Preventing Sensitive Information Leakage from Mobile Sensor Signals via IntegrativeTransformation

Zhang, Dalin; Yao, Lina; Chen, Kaixuan; Yang, Zheng; Gao, Xin; Liu, Yunhao

Published in:
IEEE Transactions on Mobile Computing

DOI (link to publication from Publisher):
10.1109/TMC.2021.3078086

Publication date:
2022

Document Version
Accepted author manuscript, peer reviewed version

Citation for published version (APA):
Zhang, D., Yao, L., Chen, K., Yang, Z., Gao, X., & Liu, Y. (2022). Preventing Sensitive Information Leakage from Mobile Sensor Signals via IntegrativeTransformation. *IEEE Transactions on Mobile Computing*, *21*(12), 4517-4528. Article 9424974. Advance online publication. https://doi.org/10.1109/TMC.2021.3078086

Preventing Sensitive Information Leakage from Mobile Sensor Signals via Integrative Transformation

Dalin Zhang *Member, IEEE,* Lina Yao *Member, IEEE,* Kaixuan Chen *Member, IEEE,*
Zheng Yang *Senior Member, IEEE,* Xin Gao *Member, IEEE* and Yunhao Liu *Fellow, IEEE*

Abstract—Ubiquitous mobile sensors on human activity recognition pose the threat of leaking personal information that is implicitly contained within the time-series sensor signals and can be extracted by attackers. Existing protective methods only support specific sensitive attributes and require massive relevant sensitive ground truth for training, which is unfavourable to users. To fill this gap, we propose a novel data transformation framework for prohibiting the leakage of sensitive information from sensor data. The proposed framework transforms raw sensor data into a new format, where the sensitive information is hidden and the desired information (e.g., human activities) is retained. Training can be conducted without using any personal information as ground truth. Meanwhile, multiple attributes of sensitive information (e.g., age, gender) can be collectively hidden through a one-time transformation. The experimental results on two multimodal sensor-based human activity datasets manifest the feasibility of the presented framework in hiding users' sensitive information (inference MAE increases $\sim$ 2 times and inference accuracy degrades $\sim$ 50%) without degrading the usability of the data for activity recognition (only $\sim$2% accuracy degradation).

Index Terms—mobile sensors, human activity recognition, sensitive information protection, neural network

1 Introduction

Human activity recognition (HAR) plays an important role in various attractive human-in-the-loop applications especially in the smart living scenario [1], [2]. A typical case is the smart healthcare, where wearable sensors are employed to capture the motions of users and a healthcare provider can use the data to support exercise and medical suggestions or emergency rescues. Our lives become safer and more convenient with the assist of these personalized and ubiquitous services. Nevertheless, the concern of privacy leakage of this kind of personal data has drawn an increasing attention in recent years [3]. Although the sensors are originally used to capture the movement of users, personal traits can also be held within the continuous signals unintentionally.

Consider a scenario where an elderly person lives alone that constant monitoring his/her health situation is required. Smart wearable devices or smartphones with multiple built-in sensors (e.g., accelerometer, gyroscope, and magnetometer) are usually used for the monitoring purpose. The collected data is continuously transmitted to the healthcare institute for analysis. An essential task of such analysis is to recognize the activities of the user, such as walking, jogging, and running. However, people perform activities in different ways because of the divergence of personal facts like age, gender, and weight. For example, a person with high body weight would walk more slowly than a person with lower body weight. Therefore, the person's weight information could be inferred through interpreting the sensor signals. Other personal information like age, gender, and height can also be figured out in the similar way [3]–[5]. This kind of information leakage is unacceptable so that the sensor data cannot be directly sent out without any privacy destruction procedure. On the contrary, the inference about human activities such as walking, jogging, and running is the purpose of collecting the sensor data and extremely critical for the downstream applications especially when for the healthcare treatment. As a result, the sensor data should not be modified to avoid degrading the activity recognition performance. In this context, we draw a conundrum, where the sensor data should not be released unchangeably to avoid privacy leakage whereas modification should also be avoided for keeping the recognition precision.

Regarding this contradiction, we report a data transformation approach that manages to separate the two sorts of information contained within the same sensor signals and then hide the user sensitive information and maintain the activity information at the same time. Figure 1 depicts the general data transformation framework. Ideally, sensitive information is held while the desired activity information can pass through during the transformation process. However, as all information (i.e., sensitive and desired) is enclosed in the same raw signals, it is impractical to hide sensitive information yet to keep all the other information unchanged at the same time. In light of this fact, we make an assumption that the specific task, human activity recognition, is the only desired task that needs to be kept unaffected in this study. Nevertheless, our reported framework is adaptable to any desired tasks or a multi-

D. Zhang, and K. Chen (corresponding author) are with the Department of Computer Science, Aalborg University, 9220, Aalborg Øst, Denmark. Email: dalinz@cs.aau.dk, kchen@cs.aau.dk
L. Yao is with the School of Computer Science and Engineering, the University of New South Wales, Sydney NSW 2052, Australia. Email: lina.yao@unsw.edu.au
Z. Yang and Y. Liu are with the School of Software and TNLIST, Tsinghua University, Beijing, 100084, P. R. China. Email: yangzheng@tsinghua.edu.cn, yunhao@tsinghua.edu.cn
X. Gao is with the Computational Bioscience Research Center, Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, 23955-6900, Kingdom of Saudi Arabia. Email: xin.gao@kaust.edu.sa

Manuscript received xx xx, 2020.

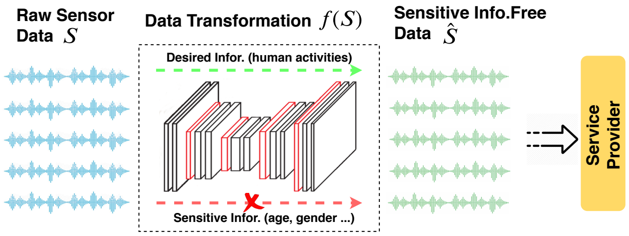

Raw Sensor Data S
Data Transformation $f(S)$
Sensitive Info.Free Data $\hat{S}$
Desired Infor. (human activities)
Sensitive Infor. (age, gender ...)
Service Provider

Figure 1. Data transformation framework for preventing sensitive information leakage. The transformation module modifies the raw sensor data to hide user sensitive information but to keep the desired information unaffected, and then sends the transformed data to a service provider.

task scenario, where more than one desired information needs to be untouched.

There are some factors that make the task of hiding sensitive information challenging. First, as mentioned above, the sensitive information and desired information are both from the same chunk of data that it is painful to modify one in isolation from the other [6]. Moreover, the desired information is the basis to the subsequent downstream applications, such as users' daily activities (desired information from mobile sensor signals) are the basis to their health monitoring (an application). Thus, the impact on the desired information should be minimized. Second, the inherent noises encompassed in the sensor signals deteriorate the data quality severely [7]. In addition, different sensing modes embed information in different ways that developing a unified sensitive information protection approach for all modes is full of challenges, especially for traditional feature engineering. Existing solutions can be divided into two main streams: (1) ceasing sensors based on pre-defined conditions or user specifications, and (2) modifying raw data to hide sensitive information before being sent out to a service provider. The first solution is non-intelligent that are not only damaging to the downstream applications yet not able to fully prevent information leakage at all time [8], [9]. In contrast, the latter solution is a more practical and promising approach that engages boosting attentions [3], [10], [11].

As illustrated in Figure 1, the main idea of the data transformation solution is to seek a data conversion algorithm that makes a better trade-off between hiding the sensitive information and retaining the desired information. Previous works have two principle drawbacks in this context. First, most current approaches can hide only one specific sensitive information (e.g., such as gender [11] or user ID [3]) at one transformation process. This kind of methods would fail in most practical scenarios where it is required to hide multiple sorts of sensitive information. Second, user sensitive information is usually a necessity to train the data transformation network [3], [10], [11]. This necessity would leak the sensitive information to a second party directly, and thus introduce a new leaking risk, which is more severe than only revealing raw data.

