage attacks. We compare the performance of four different anonymizer models: UNet and ResNet from [16], CMR and DMR. To defend a linkage attack, a perfect anonymizer should trick the attack model into predicting all skeletons are the same, i.e., a Recall score of 1.0.

Table 2 shows the results of the linkage attack as well as the performance of the anonymizer models. The base linkage attack achieves an impressively high F-1 score at 74.27%. All anonymizer models tested fool the LAN model into believing most actors were the same, i.e., the anonymizers are great at hiding PII. This means that the LAN model focuses more heavily on spatial rather than temporal information. This aligns with human perception, as people tend to recognize others based on appearance rather than movement.

Action recognition utility. Upon testing the utility, we find the raw data preserves all the action class information well. The SGN action recognition model on the raw data achieves an accuracy of 94.25% on the NTU60+120 dataset. However, the anonymized data preserves little utility about action recognition. The highest utility for the anonymizer is DMR with an action recognition accuracy of 4.55%, followed by CMR with an action recognition accuracy of 3.2%. The UNet and ResNet anonymizers both achieve an accuracy of only 0.84%, which is about the same as random choices (1/120). The high action recognition utility presented in [16] is only because the anonymizer's utility is only preserved when evaluated with the pre-trained utility adversary. The DMR had the highest utility but the lowest privacy performance indicating the necessity of a privacy/utility trade off. It suggests that further research and development are necessary to improve the utility performance of anonymizers without compromising privacy protection.

4 CONCLUSION

In this work we presented a novel linkage attack method, called Linkage Attack Neural Network (LAN), that detects if a target and a reference skeleton are the same individual. We base the model on the structure of Siamese Networks and utilize the semantic guided encoders to create a low dimensional PII encoding. Our experiment reveals that there is a privacy leakage that the LAN can detect. We also present two MR based defense models and compare their results to established anonymizer frameworks. In future works, we will develop a deep motion retargeting framework purpose built to mitigate PII leakage and anonymize the skeleton while preserving its action recognition utility.

ACKNOWLEDGEMENTS

This work was supported in part by UNC Charlotte startup fund and NSF grant 1840080.

REFERENCES

[1] Kfir Aberman, Peizhuo Li, Dani Lischinski, Olga Sorkine-Hornung, Daniel Cohen-Or, and Baoquan Chen. 2020. Skeleton-aware networks for deep motion retargeting. *ACM Trans. Graph.* 39, 4 (2020), 62.

[2] Kfir Aberman, Rundi Wu, Dani Lischinski, Baoquan Chen, and Daniel Cohen-Or. 2019. Learning character-agnostic motion for motion retargeting in 2D. *ACM Trans. Graph.* 38, 4 (2019), 75:1–75:14.

[3] Luca Bertinetto, Jack Valmadre, João F. Henriques, Andrea Vedaldi, and Philip H. S. Torr. 2016. Fully-Convolutional Siamese Networks for Object Tracking. In *Computer Vision - ECCV 2016 Workshops - Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 9914)*, Gang Hua and Hervé Jégou (Eds.). 850–865.

[4] Jane Bromley, James W. Bentz, Léon Bottou, Isabelle Guyon, Yann LeCun, Cliff Moore, Eduard Säckinger, and Roopak Shah. 1993. Signature Verification Using A "Siamese" Time Delay Neural Network. *Int. J. Pattern Recognit. Artif. Intell.* 7, 4 (1993), 669–688.

[5] Davide Chicco. 2021. Siamese Neural Networks: An Overview. In *Artificial Neural Networks - Third Edition*, Hugh M. Cartwright (Ed.). Methods in Molecular Biology, Vol. 2190. Springer, 73–94.

[6] Carl D. Crane, III and Joseph Duffy. 1998. *Kinematic Analysis of Robot Manipulators.* Cambridge University Press. https://doi.org/10.1017/CBO9780511530159

[7] Sounak Dey, Anjan Dutta, Juan Ignacio Toledo, Suman K. Ghosh, Josep Lladós, and Umapada Pal. 2017. SigNet: Convolutional Siamese Network for Writer Independent Offline Signature Verification. *CoRR* abs/1707.02131 (2017).

[8] Xingping Dong and Jianbing Shen. 2018. Triplet Loss in Siamese Network for Object Tracking. In *Proceedings of the European Conference on Computer Vision (ECCV).*

[9] Yong Du, Wei Wang, and Liang Wang. 2015. Hierarchical recurrent neural network for skeleton based action recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015.* IEEE Computer Society, 1110–1118.

[10] Sean Ryan Fanello, Ilaria Gori, Giorgio Metta, and Francesca Odone. 2013. Keep it simple and sparse: real-time action recognition. *J. Mach. Learn. Res.* 14, 1 (2013), 2617–2640.

[11] Qing Guo, Wei Feng, Ce Zhou, Rui Huang, Liang Wan, and Song Wang. 2017. Learning Dynamic Siamese Network for Visual Object Tracking. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV).*

[12] Anfeng He, Chong Luo, Xinmei Tian, and Wenjun Zeng. 2018. A Twofold Siamese Network for Real-Time Object Tracking. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018.* Computer Vision Foundation / IEEE Computer Society, 4834–4843.

