Anonymization for Skeleton Action Recognition

Saemi Moon[1*], Myeonghyeon Kim[3*], Zhenyue Qin[4,5], Yang Liu[4], Dongwoo Kim[1,2]

[1]Computer Science and Engineering, Pohang University of Science and Technology
[2]Graduate School of Artificial Intelligence, Pohang University of Science and Technology
[3]Scatter Lab
[4]Australian National University
[5]Tencent
{saemi, dongwookim}@postech.ac.kr, nessunkim@gmail.com, {zhenyue.qin, yang.liu3}@anu.edu.au

Abstract

Skeleton-based action recognition attracts practitioners and researchers due to the lightweight, compact nature of datasets. Compared with RGB-video-based action recognition, skeleton-based action recognition is a safer way to protect the privacy of subjects while having competitive recognition performance. However, due to improvements in skeleton recognition algorithms as well as motion and depth sensors, more details of motion characteristics can be preserved in the skeleton dataset, leading to potential privacy leakage. We first train classifiers to categorize private information from skeleton trajectories to investigate the potential privacy leakage from skeleton datasets. Our preliminary experiments show that the gender classifier achieves 87% accuracy on average, and the re-identification classifier achieves 80% accuracy on average with three baseline models: Shift-GCN, MS-G3D, and 2s-AGCN. We propose an anonymization framework based on adversarial learning to protect potential privacy leakage from the skeleton dataset. Experimental results show that an anonymized dataset can reduce the risk of privacy leakage while having marginal effects on action recognition performance even with simple anonymizer architectures. The code used in our experiments is available at https://github.com/ml-postech/Skeleton-anonymization/

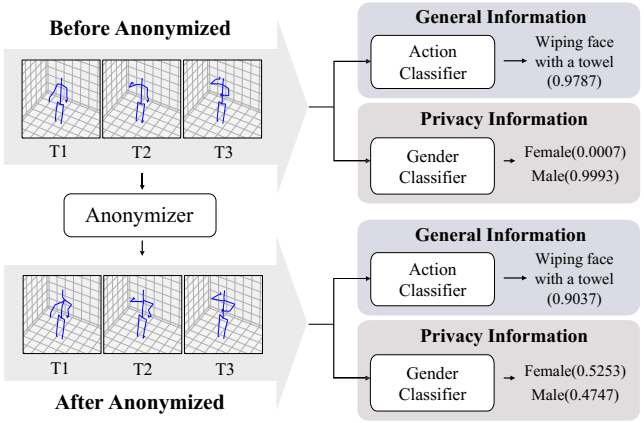


Figure 1: One example of privacy leakage from the ETRI-activity3D dataset. A gender classifier is employed to identify the gender from a skeleton sequence, exposing the private gender information with high accuracy. After being anonymized, the action classifier can still recognize the action accurately while the privacy classifier hides the original gender information.

Introduction

Action recognition has been widely studied in many applications such as sports analysis (Tran et al. 2018), human-robot interaction (Fanello et al. 2013), and intelligent healthcare services (Saggese et al. 2019). Due to the success of convolutional neural networks, many recognition approaches are proposed based on a sequence of video frames. Action recognition can further be used for the public good. For example, with surveillance cameras in a public area or a school, we can detect violent actions.

To employ the recognition system appropriately, one must ensure that private information is not abused before and after analysis. Skeleton-based action recognition can be alternative to video-based recognition. Due to the advance in depth and motion sensors, details of motion characteristics can be preserved in the skeleton dataset. Compared with RGB videos, the skeleton dataset seems to expose fewer details on participants. It is often challenging to identify sensitive information such as gender or age from a skeleton to compare with the RGB video to the naked eye.

We raise a question about the privacy-safeness of skeleton datasets. To check potential privacy leakage from skeletons, we conduct experiments on identifying gender or identity with Shift-GCN (Cheng et al. 2020), MS-G3D (Liu et al. 2020), and 2s-AGCN (Shi et al. 2019). Based on our analysis, a properly trained classifier can predict private information accurately. Therefore, the skeletons are not safe from the privacy leakage problem. A previous study (Sinha, Chakravarty, and Bhowmick 2013) confirms the possibility of identifying a person from skeletons extracted from Kinect. Also, prior work (Wang and Wang 2018) suggests an end-to-end framework to predict both action and identity recognition tasks.

This work aims to develop a framework that can anonymize skeleton datasets while preserving critical action

*These authors contributed equally.

features for recognition. To this end, we propose a minimax framework to anonymize the skeletons. With RGB-video datasets, object detection followed by blurring or inpainting with pre-trained generative models is often employed to anonymize datasets (Yang et al. 2021; Hukkelås, Mester, and Lindseth 2019). However, these methods cannot be directly applied to the skeleton dataset.

The minimax framework consists of an anonymizer network with two sub-networks designed to predict action and private information. The anonymizer removes private information from skeletons, and then the output skeleton is fed into action and privacy classifiers separately. We maximize the accuracy of the action classifier while minimizing the identifiability of private information with the other classifier. In addition, we enforce the anonymized skeleton similar to the original one to make sure they are visually indistinguishable from each other. To solve the minimax problem, we propose an adversarial learning algorithm. Experimental results show that the proposed algorithm results in an effective anonymizer.