Targeting the above defects, this paper presents a unified data transformation framework that can hide multiple kinds of sensitive information at a one-time transformation and does not require users to provide any private information for building the framework. The main idea is originated from the image style transformation research [12] where an image contains two aspects of information: style and content. The style of an image tells *how* it is viewed and the content of an image represents *what* it shows. Inspired by this observation, we argue that a human activity also has such two kinds of abstract information: style (*how*) and content (*what*). The "style" of an activity tells *how* the activity is performed. For example, a walking activity can be in different styles, like limping, wobbly walking, and slow stride. The "content", on the other hand, tells *what* activity a person is performing. In this context, a user's private information, like gender or weight, influences the activity style. As an illustration, a person with higher body weight commonly moves more slowly than a person with low body weight. Besides, the desired information typically requires what a user is doing so can be accounted for "content" information. In light of this intuition, we propose to transform raw sensor data into a new representation that does not have a specific "style" (sensitive information) like random noise and has a "content" (desired information) as raw data. The transformed data should satisfy such conditions: when an adversary tries to infer user sensitive information from the transformed data, the results should be as unreliable as drawn from random noise; whereas a service provider can make inferences about desired information from the transformed data with as high accuracy as from the raw data. Concretely, we design a fully convolutional $TransNet$ that is responsible to carry out the data transformation process and an auxiliary $LossNet$ for defining the training targets of the $TransNet$. The $LossNet$ determines a *style loss* and a *content loss* that try to minimize the differences between the transformed data and random noise as well as raw data, respectively. During the training process, none of user sensitive information is required and only sensor data is presented. Experiments are conducted on two multimodal activity recognition datasets to hide five types of sensitive information (i.e., age, gender, height, weight, and ID). The empirical validation results manifest that the reported approach can successfully hide multiple sensitive information simultaneously at a one-time transformation while supporting a high preservation level of desired information with regard to activity recognition accuracy. This paper extends the preliminary report [13] from four aspects. First, we present new experiments to further study the impact of each proposed loss function and to visualize the transformation process; we also elaborate experimental results and more implementation details in the *Experiment and Results* section; in addition, we use MAE instead of MSE to better illustrate the worst-case scenario performance. Second, a new *Discussion* section is added to explore more details on research significance and limitations. Moreover, we extend the *Introduction* section and add a new section *Related Work* to present more details on research background and intuitions. Lastly, we elaborate the *Methodology* section to give more details on the training process. The implementation code is made publicly available online [1].

2 Related Work

2.1 Activity Recognition from Mobile Sensor Signals

The mobile sensor-based HAR aims to recognize human activities from multimodal time-series signals that are originated from mobile sensors. In this context, machine learning technologies have shown dominant performance since it is hard to uncover the latent traits of sensor signals and their complex correlations by computing methods. Popular methods like Support Vector Machine (SVM) [14] have demonstrated effective in dealing with subject-dependent scenarios. However, they get into trouble when

[1] https://github.com/dalinzhang/SensePrivacy

facing the subject-independent applications [15]. Recent years, deep neural networks have been proven superior to traditional machine learning approaches in various fields including the HAR. They are first applied to mobile sensor-based human activity recognition by [16]. Restricted Boltzmann Machines (RBM) is leveraged for automatic feature extraction and demonstrated significant performance improvement over classical feature engineering. In addition to unsupervised feature learning with deep neural networks, supervised learning with deep neural networks were much more widely explored by researchers. Such as in [17], [18], convolutional neural networks were adopted to fuse different sensing modalities and thus enhanced activity recognition. The advanced attention-based mechanisms were introduced by [19], [20] to achieve remarkable recognition accuracy on different evaluation datasets. The attention-based methods could also provide explicable features of neural networks for HAR [19]. Chen et al. proposed a multi-agent attention model to extract attentive features from different angles and achieved promising performance on several public datasets [21].

2.2 Hide Sensitive Information from Mobile Sensor Signals

The main application of human activity recognition is to monitor human behaviors in a smart environment [22], [23]. Thus, the wearable sensors need to capture the physiological signals of users continuously. Since the way an activity is taken varies among users (due to age, gender, and weight) [24], an adversary could infer user this sensitive information through the time series signals [3]. Although deep neural networks demonstrate powerful abilities in mobile sensor data analysis, the privacy concern limits their further applications.

One naive way of hiding user sensitive information is to stop sensor working based on predefined settings or user specifications [8], [9], [25], [26]. For example, Olejnik et al. [26] proposed an automatic runtime control based on a smartphone usage context. However, this approach dramatically influences the application usage, especially for health monitoring and elderly care. Adding random noise to sensor data is an alternative way to hiding sensitive information [27]–[29]. This kind of methods works well for hiding sensitive information. However, it destroys the usability of data severely because the noise usually applies undifferentiated perturbations to all information, including both desired and sensitive ones [27]. A more advanced approach is to transform raw sensor signals into a new representation without user sensitive information embedded, but with data usability kept. For example, the authors of [3] proposed an adversarial training strategy that transformed raw sensor data into neural network features to hide user-discriminative information. The authors of [11] reported transforming raw data into a new representation with the same data size to hide the gender information. Our work combines the merits of noise perturbation methods and data transformation methods. It conditionally perturbs the raw sensor signals with the desired information unaffected, but all sensitive information is disturbed to be like random noise.

3 METHODOLOGY

3.1 Problem Statement and Definition

Since there are commonly more than one sensing modality deployed for HAR, we first assume all sensor signals have been processed to be synchronized and to have a same frequency. At time point t, the vector $X(t) = [x_1(t), x_2(t), ..., x_m(t)]$ represents the readings of m sensor components (each component could be an axis of a mobile sensor). Complying with this definition, for a time duration of d in length starting from time point t, we conclude the time series sensor signals $S_d(t) = [X(t); X(t+1); ...; X(t+d-1)]$. For simplicity, we use S_d instead of $S_d(t)$ in the rest of the paper.

The S_d is two-dimensional (2D) raw sensor data with one dimension representing time and the other representing sensor components. In traditional conditions, a service provider uses an pre-defined recognition function $I_a(.)$ to infer a user's activities Y_a from S_d to provide dedicated services. Ideally, $I_a(S_d) = Y_a$. On the other hand, there also exists a certain sensitive inference function $I_s(.)$ that can be used to infer the user's sensitive information Y_s (such as gender and age) from S_d. Under an ideal condition, $I_s(S_d) = Y_s$. Our goal is to find an optimal transformation function $f^*(.)$ so that the sensitive information derived from the optimal transformed data $\hat{S}^*_d = f^*(S_d)$ is like drawn from a chunk of random data: $I_s(\hat{S}^*_d) = I_s(Z_d | Z_d = (z_{ij})_{m\times d}; z_{ij} \sim U)$, whereas the desired information about human activities can be the same as drawn from raw data: $I_a(\hat{S}^*_d) = I_a(S_d) = Y_a$. Here, Z_d is a 2D matrix with the same size of S_d, and its element z_{ij} is a random variable drawn from a uniform distribution U; $\hat{S}_d$ is the transformed data achieved through the transformation function $f(.)$ with the raw data S_d as input and $\hat{S}^*_d$ is the optimal transformed data from where the sensitive information cannot be inferred.

3.2 Overview

Figure 2 illustrates the overall architecture of our proposed framework. In order to achieve the optimal transformation function $f^*(.)$, we design the framework to comprise a *TransNet* $f(.)$ that is in charge of the data transformation process, and an auxiliary *LossNet* ϕ that defines the loss functions for preparing the TransNet. Specifically, the LossNet defines: a "style" loss that measures the "style" difference between the $transformed\ data$ $f(S_d)$ and $random\ noise\ N$, a "content" loss that measures the "content" difference between $transformed\ data\ f(S_d)$ and $raw\ data\ S_d$, and a usability loss that specifically helps to keep the inference accuracy of the desired information.