[13] Qiuhong Ke, Mohammed Bennamoun, Senjian An, Ferdous Ahmed Sohel, and Farid Boussaïd. 2017. A New Representation of Skeleton Sequences for 3D Action Recognition. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017.* IEEE Computer Society, 4570–4579.

[14] Jun Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, and Alex C. Kot. 2020. NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 10 (2020), 2684–2701.

[15] Mengyuan Liu, Hong Liu, and Chen Chen. 2017. Enhanced skeleton visualization for view invariant human action recognition. *Pattern Recognit.* 68 (2017), 346–362.

[16] Saemi Moon, Myeonghyeon Kim, Zhenyue Qin, Yang Liu, and Dongwoo Kim. 2023. Anonymization for Skeleton Action Recognition. AAAI Press. https://doi.org/10.1609/aaai.v37i12.26754

[17] Ilesanmi Olade, Charles Fleming, and Hai-Ning Liang. 2020. BioMove: Biometric User Identification from Human Kinesiological Movements for Virtual Reality Systems. *Sensors* 20, 10 (2020), 2944.

[18] Alessia Saggese, Nicola Strisciuglio, Mario Vento, and Nicolai Petkov. 2019. Learning skeleton representations for human action recognition. *Pattern Recognit. Lett.* 118 (2019), 23–31.

[19] Amir Shahroudy, Jun Liu, Tian-Tsong Ng, and Gang Wang. 2016. NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016.* IEEE Computer Society, 1010–1019.

[20] Chenyang Si, Ya Jing, Wei Wang, Liang Wang, and Tieniu Tan. 2018. Skeleton-Based Action Recognition with Spatial Reasoning and Temporal Stack Learning. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 11205)*, Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (Eds.). Springer, 106–121.

[21] Aniruddha Sinha, Kingshuk Chakravarty, and Brojeshwar Bhowmick. 2013. Person identification using skeleton information from kinect.

[22] Cunli Song and Shouyong Ji. 2022. Face Recognition Method Based on Siamese Networks Under Non-Restricted Conditions. *IEEE Access* 10 (2022), 40432–40444. https://doi.org/10.1109/ACCESS.2022.3167143

[23] Sijie Song, Cuiling Lan, Junliang Xing, Wenjun Zeng, and Jiaying Liu. 2017. An End-to-End Spatio-Temporal Attention Model for Human Action Recognition from Skeleton Data. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, Satinder Singh and Shaul Markovitch (Eds.). AAAI Press, 4263–4270.

[24] Yansong Tang, Yi Tian, Jiwen Lu, Peiyang Li, and Jie Zhou. 2018. Deep Progressive Reinforcement Learning for Skeleton-Based Action Recognition. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018.* Computer Vision Foundation / IEEE Computer Society, 5323–5332.

[25] Du Tran, Heng Wang, Lorenzo Torresani, Jamie Ray, Yann LeCun, and Manohar Paluri. 2018. A Closer Look at Spatiotemporal Convolutions for Action Recognition. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018.* Computer Vision Foundation / IEEE Computer Society, 6450–6459.

[26] Hongsong Wang and Liang Wang. 2018. Learning content and style: Joint action recognition and person identification from human skeletons. *Pattern Recognit.* 81 (2018), 23–35.

[27] Junwu Weng, Mengyuan Liu, Xudong Jiang, and Junsong Yuan. 2018. Deformable Pose Traversal Convolution for 3D Action and Gesture Recognition. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VII (Lecture Notes in Computer Science, Vol. 11211)*, Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (Eds.). Springer, 142–157.

[28] Haoran Wu, Zhiyong Xu, Jianlin Zhang, Wei Yan, and Xiao Ma. 2017. Face recognition based on convolution siamese networks. In *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI).* 1–5. https://doi.org/10.1109/CISP-BMEI.2017.8302003

[29] Wanghui Xiao and Yuting Ding. 2022. A Two-Stage Siamese Network Model for Offline Handwritten Signature Verification. *Symmetry* 14, 6 (2022), 1216.

[30] Sijie Yan, Yuanjun Xiong, and Dahua Lin. 2018. Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, Sheila A. McIlraith and Kilian Q. Weinberger (Eds.). AAAI Press, 7444–7452.

[31] Pengfei Zhang, Cuiling Lan, Junliang Xing, Wenjun Zeng, Jianru Xue, and Nanning Zheng. 2017. View Adaptive Recurrent Neural Networks for High Performance Human Action Recognition from Skeleton Data. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017.* IEEE Computer Society, 2136–2145.

[32] Pengfei Zhang, Cuiling Lan, Junliang Xing, Wenjun Zeng, Jianru Xue, and Nanning Zheng. 2019. View Adaptive Neural Networks for High Performance Skeleton-Based Human Action Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 8 (2019), 1963–1978.

[33] Pengfei Zhang, Cuiling Lan, Wenjun Zeng, Junliang Xing, Jianru Xue, and Nanning Zheng. 2021. Multi-Scale Semantics-Guided Neural Networks for Efficient Skeleton-Based Human Action Recognition. *CoRR* abs/2111.03993 (2021).

[34] Wentao Zhu, Cuiling Lan, Junliang Xing, Wenjun Zeng, Yanghao Li, Li Shen, and Xiaohui Xie. 2016. Co-Occurrence Feature Learning for Skeleton Based Action Recognition Using Regularized Deep LSTM Networks. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, Dale Schuurmans and Michael P. Wellman (Eds.). AAAI Press, 3697–3704.