We summarize our contributions as follows:

- We empirically show potential privacy leakage from widely-used skeleton datasets such as NTU60 (Shahroudy et al. 2016) and ETRI-activity3D(Jang et al. 2020).
- We develop a skeleton anonymization network based on action and sensitive variable classifiers.
- We propose a learning algorithm based on the adversarial learning method to anonymize skeletons.
- We show that the anonymized skeletons are more robust to privacy leakage while still enjoying high action recognition accuracy.

Skeleton Anonymization

In this section, we propose a framework for the skeleton anonymization model.

Anonymization framework. Let $\vec{x} \in \mathbb{R}^{T\times D\times 3}$ be 3D coordinates of D joints over T frames, and $y \in \mathcal{Y}$ be an action label for a given skeleton sequence $\vec{x}$, where $\mathcal{Y}$ is a set of actions to be recognized. Let $z \in \mathcal{Z}$ be private information related to the skeleton sequence $\vec{x}$, e.g., gender or identity, where $\mathcal{Z}$ is a set of possible private labels.

We aim to develop an anonymization network that can effectively remove private information from skeleton datasets while maintaining the recognizability of actions from the anonymized skeletons. To do this, we propose a minimax framework consisting of three different neural network components. Let $f_\theta : \mathbb{R}^{T\times D\times 3} \to \mathbb{R}^{T\times D\times 3}$ be an anonymizer network aiming to remove sensitive information from the input skeletons, $h_\psi : \mathbb{R}^{T\times D\times 3} \to \mathcal{Y}$ be an action classifier, and $g_\phi : \mathbb{R}^{T\times D\times 3} \to \mathcal{Z}$ be a privacy classifier that predicts sensitive personal information. Our goal is to train an anonymizer f_θ whose output can maximally confuse the classification performance on the private variables. On the other hand, the output of the anonymizer should keep all relevant information for recognizing action to preserve the performance of the action classifier h_ψ. In other words, the output should not be very different from the original skeletons since the anonymized skeletons can be recognizable by the naked eye. To satisfy all requirements, we formalize the anonymization via the following minimax objective:

$$\min_\theta \max_\phi \mathbb{E}\Big[\mathrm{CE}\left(y, h_\psi\left(f_\theta(\vec{x})\right)\right) - \alpha\, \mathrm{CE}\left(z, g_\phi\left(f_\theta(\vec{x})\right)\right) + \beta||\vec{x} - f_\theta(\vec{x})||_2^2 \Big], \tag{1}$$

where CE is the cross entropy, and α and β are hyperparameters controlling the importance of the privacy classification and the reconstruction error, respectively. The reconstruction error between the original and anonymized skeleton data $||\vec{x} - f_\theta(\vec{x})||_2^2$ ensures the anonymized skeletons are similar to the original ones. To maximize the objective, the private classifier needs to classify the private label z correctly. To minimize the objective, the anonymizer makes the actions easily identifiable by action classifier h_ψ while making the private classifier misclassify the private label z. To simplify the learning process, we use a pre-trained action classifier and fix the parameters of the action classifier during training. The fixed action classifier constrains the anonymized skeleton compatible with the pre-trained model. The anonymized skeletons are also likely to work well with other pre-trained classifiers available.

Minimizing the objective w.r.t θ can make the anonymizer fool the private classifier. However, one may exploit this fact to infer the true label. For example, in a binary classification problem, the true label can be obtained by choosing the opposite of the prediction. To avoid this issue, we minimize the entropy of classified outputs during the minimization step:

$$\min_\theta \mathcal{L}_{\text{adv}} = \min_\theta \mathbb{E}\Big[\mathrm{CE}\left(y, h_\psi\left(f_\theta(\vec{x})\right)\right) - \alpha H\left(g_\phi\left(f_\theta(\vec{x})\right)\right) + \beta||\vec{x} - f_\theta(\vec{x})||_2^2 \Big], \tag{2}$$

where $H\left(g_\phi\left(f_\theta(\vec{x})\right)\right)$ is the entropy of the distribution of private labels predicted from the anonymized skeleton. Therefore, the optimal anonymizer yields the most confusing skeletons to the private classifier. In the maximization step, we still maximize the negative cross entropy $-\alpha\, \mathrm{CE}\left(z, g_\phi\left(f_\theta(\vec{x})\right)\right)$ w.r.t. ϕ to train the private classifier. Figure 2 shows the overall framework for data anonymization.

Alternating minimization and maximization are often employed to solve a minimax objective as shown in the generative adversarial network (Goodfellow et al. 2014). Following previous work, we also use the alternating algorithm to optimize the objective. Algorithm 1 shows the overall training algorithm. In this work, the adversarial learning algorithm starts with pre-trained classifiers g_ϕ and h_ψ to make the learning stable.