Each loss function computes a scalar value $\ell_i(\hat{S}_d, O_i)$ measuring the difference between the transformed data $\hat{S}_d$ and a transformation target O_i (e.g., random noise or raw data). The TransNet is trained with the stochastic gradient descent to minimize the weighted combination of all loss functions:

$$\operatorname{argmin} \mathbf{E}\big[\sum_{i=1} \lambda_i \ell_i(f(S_d), O_i)\big], \sum_i \lambda_i = 1, \tag{1}$$

where λ_i is the weight of each loss function, which we set experimentally in this research. It mainly controls the tradeoff between privacy and data usability (see section 4.3.2 for detailed experimental results).

3.3 Network Structure

3.3.1 LossNet

The LossNet ϕ is a traditional 2D convolutional neural network for human activity recognition. It is first trained from scratch on raw training sensor data and then fixed for the subsequent training process of the TransNet. The detailed configuration of the LossNet

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3078086, IEEE Transactions on Mobile Computing

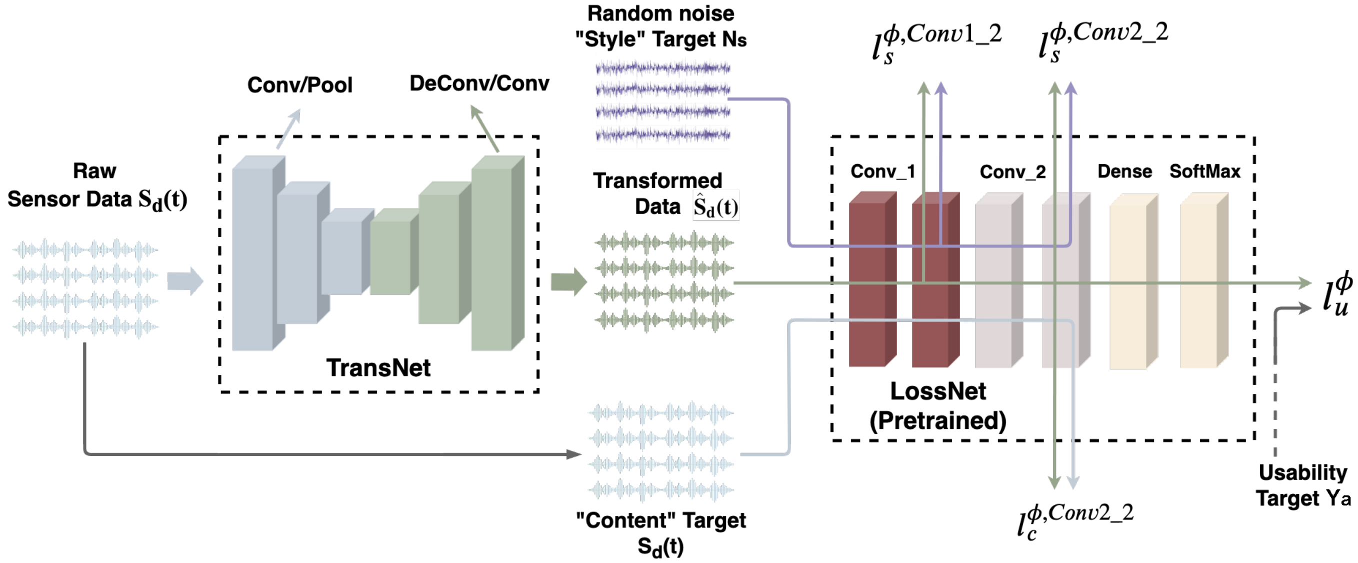


Figure 2. Framework overview. We first pretrain the LossNet on raw sensor data for inferring desired information. Then the LossNet is fixed and used to define the loss functions that measure "style" difference between transformed data and random noise, and "content" difference between transformed data and raw data. We also define a usability loss to keep the inference accuracy of the desired information explicitly. The TransNet is trained through minimizing a weighted combination of the above loss functions to hide sensitive information while simultaneously preserving the desired information.

Table 1
The configurations of the LossNet and TransNet. H, W, F, sH, and sW refer to height, width, the number of feature maps, stride height, and stride width respectively.

Layer	Input Size (H×W×F)	Kernel/Stride (H×W/sH×sW)	Padding	Activation
LossNet				
Conv1_1	m×50×1	1×3/1×1	Same	Relu
Conv1_2	m×50×16	1×3/1×1	Same	Relu
MaxPool1	m×50×16	1×2/1×2	Valid	-
Conv2_1	m×25×32	1×3/1×1	Same	Relu
Conv2_2	m×25×32	1×3/1×1	Same	Relu
MaxPool2	m×25×32	1×2/1×2	Valid	-
Dense	flat(m×12×32)	-	-	Relu
Dense	400	-	-	softmax
TransNet				
Conv1	m×50×1	1×3/1×1	Same	Relu
MaxPool1	m×50×16	1×2/1×2	Valid	-
Conv2	m×25×16	1×3/1×1	Same	Relu
MaxPool1	m×25×32	1×2/1×2	Same	-
Conv3	m×13×32	1×3/1×1	Same	Relu
DeConv1	m×13×32	1×3/1×2	Same	Relu
Conv4	m×26×32	1×3/1×1	Same	Relu
DeConv1	m×26×32	1×3/1×2	Same	Relu
Conv4	m×52×32	1×3/1×1	Valid	-

is depicted in Table 1. The LossNet has two convolutional blocks, each of which has two convolutional layers and a maxpooling layer. The input into the LossNet has a size of $m \times d \times 1$, where m is the number of sensor components, d is time period length, and 1 is the number of feature maps. The period length of d is set to 50 in this study. The convolutional kernel is always set to 1×3, and the maxpooling is always applied along the time dimension to reduce the feature map size by half. After flattening, the output of the second pooling layer is fed into a dense layer of size 400. As last, a dense layer with the softmax activation function defined as softmax(x_i)=$\frac{1}{\mathcal{Z}}$exp(x_i) with $\mathcal{Z}$=$\sum_i$exp(x_i), is appended for the final output. The loss function for training the LossNet is a cross-entropy loss for human activity classification:

$$\ell_a^{\phi} = -\sum_e Y_{a,e} log(\phi(S_d)_e), \tag{2}$$

where $Y_{a,e}$ and $\phi(S_d)_e$ is the label and the predicted probability of the activity category e, respectively. The predicted probability $\phi(S_d)_e$ is output from the LossNet with the raw data as input.

3.3.2 TransNet

The TransNet $f(.)$ is a fully convolutional neural network with downsampling first and upsampling to the original size afterward. Specifically, thesampling frequency and the number of sensor modalities should be the same as the raw data. The fully convolutional TransNet can take any size of data as input, which is another advantage of our framework. The reason we resize the transformed data into the same size as the raw data lies in two folds. (1) It would add extra uncertainty when an adversary was trying to infer sensitive information from the revealed data. The adversary cannot tell whether the revealed data is raw or transformed by simply observation so an extra step would be required to make a judgement before any manipulations to recover raw data. This judgement step would accumulate extra uncertainty to the final results of sensitive inference. (2) When the transformed data and raw data had the same size, the detailed fluctuation of every sensor modality would be kept. This information is critical to specific data analysis scenarios, such as false recognition analysis or abnormal activity analysis. Thus, keeping consistent data sizes in both the modality and time series dimension is an important manner to keep as abundant desired information as possible.