Anonymizer networks. The anonymizer f_θ can be any prediction model that modifies skeletons while preserving

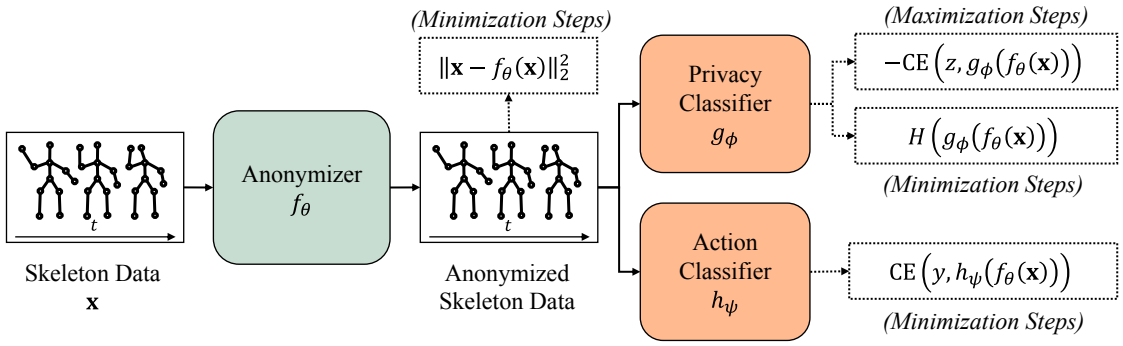


Figure 2: Anonymization framework. The framework consists of three sub-networks: 1) anonymizer f_θ, 2) privacy classifier g_ϕ, and 3) action classifier h_ψ. The dashed box represents the losses used in the minimization and maximization steps with adversarial learning. Note that the privacy classifier uses a separate loss for minimization and maximization in an adversarial learning setup. The parameter of the action classifier ψ is pre-trained and not updated during anonymizer training.

Algorithm 1: Adversarial Anonymization

Require: Pre-trained classifiers h_ψ and g_ϕ, E: # of epochs, m: minibatch size, k: # of minimization steps
while until convergence **do**
  for $t \leftarrow 1$ to k **do**
    Sample minibatch of m samples $\{(\vec{x}_i, y_i, z_i)\}_{i=1}^m$
    Compute $\nabla_\theta \mathcal{L}_{\text{adv}}$ with minibatch ▷ Equation 2
    Update $\theta \leftarrow \theta - \nabla_\theta \mathcal{L}_{\text{adv}}$
  end for
  Sample minibatch of m samples $\{(\vec{x}_i, y_i, z_i)\}_{i=1}^m$
  Compute $\nabla_\phi \alpha \,\mathrm{CE}\left(z, g_\phi\left(f_\theta(\vec{x})\right)\right)$ with minibatch
  Update $\phi \leftarrow \phi - \nabla_\phi \alpha \,\mathrm{CE}\left(z, g_\phi\left(f_\theta(\vec{x})\right)\right)$
end while

the original dimension. We employ two simple neural network architectures for the anonymizer: 1) residual networks and 2) U-net architectures.

First, the residual network (He et al. 2016) anonymizer adopts a simple residual connection from the input skeletons to the output skeletons. Specifically, the model can be formalized as

$$f_\theta(\vec{x}) = \mathrm{MLP}_\theta(\vec{x}) + \vec{x},$$

where $\mathrm{MLP}_\theta : \mathbb{R}^{D\times 3} \rightarrow \mathbb{R}^{D\times 3}$ is a simple multi-layered perceptron parameterized by θ. The residual connection keeps the position similar to the original skeleton while the MLP layer models the disposition of joints to anonymize. We use two fully-connected layers to model the disposition. The anonymizer is applied to each frame of a skeleton sequence. Although the anonymizer is applied to each frame independently, the back-propagated signals from action and private classifiers make the entire sequence coherent. By initializing θ with weights close to zero, we make the anonymizer add a small random noise to the original skeleton in the early stage of learning.

Second, the U-Net (Ronneberger, Fischer, and Brox 2015) architecture is adopted to our anonymizer network. The U-Net consists of two paths: the contracting path, and the expanding path. In the contracting path, it repeats downsampling and maxpool an input skeleton data to encode it to the feature map. In the expanding path, U-Net repeats upsampling and concatenating feature maps via skip connection. Especially, skip connections concatenates the features from the contracting path to the corresponding level in the expanding path. It makes the output skeleton position similar to the original skeleton.

Experiments

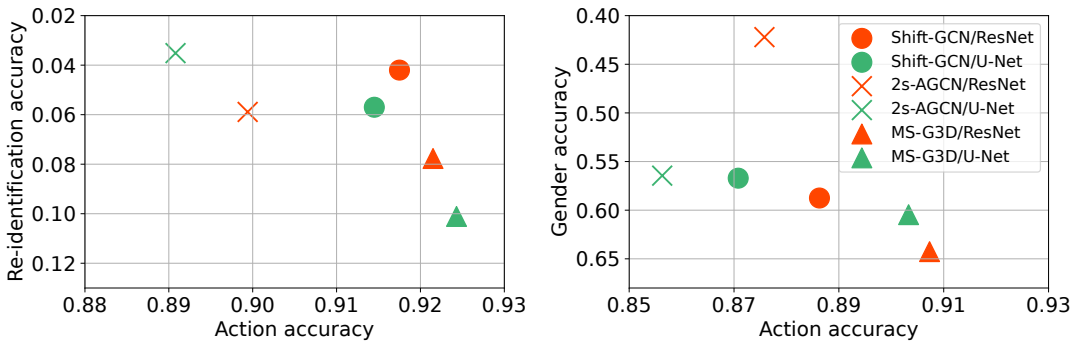
In this section, we demonstrate the performance of the proposed framework for anonymizing skeleton datasets. We use two publicly available datasets and anonymize two different types of private information: gender and identity.

Datasets

We use two datasets: ETRI-activity3D (Jang et al. 2020) and NTU RGB+D 60 (NTU60) (Shahroudy et al. 2016). For the ETRI-activity3D dataset, we anonymize the gender information from the skeletons. For the NTU60 dataset, we anonymize the identity of the skeletons. The detailed experimental setups for these datasets are as follows.

ETRI-activity3D. ETRI-activity3D is an action recognition dataset originally published for recognizing the daily activities of the elderly and youths. It contains 112,620 skeleton sequence samples observed from 100 people, half of whom were between the ages of 64 and 88 and the rest were in their 20s. The elderly consist of 33 females and 17 males, and the young adults consist of 25 females and 25 males. The samples are categorized into 55 classes based on the activity type. Each sequence consists of 3D locations of 25 joints of the human body.

With the ETRI-activity3D dataset, we anonymize the gender information from the skeletons. We drop samples from 5 classes for the following experiments, e.g., handshaking, containing two people, so only one person appears in the remaining samples. After removing malformed and two-person samples, we split the remaining samples into 68,788 and 34,025 training and validation, respectively. We split the



(a) Re-identification task

(b) Gender classification

Figure 3: Action and privacy accuracy of three baseline models with two different anonymizers after anonymization. y-axis is reversed. Note that, before anonymization, the average top-1 re-identification accuracy is 80%, and the average gender classification accuracy is 87%.

dataset according to the subject ID. In other words, the subjects in the validation set do not appear in the training set. Through this split, we measure the generalizability of the gender classifier to the unknown subjects.

NTU60. NTU60 is an action recognition dataset that contains 60 action classes and 56,880 skeleton sequences taken from 40 subjects. The format of skeleton data is the same as ETRI-activity3D, which includes the 3D positions of 25 human body joints. After removing malformed samples, we split the remaining samples into 37,646 and 18,932 as training and validation sets, respectively. Following the original work (Shahroudy et al. 2016), we split it according to the camera ID so that both sets contain identical subjects with different views.

Privacy Leakage

To verify privacy leakage from each dataset, we first check the performance of the gender classification and re-identification task. To train gender classifier and re-identification task, three popular baseline models, Shift-GCN (Cheng et al. 2020), MS-G3D (Liu et al. 2020), and 2s-AGCN (Shi et al. 2019), are adopted. For the gender classifier with MS-G3D, we use MS-G3D without a G3D module. This makes training faster without losing too much accuracy. We train multiple times for the gender classifier and re-identification task. Each model is trained with a different random initialization.

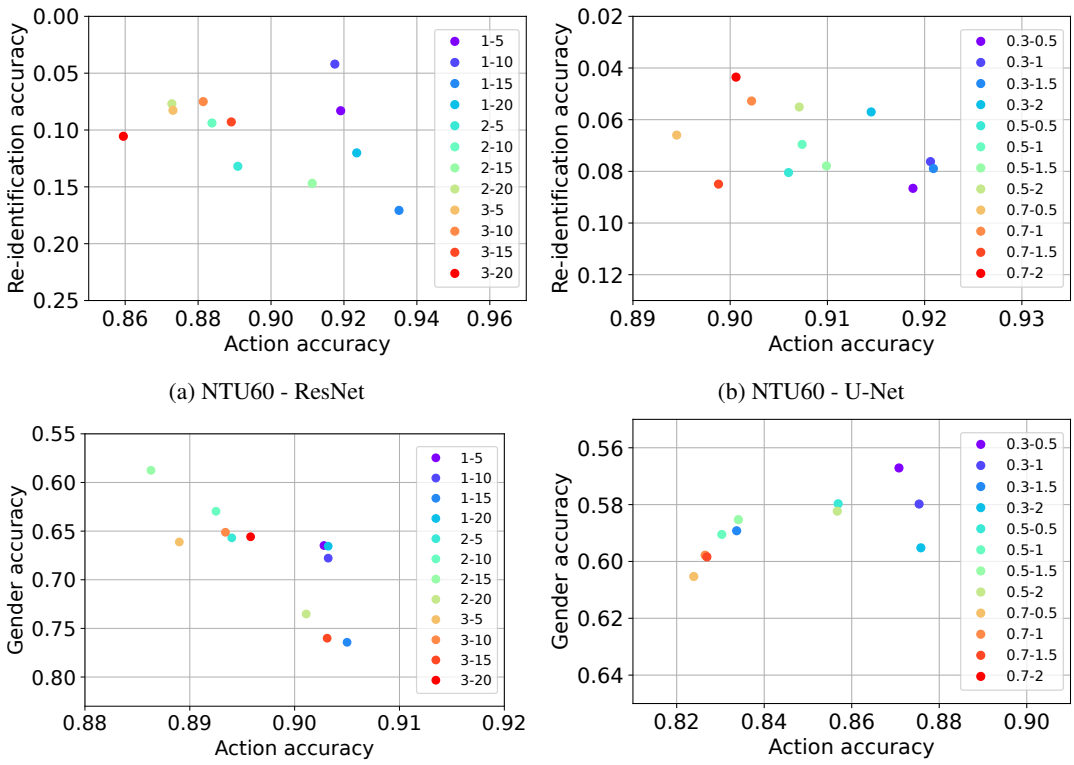
After training, the gender classifier achieves 87% accuracy on average. The re-identification task achieves 80% and 97% for top-1 and top-5 accuracy, respectively. The detailed results are available in Appendix. As the results suggest, the privacy information can be easily predicted by a classification model trained with private labels. Note that the test splits do not contain the subject used for gender classification. This reveals the generalizability of gender classification to unseen subjects. Also, for the re-identification task, the train split and test split have different camera IDs, so the same person appears in both sets with different views. This result indicates that the joint trajectory contains personal traits that can be easily exploited to identify a person.