The detailed configuration of the TransNet is also illustrated in Table 1. We use two convolution/maxpooling pairs to downsample the input data followed by two deconvolution/convolution blocks to upsample to the original size. Rather than depending on an interpolating upsampling, deconvolution allows the upsampling process to be learned jointly with the rest of the network. Although

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3078086, IEEE Transactions on Mobile Computing

the input and output have the same size, there are several benefits to the networks that first downsample and then upsample. The first benefit is the computational cost, which is positively correlated with the size of feature maps [30]. The second benefit comes from the receptive field sizes. With downsampling by a factor of D, each additional convolutional layer increases the receptive field size by $2D$ without extra computational cost. Otherwise, it only increases the receptive field size by 2 [31]. The input and output of the TransNet both have a size of $m \times 50 \times 1$. To achieve the size unchanged, we tune the padding options of both the first pooling layer and the last convolutional layer to "Valid". The kernel size setting is consistent with the LossNet. Considering that the raw sensor data can be either positive or negative, we do not apply any activation functions to the last convolutional layer of the TransNet so that the transformed data is not restricted to an activation function's bound (e.g., *relu* has a lower bound of 0) and can be any positive or negative values.

3.4 "Style" and "Content" Consistency

3.4.1 Content Consistency

We define a "content" loss function for measuring the "content" consistency between the transformed data and raw data. The "content" information describes what a user does during the data recording period d, which is human activities in this study. As deeper layers help extract better features, we encourage the raw data S_d and the transformed data $\hat{S}_d$ to have similar feature representations as computed by a deeper convolutional layer of the LossNet ϕ. Formally, let $\phi_j(\hat{S}_d)$ and $\phi_j(S_d)$ be the outputs of the jth layer of the network ϕ when the input of ϕ is the transformed data $\hat{S}_d$ and raw data S_d respectively. If j is a convolutional layer, then $\phi_j(\hat{S}_d)$ will be feature maps of shape $C_j \times H_j \times W_j$. The "content" difference of layer j is defined as the Euclidean distance between the feature representations of the transformed data $\hat{S}_d$ and raw data S_d:

$$\ell_c^{\phi,j} = \frac{1}{C_j H_j W_j} ||\phi_j(\hat{S}_d) - \phi_j(S_d)||_2^2. \quad (3)$$

We use the "content" difference of the layer *Conv2_2* of the LossNet to produce the "content" loss:

$$\ell_c^{\phi} = \ell_c^{\phi,Conv2\_2}. \quad (4)$$

Using a "content" loss from the intermediate layer of the LossNet to train the TransNet encourages the transformed data to keep the "content" similar to the raw data but does not force them to match exactly.

3.4.2 Style Consistency

Besides encouraging similar "content" to raw data, we also would like the transformed data to have no specific "styles". In practical, we use random noise N_s as the style transformation target as random noise can be regarded as a special "style" of "no style". The "style" represents the manner a user performs an activity, which is impacted by personal information like age, gender, and weight [32]. These kinds of personal information are sensitive to users and should not be leaked. Previous research has reported that a convolutional neural network that is originally trained for human activity recognition has the possibility of learning features that could be used for accurately estimating the user's sensitive information, without any intentional design [3]. Therefore, we here use the LossNet to generate the "style" loss for training the TransNet to prevent such leakage.

Inspired by the image style transformation process [12], we utilize the Gram matrix to measure the "style" difference. Essentially, matching the Gram matrices of shallow-layers is to minimize the maximum mean discrepancy (MMD) between the raw data and transformed data so the style transfer can be regarded as a special domain adaptation problem [33]. The transformation process aligns the source domain distribution (i.e., raw data) to the target domain distribution (i.e., random noise). Therefore, we assume that the Gram matrix of shallow-layers can capture all domain information (i.e., sensitive information in this work). We first give the definition of the Gram matrix. Let $\phi_j(x)$ be the output of the jth convolutional layer of the LossNet ϕ when the input of ϕ is x. The shape of $\phi_j(x)$ is $C_j \times H_j \times W_j$. Then the Gram matrix G_j^{ϕ} is defined as a matrix of shape $|C_j| \times |C_j|$ with its elements as:

$$G_j^{\phi}(x)_{c,c'} = \frac{1}{C_j H_j W_j} \sum_{h=1}^{H_j} \sum_{w=1}^{W_j} \phi_j(x)_{h,w,c} \phi_j(x)_{h,w,c'}. \quad (5)$$

It captures information about which feature maps tend to activate together. In practice, the Gram matrix can be computed easily via $G_j^{\phi}(x) = \Psi\Psi^T / C_j H_j W_j$, where Ψ can be obtained by reshaping ϕ_j into a 2D matrix of shape $C_j \times H_j W_j$. The "style" difference is the squared Frobenius norm of the difference between the Gram matrices of the transformed data $\hat{S}_d$ and the random noise N_s:

$$\ell_s^{\phi,j}(\hat{S}_d, N_s) = ||G_j^{\phi}(\hat{S}_d) - G_j^{\phi}(N_s)||_F^2. \quad (6)$$

The layer Conv1_2 and Conv2_2 of the LossNet are used to produce the "style" loss, which is the sum of the "style" difference of each layer. Therefore, we have the final "style" loss:

$$\ell_s^{\phi} = \ell_s^{\phi,Conv1\_2}(\hat{S}_d, N_s) + \ell_s^{\phi,Conv2\_2}(\hat{S}_d, N_s). \quad (7)$$

3.4.3 Usability Loss.

We also define a usability loss ℓ_u^{ϕ} to strengthen maintaining specific desired information during the data transformation process. The usability loss is a cross-entropy loss that measures the difference between the prediction from the pretrained LossNet with the transformed data as input and the ground truth of the desired information:

$$\ell_u^{\phi} = -\sum_e Y_{a,e} log(\phi(\hat{S}_d)_e), \quad (8)$$

where $Y_{a,e}$ and $\phi(\hat{S}_d)_k$ are the label and the predicted probability of the activity category k, respectively. The activity predicted probability $\phi(\hat{S}_d)_e$ is output from the pretrained LossNet with the transformed data as input.

The final loss function is the weighted summation of all individual losses ℓ_c, ℓ_s^{ϕ}, and ℓ_u^{ϕ}.

$$\ell^{\phi} = \mathbf{E}\big[\lambda_c^{\phi} \ell_c(\hat{S}_d, S_d) + \lambda_s \ell_s^{\phi}(\hat{S}_d, N_s) + \lambda_u \ell_u^{\phi}(\hat{S}_d, Y_a)\big]. \quad (9)$$

The weight of each loss λ_i is set experimentally and kept adding up to 1.

3.5 Training Process

The LossNet ϕ is first trained from scratch on raw training data for inferring desired information that is human activity in this study and then fixed during the subsequent training process of the TransNet. Note that when training the LossNet, only raw training data and the corresponding labels of human activities are provided; the labels of user sensitive information are not required. After training the LossNet, we start to train the TransNet. The goal of training the TransNet is to let the transformed data have no specific "styles" like random noise and have a "content" of raw data. Thus the transformed data $\hat{S}_d$, raw data S_d and random noise N_s are input into the pretrained LossNet, respectively. Then the final loss calculated by Eq. (9) is obtained, which at last the TransNet is trained to minimize.

Algorithm 1 describes the detailed training process of our proposed data transformation mechanism. It only takes raw training data S_d, and its paired human activity ground truth Y_a as input. The training process comprises training both the LossNet and the TransNet separately. In contrast, only the TransNet is used to take raw test data as input, and outputs transformed test data during the test stage.