Anonymization Results

Our preliminary study indicates that gender and identification can leak enough from training. Based on the results obtained in the previous experiments, we evaluate the performance of anonymization with an adversarial learning algorithm. As mentioned earlier, experiments are conducted with two anonymizer networks (ResNet (He et al. 2016), U-Net (Ronneberger, Fischer, and Brox 2015)), two datasets (NTU60 (Shahroudy et al. 2016), ETRI-activity3D (Jang et al. 2020)), and three baseline models (Shift-GCN (Cheng et al. 2020), MS-G3D (Liu et al. 2020), and 2s-AGCN (Shi et al. 2019)). For each task, we use two pre-trained classifiers for action and privacy, respectively. One classifier is used to initialize the adversarial algorithm, and the other is used to measure the accuracy after anonymization.

Figure 3 shows the results of anonymization with re-identification task and gender classification. In general, we observe that we can dramatically decrease privacy accuracy while minimally sacrificing action recognition accuracy. We also observe more leakage of private information when the action accuracy is relatively higher. Note that we use a balanced test set for identity anonymization. Since NTU60 has 40 subjects, one can achieve 2.5% accuracy with random classification. For the gender classification, one can achieve 50% accuracy with a random classifier on the test set.

One would expect the trade-off between action accuracy and privacy accuracy based on the choice of hyperparameters α and β. The choice of the best anonymization model may vary depending on the application. In this work, we report the performance of the best model based on *action accuracy* $\times$ *(1 - re-identification accuracy)* from the results of various configurations. According to our metric, ResNet (α:1, β:10) and U-Net (α:0.3, β:2) models are chosen as representative model. We provide additional results with different configurations in the next section. Note that one may use a different metric to select a model given different application scenarios. We use the best configuration obtained from the baseline model, Shift-GCN, to train the other baseline



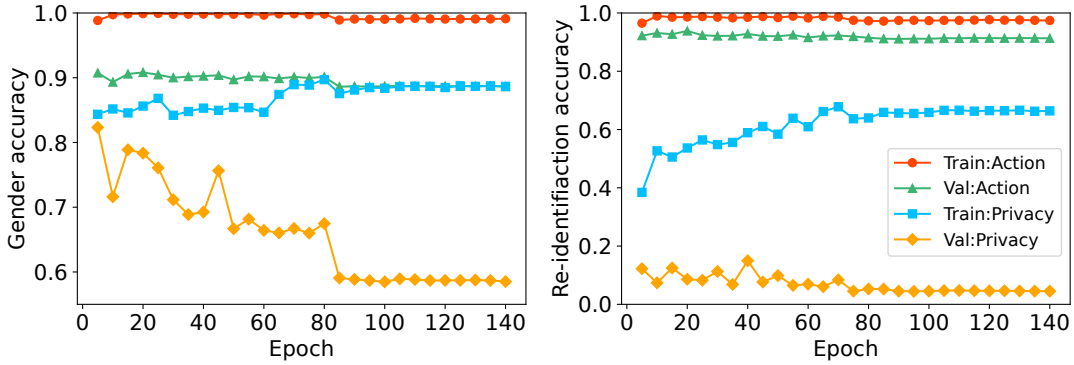
(a) NTU60 - ResNet

(b) NTU60 - U-Net

(c) ETRI-activity3D - ResNet

(d) ETRI-activity3D - U-Net

Figure 4: The trade-off between action accuracy and privacy accuracy based on a different configuration of hyperparameter α and β on NTU60 and ETRI-activity3D with two anonymizer networks (legend: α-β). Note that the y-axis is reversed.



(a) Gender classification

(b) Re-identification task

Figure 5: Accuracy and reconstruction error over epochs with the residual anonymizer. 'Train:Action' and 'Val:Action' indicate the training and validation accuracy of the action classification, and 'Train:Privacy' and 'Val:Privacy' indicate the training and validation accuracy of the privacy classification.

models. The detailed configuration of hyperparameters used to plot Figure 3 is available in Appendix.

Trade-off analysis. We vary the value of α and β to check the trade-off between action accuracy and privacy leakage based on different configurations of hyperparameters. We use Shift-GCN as a baseline model for analysis. Figure 4a and Figure 4b show the result with various hyperparameter configurations on the identity anonymization task. Figure 4c and Figure 4d show the result of the gender anonymization task. We can observe that given a fixed α, increasing β increases the chance of privacy leakage as well as the action accuracy showing the presence of the trade-off between the action accuracy and privacy leakage.

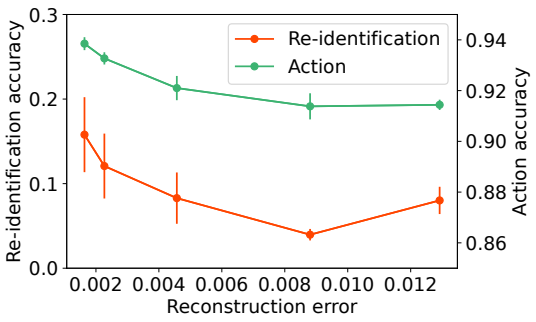


Figure 6: The trade-off between reconstruction error and accuracy with the residual anonymizer on NTU60.

We additionally plot how the validation accuracy changes over training procedure with adversarial learning in Figure 5. For both datasets, the action accuracy remains high over the epochs on both training and validation sets. However, there is a gap between the training and validation accuracy on privacy, which shows the overfitting in the classifier. Specifically, for gender classification, gender accuracy starts with similar values. Then gender accuracy increases on the train set and decreases on the validation set. The re-identification accuracy drops first on both training and validation sets for the re-identification task. Note that the re-identification task achieves 80% accuracy. The validation accuracy at the first epoch indicates that the re-identification task is more sensitive to the additional noise introduced by random weights of the residual network than the gender classification.

Reconstruction error analysis. A reconstruction error directly shows the difference between the original and anonymized skeletons. Although we cannot directly set the level of reconstruction error, we vary the parameters to obtain different levels of reconstruction error and corresponding prediction accuracy. Please check the Appendix for the detailed hyperparameter settings. As shown in Figure 6, there is a trade-off between reconstruction error and re-identification accuracy. As we increase the reconstruction error, we can reduce the re-identification accuracy. However, high reconstruction error yields low action accuracy as well.

Comparison with alternative approaches. Since we propose privacy leakage for the first time, no anonymization method removes privacy information while remaining action accuracy high. Therefore, we consider two alternative approaches to anonymize privacy information by modifying skeleton data potentially. (1) Random noise: As a baseline, we randomly inject white noise drawn from the zero mean normal distributions with varying variances to the original skeleton. (2) Adversarial attack method: Adversarial attack is a technique that makes a model fool by perturbing input data. There are several adversarial attack research on skeleton action recognition (Liu, Akhtar, and Mian 2020; Wang et al. 2021; Diao et al. 2021; Tanaka, Kera, and Kawamoto 2022; Zheng et al. 2020). We use Wang et al. (2021) method to attack privacy information. Note that we select Shift-GCN as a baseline model in this experiment.

Method		Action.	Iden.
Not-anonymized		0.9510	0.8095
Random noise	$\sigma = 0.001$	0.7565	0.7450
	$\sigma = 0.005$	0.4430	0.3240
	$\sigma = 0.010$	0.2660	0.1735
	$\sigma = 0.020$	0.1265	0.1020
	$\sigma = 0.050$	0.0455	0.0840
	$\sigma = 0.100$	0.0450	0.0715
Adversarial attack	Attacked	0.9435	0.0000
	Non-Attacked	0.9435	0.3621
Our method		0.9175	0.0420

Table 1: Comparison between our method and other alternative approaches. This experiment is conducted with the residual anonymizer on NTU60.

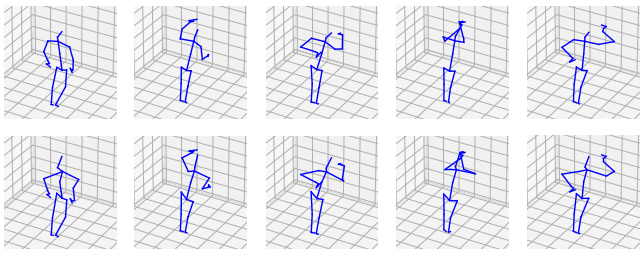


Figure 7: Five frames of the original (top) and the gender anonymized (bottom) skeletons for an action "wiping face with a towel" from ETRI-activity3D. The subject is an elderly female.

Table 1 shows the results of comparing our method to other approaches. The random noise cannot preserve action information while reducing privacy leakage. The results with an adversarial attack show that attacked skeleton data succeeds in removing privacy information from the target model, i.e., the identification accuracy of the attacked model is zero. However, identification accuracy remains relatively high for the other pre-trained model, which has not been attacked. The adversarial attack-based anonymization is model-specific and difficult to generalize to the unseen models, whereas anonymized skeleton data with our proposed framework performs relatively well with any pre-trained model.

Qualitative analysis. To qualitatively understand the effect of anonymization, we visualize one example from the ETRI-activity3D dataset before and after anonymization in Figure 7. The top and bottom rows show five selected frames before and after anonymization for each figure, respectively. We can find some interesting patterns in the visualization. For example, the length of the neck bone is slightly increased, and the bone is moved to the upright position after anonymization. Given that an elderly female acts, we can conjecture that the adjustment makes gender unrecognizable. More visualization results are provided in Appendix.

Related Work

Our work lies in public data anonymization and skeleton-based action recognition. In this section, we provide the previous attempts at dataset anonymization and skeleton-based action recognition. Also, we mention other research about adversarial attacks in skeleton data

Public Dataset Anonymization

Researchers have pointed out privacy issues with public visual datasets and tried to mitigate them. Caesar et al. (2020); Frome et al. (2009); Yang et al. (2021) propose a blurring approach where the privacy-sensitive regions are blurred with an object detection method. Flores and Belongie (2010); Uittenbogaard et al. (2019) propose an inpainting method to remove potentially problematic objects such as pedestrians and vehicles. A large body of prior work has used GANs(Goodfellow et al. 2014) to preserve visual private information. Ren, Lee, and Ryoo (2018); Maximov, Elezi, and Leal-Taixé (2020); Hukkelås, Mester, and Lindseth (2019) use GANs to generate fake faces to replace real ones. Also, Gu et al. (2020) proposes a face identity transformer that anonymizes face information according to the given password.