Algorithm 1 Training Process of the Data Transformation Framework for Multiple Sensitive Information Protection

```
Input: raw training data $S_d$, human activity labels $Y_a$
Pretrain LossNet $\phi$
    LossNet $\phi \Leftarrow$ random initializing
    Train LossNet $\phi \Leftarrow \ell_a^\phi = -\sum_e Y_{a,e} log(\phi(S_d)_e)$
    Fix weights and biases of LossNet $\phi$
Train TransNet $f(.)$
    TransNet $f(.) \Leftarrow$ random initializing
    Generate transformed training data $\hat{S}_d$ = $f(S_d)$
    Generate random noise $N_s$
    Feed transformed training data $\hat{S}_d$, random noise $N_s$, raw data $S_d$ into pretrained LossNet $\phi$ respectively
    Calculate "content" loss $\ell_c^\phi = ||\phi_j(\hat{S}_d) - \phi_j(S_d)||_2^2$, "style" loss $\ell_s^\phi = \sum_j ||G_j(\hat{S}_d) - G_j(N_s)||_F^2$, and usability loss $\ell_u^\phi = -\sum_k Y_{a,k} log(\phi(\hat{S}_d)_k)$, where $G(.)$ is the Gram matrix
    Calculate the weighted summation of all loss functions $\ell^\phi = \mathbf{E}\big[\lambda_c \ell_c^\phi(\hat{S}_d, S_d) + \lambda_s \ell_s^\phi(\hat{S}_d, N_s) + \lambda_u \ell_u^\phi(\hat{S}_d, Y_a)\big]$.
    Train the TransNet $f(.)$ w.r.t $\ell^\phi$ in a stochastic gradient descent manner
```

4 Experiment and Results

This section first describes the details of the two evaluation datasets and experiment setup. Then we give the overall experimental results to show the capability of the proposed framework to collectively hide multiple kinds of sensitive information through a unified transformation. At last, we further discuss the usability-privacy tradeoff and loss function design concerns of the proposed framework.

4.1 Datasets

To satisfy the evaluation scenario, we select datasets according to the following criteria:

- The dataset should have at least two kinds of user sensitive information available for validating whether the proposed model is able to hide multiple kinds of sensitive information by a unified transformation. This point is solely for the evaluation purpose;
- The dataset should have at least one kind of desired information (e.g., human activity) available for validating whether the desired information is properly unaffected.
- The dataset should have multiple subjects for validating the framework's generalization over subjects. This point is optional but highly preferred.

We select two public inertial sensor-based human activity recognition datasets: MotionSense [11] and MobiAct [34], which meet all the above criteria. Both datasets have five kinds of user sensitive information available: gender (M/F), identification (ID), height (mm), weight (kg), and age (years old). The desired information of both datasets is the activity that a user performs. Therefore, the goal of the presented framework is to prevent the inference of sensitive information, namely gender, ID, height, weight, and age, while to keep the desired information, namely human activities, still being inferred successfully after data transformation. For the train-test split, we follow the convention in [11] to use the trial-independent manner instead of the subject-independent manner. This is the requirement of testing the *ID* information that represents a multi-user classification task; the main goal is to classify different users [35].

4.1.1 *MotionSense Dataset*

The MotionSense dataset is collected from two inertial sensors, accelerometer and gyroscope, which are integrated within an iPhone 6s smartphone and kept in a user's front pocket. Four sorts of time-series signals are obtained from the inertial sensors, namely attitude, rotation rate, user acceleration, and gravity. Each sort of signal has three dimensions: roll, pitch, and yaw of the attitude data and x, y, and z of the others. Thus there are 12 dimensions in the recording of each time point. A total of 24 users (10 females, 14 males) in a range of gender, age, weight, and height participate in the experiments and collect data of four daily activities: downstairs, upstairs, jogging, and walking. We remove the recordings with incomplete data or labels through data inspection and finally achieve 264 trials of 767,660 recordings. Following the trial-independent manner [11], we select 168 long trials of 2 to 3 minutes each for training and the remaining 96 short trials of 0.5 to 1 minutes each for test. After trial segmentation by a 50-length sliding window, we obtain the sample size (12×50). Finally, there are 61,728 samples (∼80%) for training and 14,098 samples (∼20%) for test.

4.1.2 *MobiAct Dataset*

The MobiAct dataset comprises data recorded from the accelerometer, gyroscope, and orientation sensors of a Samsung Galaxy S3 smartphone for fifty-seven subjects performing nine different types of Activities of Daily Living (ADLs). The main characteristic of this dataset is that it attempts to simulate ADLs with the smartphone located with random orientation in a loose pocket chosen by the participants. The orientation sensor is software-based and derives its data from the accelerometer and the geomagnetic field sensor. Different from the MotionSense dataset, there are three kinds of time series signals obtained from the sensors, namely orientation, rotation rate, and acceleration (including gravity). Each sort of data has three axes: roll, pitch, and azimuth of the orientation signals and x, y, and z of the others. Therefore, the recording of each time point has nine dimensions. After data inspection, we select the data of 44 subjects (14 females, 30 males) performing four ADLs, downstairs, upstairs, walking and jogging, without data and labels of user information missing, in the form of 704 trials of 1,121,296 recordings. As there is no duration difference between the trials of the same activity, we randomly select ∼66% trials for each activity for training and the remaining ∼33% trials for test. Similar to the trial segmentation process of the MotionSense dataset, we cut each trial with a 50-length sliding window and obtain the sample size of (9×50). Finally, there are 88,412 samples (∼80%) for training and 22,212 samples (∼20%) for tests.

4.2 Experimental Setup

4.2.1 Evaluation Setup

Following the conventions in previous researches [3], we use the changes of inference performance before and after transformation to validate the proposed framework. Concretely, if the accuracy of inferring human activities decreases marginally (relative decrease of less than 5%) after data transformation, the proposed framework is regarded as successfully retaining desired information. Otherwise, the proposed framework is regarded as failing to preserve the desired information. On the other hand, if the accuracy of inferring user ID decreases considerably (relative decrease of more than 50%) after data transformation, the proposed framework is regarded as successfully hiding gender information. Otherwise, the proposed framework fails to hide user ID information. For continuous information like height, a considerable increase of the inference error is defined as a relative change of more than 100%. Similar evaluation criteria apply to the other sensitive information.

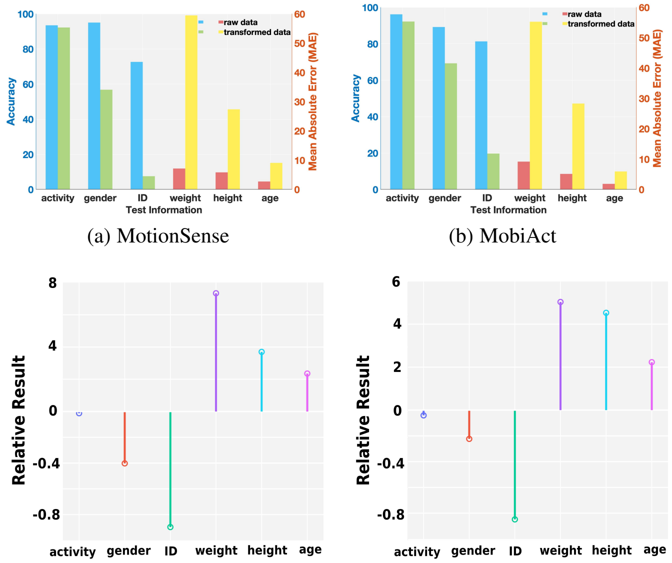
To validate that the sensitive information is hidden and the desired information is retained after data transformation, we build six evaluation neural networks for each dataset for six kinds of information, namely human activity, gender, ID, height, weight, and age. Note that these six neural networks are built only for the evaluation purpose. All evaluation neural networks used the same raw bunch of data and their ground truth for training. For example, the neural network for evaluating whether the human activity information is unaffected after transformation is trained with the raw training data and human activity labels. Similarly, the evaluation network for gender is trained with raw training data and gender labels.

During the test phase, the raw test data first goes through the well-trained TransNet to achieve the transformed test data. Then the output is fed into each evaluation network respectively to get the evaluation results after the transformation. For comparison, the raw test data is also fed into each evaluation network, respectively, to get the evaluation results before the transformation. The evaluation results of both before and after transformation are presented in the following $Evaluation\ Results$ section.