There are also some works that exist for other domains: sound domain (Cohen-Hadria et al. 2019; Sümer et al. 2020) and text domain (Li, Baldwin, and Cohn 2018; Coavoux, Narayan, and Cohen 2018; Mosallanezhad, Beigi, and Liu 2019). Similar concerns are also made for skeleton datasets. Sinha, Chakravarty, and Bhowmick (2013) propose a method to recognize persons from skeleton data. This work focuses on gait patterns extracted from human skeletons. The authors build a model with some predefined features and tested an adaptive neural network and naïve Bayes classifier to recognize the identity of persons. This implies the potential privacy leakage from public datasets.

Skeleton-Based Action Recognition

Human skeleton data is a sequence of graphs, where joints and bones are represented as nodes and edges separately within a graph. In early times, skeleton motion trajectories are embedded into a manifold space as points. The relative distances between these points acted as clues for action recognition. However, these models do not exploit the internal spatial relationship between joints. Later, convolution neural networks (CNNs) are utilized to extract spatial co-occurrence patterns between joints. Nevertheless, CNNs cannot model a skeleton's topological information.

Then, graph convolution networks (GCNs) are introduced to model these topological relations. Nonetheless, basic GCNs are not suitable for human skeleton sequences because they contain not only the 3D position of joints but also the time series. Yan et al. introduce the spatial-temporal graph convolutional networks (ST-GCN) (Yan, Xiong, and Lin 2018). They conduct graph convolution for extracting spatial features and perform 1×1 convolution over each joint for capturing temporal variations. Following this line, various graph neural architectures are proposed to extract features from the graphs. AS-GCN (Li et al. 2019) applies parametric adjacency matrices to substitute for the fixed skeleton graph. Subsequently, AGC-LSTM (Si et al. 2019) incorporates graph convolution layers into a long short-term memory network (LSTM) as gate operations to capture long-range temporal movements in action sequences. The 2s-AGCN model (Shi et al. 2019) proposes bone features and learnable residual masks to enhance more flexibly extracting skeletons' structural information and ensembles the models trained separately with joints and bones to improve the classification accuracy. MS-G3D (Liu et al. 2020) introduces cross-spacetime skip connections, which additionally connect all 1-hop neighbors across all time frames of a dilated sliding window for direct information flow, and show improved recognition performance. Shift-GCN (Cheng et al. 2020) applies shift graph operations and lightweight pointwise convolutions to overcome computational complexity and inflexible receptive fields of prior GCN-based studies.

In this work, we use Shift-GCN(Cheng et al. 2020), MS-G3D(Liu et al. 2020), and 2s-AGCN(Shi et al. 2019) as a baseline recognition model for private information. Although the original model is developed to recognize the actions of skeletons, we empirically show the model can successfully classify private information with a proper training procedure.

Adversarial Attacks in Skeleton Data

Although deep neural network models achieve high performance in many tasks, there are main concerns about robustness. Researchers have investigated that imperceptible perturbed input data can easily deceive the deep neural network(Athalye et al. 2018). Recently, adversarial attacks have been attempted at skeleton-based action recognition(Liu, Akhtar, and Mian 2020; Wang et al. 2021; Diao et al. 2021; Tanaka, Kera, and Kawamoto 2022; Zheng et al. 2020). These researches show that adversarial attacks successfully fool the state-of-art models with small perturbations. In this work, we use the adversarial attack method to attack privacy information and compare the result with our method.

Conclusion

In this work, we investigate privacy leakage from publicly available skeleton datasets. We show that although skeleton data may seemingly be privacy-protective, recently proposed skeletal action recognizers are surprisingly capable of extracting sensitive and identity information from these data. To address this privacy leakage problem, we propose a learning framework. Our experimental results reveal that the proposed method effectively removes the privacy information while preserving the movement patterns. Note that the anonymizers used in this work employ relatively simple architectures. Experiments show that private information can be removed effectively even with simple architectures. We leave the study of more advanced architectures for future work since our goal is to show the potential vulnerability of the skeletons and to provide a general framework to overcome.

Acknowledgements

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2019-0-01906, Artificial Intelligence Graduate School Program(POSTECH)) and National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2021R1C1C1011375). Dongwoo Kim is the corresponding author.

References

Athalye, A.; Engstrom, L.; Ilyas, A.; and Kwok, K. 2018. Synthesizing robust adversarial examples. In *International conference on machine learning*, 284–293. PMLR.

Caesar, H.; Bankiti, V.; Lang, A. H.; Vora, S.; Liong, V. E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; and Beijbom, O. 2020. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11621–11631.

Cheng, K.; Zhang, Y.; He, X.; Chen, W.; Cheng, J.; and Lu, H. 2020. Skeleton-based action recognition with shift graph convolutional network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 183–192.

Coavoux, M.; Narayan, S.; and Cohen, S. B. 2018. Privacy-preserving Neural Representations of Text. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 1–10. Brussels, Belgium: Association for Computational Linguistics.

Cohen-Hadria, A.; Cartwright, M.; McFee, B.; and Bello, J. P. 2019. Voice anonymization in urban sound recordings. In *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, 1–6. IEEE.

Diao, Y.; Shao, T.; Yang, Y.-L.; Zhou, K.; and Wang, H. 2021. BASAR: Black-box attack on skeletal action recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7597–7607.