Except for the activity classifier connected to the dense layer, the architecture of all evaluation networks is the same as that of LossNet. The evaluation networks of activity, gender, and ID use the softmax output layer for classification. For the numerical information, height, weight, and age, the linear output layer for regression analysis is used. All evaluation networks are trained using the Adam updating rule [36] with a learning rate of 10^{-3}.

4.2.2 Training Setup

The LossNet is first trained in the trial-independent manner [11] using the raw training data and its paired human activity labels. Afterward, the parameters of the trained LossNet are fixed. Note that the LossNet is only trained with human activity labels since this information is non-sensitive and can be public. To achieve the "style" loss ℓ_s^ϕ for training the TransNet, we generate the random noise N_s by random sampling from a uniform distribution between range [-20, 20] to have the same size of the raw data (i.e., 12×50 of the MotionSense dataset and 9×50 of the MobiAct dataset). The random range is set based on the reasonable scope of sensor readings. The LossNet and TransNet are trained in order using the Adam updating rule [36] with a 10^{-3} learning rate. The weight of each loss function λ_i is experimentally set as $\lambda_s = 0.55$, $\lambda_c = 0.35$, and $\lambda_u = 0.1$.



(a) MotionSense (b) MobiAct

(c) MotionSense (d) MobiAct

Figure 3. Overall evaluation results on both evaluation datasets. The upper part of this figure displays the absolute results; the bottom part presents the relative results. The activity, gender, and ID are evaluated by accuracy; the height, weight, and age are evaluated by MAE.

4.3 Evaluation Results

4.3.1 Overall Performance

Figure 3 plots the evaluation results of the proposed framework on two evaluation datasets. The numerical details are summarised in Table 2. The upper part of Figure 3 displays the absolute results of inferring both desired and sensitive information. We use the classification accuracy as the evaluation criteria for inferring categorical information, namely activity, gender, and ID, and plot the results with a normal scale to the left axis of Figure 3a and 3b. For continuous information (i.e., height, weight, and age), we introduce the mean absolute error (MAE) as the evaluation criterion. Compared with MSE, MAE is more robust to small errors so it is more suitable to show the worst-case scenario performance. We plot the results to the right axis of Figure 3a and 3b. The bottom part of Figure 3 exhibits the relative results of the transformed data compared to the raw data. The relative result is defined as:

$$\Delta_r = \frac{R(After) - R(Before)}{R(Before)}, \tag{10}$$

where $R(After)$ and $R(Before)$ are the absolute results of transformed data and raw data. As the sensitive information has

Table 2
Evaluation results of the proposed framework with optimal model parameter settings on two evaluation datasets.

Measurement Criterion	Desired Information	Sensitive Information	MotionSense			MobiAct		
			Before	After	Δ_r	Before	After	Δ_r
Accuracy (%)	Activity	-	93.49	92.25	-0.0133	96.17	92.56	-0.0375
	-	Gender	95.05	56.79	-0.4025	89.20	69.26	-0.2235
	-	ID	72.64	7.512	-0.8966	81.30	19.70	-0.8480
Mean Absolute Error (MAE)	-	Weight	7.133	59.48	+7.339	9.160	55.26	+5.033
	-	Height	5.823	27.37	+3.700	5.120	28.30	+4.527
	-	Age	2.699	9.060	+2.356	1.832	5.920	+2.231

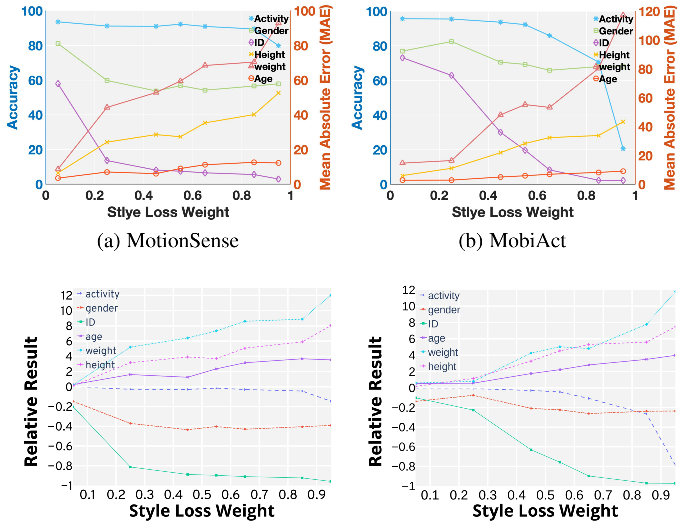
real-world reasonable ranges, the absolute results give the intuitive sense about the extent that the proposed framework perturbs the sensitive information. On the contrary, the relative results depict the change extent after data transformation compared to raw data.

Our framework obtains satisfactory performance of hiding sensitive information on both datasets with a marginal decrease of HAR accuracy, but significant error increases in inferring all test sensitive information. Specifically, after data transformation, the HAR accuracy can still maintain above 90% with only less than a 4% drop. The HAR accuracy of the MobiAct dataset has a relatively larger drop than the MotionSense dataset ($\sim$ 4% vs. $\sim$ 1%). It is noticeable that this performance can be optimized by tuning the loss weights λ_i (more in the *Privacy-Usability Tradeoff* section), and the settings of the loss weights for evaluating both datasets are identical so the reported results are not optimal separately. This demonstrates that our framework is robust to the settings of loss weights across different datasets. In contrast, when inferring user gender, the accuracy declines dramatically nearly to the random guess level. Note that due to the gender imbalance of the evaluation datasets, the random guess level of gender inference is 58% and 68% for the MotionSense and MobiAct dataset, respectively. There is also a considerable drop of ID inference accuracy after data transformation with the relative result decrease of around 85% for both datasets. Thus the sensitive information of user gender and ID has been changed to have a random "style" and hard to be precisely inferred after data transformation. The inference errors of numerical sensitive information, height, weight, and age, also rise remarkably after data transformation. In particular, the inference of weight experiences the most significant performance degradation with relative MAE increases 7.339 times and 5.033 times after transformation for the MotionSense and MobiAct datasets respectively. Even the smallest performance degradation of inferring user age still suffers MAE increase more than two times. Considering the user age has a relatively small reasonable range, the increase of the inference error is significant. The overall results exhibit that our framework is able to transform raw mobile sensor signals into a new representation that does not have a specific "style" (sensitive information) like random noise, yet the "content" (desired information) same with raw data.

4.3.2 Privacy-Usability Tradeoff

The loss weight λ_i controls the tradeoff between privacy protection and the usability of transformed data. In this section, we perform experiments by varying the weight of the style loss λ_s^ϕ from 0.05 to 0.95 to investigate the privacy-usability tradeoff of the proposed framework. To make the summation of all loss weights as a constant, the weights of content loss and usability loss have to be changed as well. We equally change the weights of content loss and usability loss. For example, if the weight of style loss decreased 0.2, the weights of content loss and usability loss would increase 0.1, respectively. Figure 4 shows the results of both evaluation datasets. The upper part shows the absolute value results, and the bottom part displays the relative results.

It is evident that with the weight of the style loss growing, the inference error of all test sensitive information increases; thus, the risk of sensitive leakage decreases. However, the accuracy of human activity recognition does not change too much until the weight of style loss higher than 0.85. Especially for the MotionSense dataset, the activity recognition accuracy remains about 90% at the style loss weight of 0.85. At the lower side, the MAE and inference accuracy of the sensitive information change obviously at 0.25 and fluctuate smoothly afterward. There is a sharp change after the style loss larger than 0.85. Similarly, the activity recognition accuracy drops clearly at the tail part, which indicates that the usability of data decreases significantly at a large style loss weight. The MobiAct dataset has a similar privacy-usability tradeoff trend as the MotionSense dataset with the efficiency of hiding sensitive information goes up apparently after the style weight of 0.45. However, the activity recognition accuracy drops a noticeable amount of about 15% with the style weight from 0.65 to 0.85. Finally, the activity inference accuracy sharply downs to only 20% at the end, where the transformed data is useless. Thus, it is crucial to carefully select the weight of the style loss to keep a satisfactory privacy-usability tradeoff.