Fanello, S. R.; Gori, I.; Metta, G.; and Odone, F. 2013. Keep It Simple And Sparse: Real-Time Action Recognition. *Journal of Machine Learning Research*, 14.

Flores, A.; and Belongie, S. 2010. Removing pedestrians from google street view images. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, 53–58. IEEE.

Frome, A.; Cheung, G.; Abdulkader, A.; Zennaro, M.; Wu, B.; Bissacco, A.; Adam, H.; Neven, H.; and Vincent, L. 2009. Large-scale privacy protection in google street view. In *2009 IEEE 12th international conference on computer vision*, 2373–2380. IEEE.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. *Advances in neural information processing systems*, 27.

Gu, X.; Luo, W.; Ryoo, M. S.; and Lee, Y. J. 2020. Password-conditioned anonymization and deanonymization with face identity transformers. In *European Conference on Computer Vision*, 727–743. Springer.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hukkelås, H.; Mester, R.; and Lindseth, F. 2019. Deepprivacy: A generative adversarial network for face anonymization. In *International Symposium on Visual Computing*, 565–578. Springer.

Jang, J.; Kim, D.; Park, C.; Jang, M.; Lee, J.; and Kim, J. 2020. ETRI-activity3D: A large-scale RGB-D dataset for robots to recognize daily activities of the elderly. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 10990–10997. IEEE.

Li, M.; Chen, S.; Chen, X.; Zhang, Y.; Wang, Y.; and Tian, Q. 2019. Actional-structural graph convolutional networks for skeleton-based action recognition. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 3595–3603.

Li, Y.; Baldwin, T.; and Cohn, T. 2018. Towards Robust and Privacy-preserving Text Representations. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 25–30. Melbourne, Australia: Association for Computational Linguistics.

Liu, J.; Akhtar, N.; and Mian, A. 2020. Adversarial attack on skeleton-based human action recognition. *IEEE Transactions on Neural Networks and Learning Systems*.

Liu, Z.; Zhang, H.; Chen, Z.; Wang, Z.; and Ouyang, W. 2020. Disentangling and Unifying Graph Convolutions for Skeleton-Based Action Recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Maximov, M.; Elezi, I.; and Leal-Taixé, L. 2020. Ciagan: Conditional identity anonymization generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5447–5456.

Mosallanezhad, A.; Beigi, G.; and Liu, H. 2019. Deep reinforcement learning-based text anonymization against private-attribute inference. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2360–2369.

Ren, Z.; Lee, Y. J.; and Ryoo, M. S. 2018. Learning to anonymize faces for privacy preserving action detection. In *Proceedings of the european conference on computer vision (ECCV)*, 620–636.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241. Springer.

Saggese, A.; Strisciuglio, N.; Vento, M.; and Petkov, N. 2019. Learning skeleton representations for human action recognition. *Pattern Recognition Letters*, 118: 23–31.

Shahroudy, A.; Liu, J.; Ng, T.-T.; and Wang, G. 2016. Ntu rgb+ d: A large scale dataset for 3d human activity analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1010–1019.

Shi, L.; Zhang, Y.; Cheng, J.; and Lu, H. 2019. Two-stream adaptive graph convolutional networks for skeleton-based action recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12026–12035.

Si, C.; Chen, W.; Wang, W.; Wang, L.; and Tan, T. 2019. An attention enhanced graph convolutional lstm network for skeleton-based action recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1227–1236.

Sinha, A.; Chakravarty, K.; and Bhowmick, B. 2013. Person identification using skeleton information from kinect. In *Proc. Intl. Conf. on Advances in Computer-Human Interactions*, 101–108.

Sümer, Ö.; Gerjets, P.; Trautwein, U.; and Kasneci, E. 2020. Automated anonymisation of visual and audio data in classroom studies. *arXiv preprint arXiv:2001.05080*.

Tanaka, N.; Kera, H.; and Kawamoto, K. 2022. Adversarial Bone Length Attack on Action Recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 2335–2343.

Tran, D.; Wang, H.; Torresani, L.; Ray, J.; LeCun, Y.; and Paluri, M. 2018. A closer look at spatiotemporal convolutions for action recognition. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 6450–6459.

Uittenbogaard, R.; Sebastian, C.; Vijverberg, J.; Boom, B.; Gavrila, D. M.; et al. 2019. Privacy protection in street-view panoramas using depth and multi-view imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10581–10590.

Wang, H.; He, F.; Peng, Z.; Shao, T.; Yang, Y.-L.; Zhou, K.; and Hogg, D. 2021. Understanding the robustness of skeleton-based action recognition under adversarial attack. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14656–14665.

Wang, H.; and Wang, L. 2018. Learning content and style: Joint action recognition and person identification from human skeletons. *Pattern Recognition*, 81: 23–35.

Yan, S.; Xiong, Y.; and Lin, D. 2018. Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Thirty-second AAAI conference on artificial intelligence*.

Yang, K.; Yau, J.; Fei-Fei, L.; Deng, J.; and Russakovsky, O. 2021. A Study of Face Obfuscation in ImageNet. *arXiv preprint arXiv:2103.06191*.

Zheng, T.; Liu, S.; Chen, C.; Yuan, J.; Li, B.; and Ren, K. 2020. Towards understanding the adversarial vulnerability of skeleton-based action recognition. *arXiv preprint arXiv:2005.07151*.