(a) MotionSense (b) MobiAct

(c) MotionSense (d) MobiAct

Figure 4. Privacy-usability tradeoff with different weights of the style loss on both evaluation datasets. The upper part of this figure presents the results of absolute values; the bottom part displays the details of relative results.

4.3.3 Visualization

To gain an intuitional sense of the transformation process, we compare the spectrograms of acceleration signals before and after transformation. Figure 5 shows the visualization results. We can observe that the transformation process introduces new periodic components that cover the original ones and differ across activities. Since the periodic information of the motion sensor data encodes abundant signatures of users, the perturbation on such components reduces the possibility of sensitive information leakage. The downstairs and upstairs visualization shows similar perturbation patterns because people act slightly differently when going upstairs and downstairs concerning the acceleration of the X-axis. In contrast, the visualization patterns of walking and jogging are significantly different, and both different from that of upstairs and downstairs.

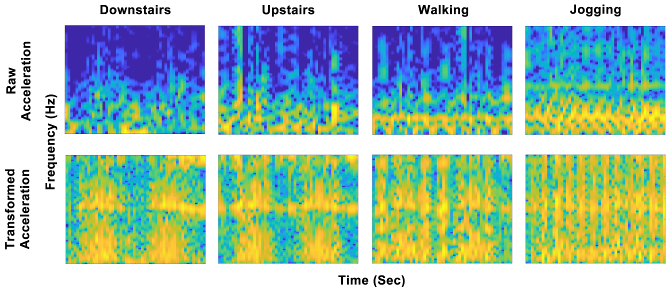


Figure 5. Spectrum visualization of raw (top) and transformed (bottom) X axis user acceleration data of different activities from one user of the MotionSense dataset.

4.3.4 Impact of Loss Functions

In this section, we discuss the impact of different loss functions ℓ_i^ϕ on the performance of the presented framework. A reconstruction loss defined as $\ell_r^f = \frac{1}{md}||S_d - \hat{S}_d||_2^2$ that encourages the transformed data to be the same with the raw data is also introduced for a comparison study as this loss function is commonly used in previous studies for sensitivity-hiding data transformation [11].

Figure 6 shows the sensitive information hiding performance of the proposed framework trained with different loss functions on the MotionSense dataset. When training with the usability loss ℓ_u^ϕ only, the transformed data can keep approximately the same usability characteristics as the raw data. However, even though the transformation process has modified the raw data that the transformed data and the raw data have different distributions, the sensitive information is still retained unintentionally after data transformation. This result is in consonance with that reported in [3]. When looking into the results achieved with the content loss ℓ_c^ϕ only, we find it promotes more information preservation than the usability loss ℓ_u^ϕ since ℓ_c^ϕ encourages the extracted features of the transformed data to be the same with the raw data rather than only the final recognition accuracy. Besides, the content loss can be used when the ground truth of the desired information is unavailable, and the LossNet is a general pretrained neural network that is not specifically designed for inferring the desired information. In contrast, as indicated in Figure 6, the style loss ℓ_s^ϕ advocates all test information damaged to a massive extent that the activity recognition accuracy drops to a random guess level.

Although the reconstruction loss helps to retain more raw information even including some hidden ones than the other loss functions, it is such a tight constraint that degrades the sensitive protection. On the contrary, the content loss is a relatively loose constraint compared to the reconstruction loss, yet also has potential abilities to preserve unknown information. Thus we use the content loss rather than the reconstruction loss, which is often used in previously reported transformation-based sensitive information hiding strategies.

5 Discussion

5.1 Release Privacy-free Data versus Release Recognition Results

This paper aims to solve the privacy leakage issue and proposes to transform raw signals at the user-end to hide private information and then release the privacy-free data. Another simple and intuitive solution to such a privacy leakage problem is to perform the human activity recognition at the user-end directly and release mere activity recognition results instead of transformed signals. However, this solution only works for the simplest scenario while it is not flexible to deal with complicated real-world scenarios. One ordinary scenario that merely releasing recognition results cannot handle is to collect a dataset for algorithm development and validation. In contrast, releasing privacy-free data not only satisfies the requirement properly but also helps to hide participants' private information. Another common scenario that requires signal data instead of mere recognition results is to manually inspect data when an abnormal recognition result occurs. Experts would request to inspect signal data to examine whether an unusual activity is attributed to the failure of a recognition algorithm or the occurrence of an emergency. Moreover, it is also desirable to having the signal data, when an algorithm fails to recognize a critical activity (e.g., falling), for developing and testing new algorithms. Therefore, simply releasing recognition results is severely limited in real-world applications.

5.2 Impact of LossNet

In the proposed framework, the LossNet is trained to use raw sensor data to categorize human activities. It has the explicit ability to extract features of activities and the implicit ability to extract features of user personal information [3]. These two properties are utilized to define the loss function for training the TransNet so that the transformed data can hide sensitive information while keep activity information. The LossNet is not used during the transformation process. Thus, even though the LossNet can extract sensitive information, it will not let the transformed data inherit any sensitive information during the transformation process. Moreover, it is only the loss function that determines whether the TransNet can hide sensitive information or not. The LossNet is merely used to provide style features for defining the loss. We design the loss function to have the style features extracted from transformed data similar to those extracted from random noise. The LossNet has the ability to extract style features but not to preserve any styles in it. When the input is random noise, the LossNet can only extract style features from random noise, that has no specific styles. Therefore, having random noise as the style training target, the transformed data will not contain any specific styles (sensitive information).

5.3 Limitations and Future Directions

Although the experimental results have demonstrated the efficacy of the proposed method, there are some scenarios where the

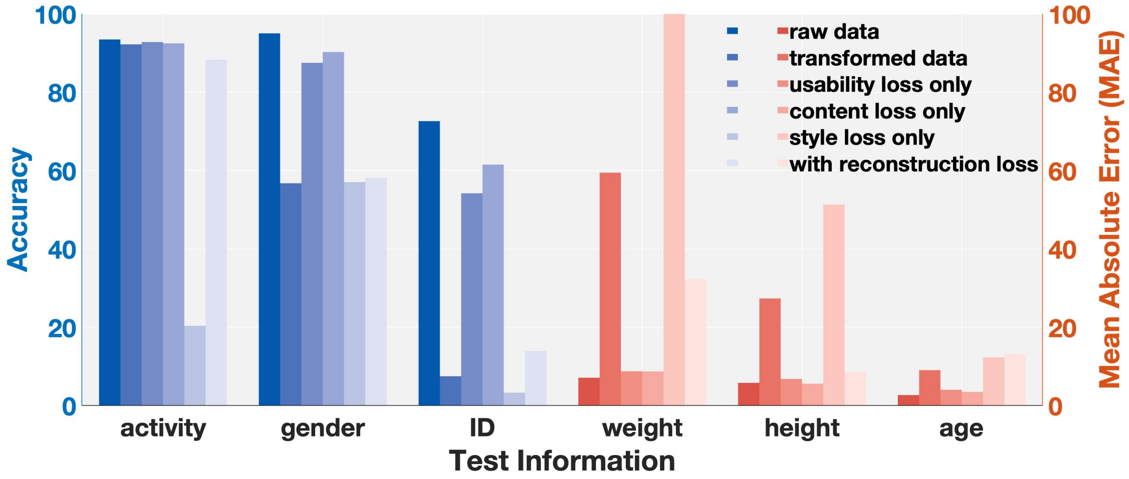


Figure 6. Impact of Different Loss Functions. The activity, gender, and ID are evaluated with classification accuracy; the height, weight, and age are evaluated with MAE.

proposed approach is not able to handle or requires further exploration and development. A limitation of this paper is that the transformation process not only hides users' private information but also unintendedly removes possible unique activity patterns. This is unfavorable for some special applications. For example, it is desirable to retain the unique walking patterns of leg-disordered patients for physicians to analyze the patients' rehabilitation process. A potential solution is to design an auxiliary module to capture such unique patterns and release the unique patterns along with common patterns to service providers. Another limitation of this work is that there are no mathematical supports for the assumption that all sensitive information can be captured by the Gram matrix. This assumption is based on that matching Gram matrices is actually a domain adaptation process that aligns the distribution of raw data to the distribution of random noise. A future direction is to verify this assumption for different sensitive attributes and investigate other matrices that can achieve the similar goal. Theoretical analysis of the sensitive information removal process is also an important future research direction. [37] proposed that the task-irrelevant and dataset-specific information could be minimized by adding a regularization term to the loss function in a supervised training process. Although the reference did not address the domain adaptation problem, extending the theoretical analysis in [37] to the domain adaptation in this work is an interesting and important subject of future investigations. A preliminary idea is that the style transformation loss can be regarded as a regularization term to control the task-irrelevant information. Regarding the framework structure, a potential direction is to take the advantage of GAN-style settings, where LossNet and TransNet are trained alternatively and repeatedly, to update the LossNet and TransNet simultaneously instead of in a sequential setting. In such a way, a more robust transformation function can be achieved.

6 Conclusion

In light of the drawbacks of being effective only on one specific sensitive information and requiring user information for training, this paper targets to resolve the limitations of previous work on hiding user sensitive information from mobile sensing signals. Other than hiding a dedicated sensitive trait, we introduce to detach multiple sensitive information from the signals and convert it to be stochastic at a one-time transformation. Meanwhile, the desired information, which is human activities in this research, is kept as constant. To achieve such a goal, we adopt the idea of style transfer to transform raw sensor data into a new representation that has no specific "styles" (sensitive information) like random noise and has "content" (desired information) same as raw signals. As a result, various user sensitive traits are disturbed simultaneously at a one-time transformation and no user personal information is required for training. We carry out experiments on two multimodal human activity recognition datasets to validate the empirical effectiveness of the reported approach. With regard to hiding five sorts of sensitive information (i.e., gender, ID, height, weight, and age), our proposed mechanism exhibits satisfactory performance on deteriorating their inference precision while holding the activity recognition accuracy with only marginal drop.

7 Acknowledgement

We wish to thank the reviewers and editors for their insightful comments and suggestions.

The research reported in this paper was supported by Innovation Fund Denmark centre, DIREC and King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under award numbers BAS/1/1624-01, REI/1/0018-01-01, URF/1/3412-01-01, and URF/1/4098-01-01.

References

[1] S. Chatterjee, B. Mitra, and S. Chakraborty, "Type2motion: Detecting mobility context from smartphone typing," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking (MobiCom)*. ACM, 2018, pp. 753–755.

[2] X. Xu, J. Yu, Y. Chen, Y. Zhu, and M. Li, "Steertrack: Acoustic-based device-free steering tracking leveraging smartphones," in *2018 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2018, pp. 1–9.

[3] Y. Iwasawa, K. Nakayama, I. E. Yairi, and Y. Matsuo, "Privacy issues regarding the application of dnns to activity-recognition using wearables and its countermeasures by use of adversarial training," in *2017 International Joint Conference on Artificial Intelligence (IJCAI)*, 2017, pp. 1930–1936.

[4] J. Lu, G. Wang, and P. Moulin, "Human identity and gender recognition from gait sequences with arbitrary walking directions," *IEEE Transactions on information Forensics and Security (TIFS)*, vol. 9, no. 1, pp. 51–61, 2013.

[5] A. Jain and V. Kanhangad, "Investigating gender recognition in smartphones using accelerometer and gyroscope sensor readings," in *2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT)*. IEEE, 2016, pp. 597–602.

[6] A. Raij, A. Ghosh, S. Kumar, and M. Srivastava, "Privacy risks emerging from the adoption of innocuous wearable sensors in the mobile environment," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM, 2011, pp. 11–20.

[7] J. Fridolfsson, M. Börjesson, C. Buck, Ö. Ekblom, E. Ekblom Bak, M. Hunsberger, L. Lissner, and D. Arvidsson, "Effects of frequency filtering on intensity and noise in accelerometer-based physical activity measurements." *Sensors*, vol. 19, no. 9, pp. 2186:1–12, 2019.

[8] K. R. Raghavan, S. Chakraborty, M. Srivastava, and H. Teague, "Override: A mobile privacy framework for context-driven perturbation and synthesis of sensor data streams," in *Proceedings of the Third International Workshop on Sensing Applications on Mobile Phones*. ACM, 2012, pp. 1–5.

[9] S. Chakraborty, C. Shen, K. R. Raghavan, Y. Shoukry, M. Millar, and M. Srivastava, "ipshield: a framework for enforcing context-aware privacy," in *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, 2014, pp. 143–156.

[10] S. A. Osia, A. Taheri, A. S. Shamsabadi, K. Katevas, H. Haddadi, and H. R. Rabiee, "Deep private-feature extraction," *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol. 32, no. 1, pp. 54–66, 2018.

[11] M. Malekzadeh, R. G. Clegg, A. Cavallaro, and H. Haddadi, "Protecting sensory data against sensitive inferences," in *Proceedings of the 1st Workshop on Privacy by Design in Distributed Systems*, 2018, pp. 1–6.

[12] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*. Springer, 2016, pp. 694–711.

[13] D. Zhang, L. Yao, K. Chen, G. Long, and S. Wang, "Collective protection: Preventing sensitive inferences via integrative transformation," in *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2019, pp. 1498–1503.

[14] Z. He and L. Jin, "Activity recognition from acceleration data based on discrete consine transform and svm," in *2009 IEEE International Conference on Systems, Man and Cybernetics*. IEEE, 2009, pp. 5041–5044.

[15] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep learning for sensor-based human activity recognition: overview, challenges and opportunities," *ACM Computing Surveys (CSUR)*, 2021.

[16] T. Plötz, N. Y. Hammerla, and P. L. Olivier, "Feature learning for activity recognition in ubiquitous computing," in *2011 International Joint Conference on Artificial Intelligence (IJCAI)*, 2011, pp. 1729–1734.

[17] S. Münzner, P. Schmidt, A. Reiss, M. Hanselmann, R. Stiefelhagen, and R. Dürichen, "Cnn-based sensor fusion techniques for multimodal human activity recognition," in *ACM International Symposium on Wearable Computers (ISWC)*. ACM, 2017, pp. 158–165.

[18] M. Z. Uddin, M. Hassan, and Mehedin, "Activity recognition for cognitive assistance using body sensors data and deep convolutional neural network," *IEEE Sensors Journal*, pp. 8413–8419, 2018.

[19] K. Chen, L. Yao, X. Wang, D. Zhang, T. Gu, Z. Yu, and Z. Yang, "Interpretable parallel recurrent neural networks with convolutional attentions for multi-modality activity modeling," in *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2018, pp. 1–8.

[20] L. Wang, J. Zang, Q. Zhang, Z. Niu, G. Hua, and N. Zheng, "Action recognition by an attention-aware temporal weighted convolutional neural network," *Sensors*, vol. 18, no. 7, pp. 1979: 1–18, 2018.

[21] K. Chen, L. Yao, D. Zhang, B. Guo, and Z. Yu, "Multi-agent attentional activity recognition," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, 2019, pp. 1344–1350.

[22] I. M. Pires., N. M. Garcia., N. Pombo., and F. Flórez-Revuelta., "Limitations of the use of mobile devices and smart environments for the monitoring of ageing people," in *Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health - HSP,*, INSTICC. SciTePress, 2018, pp. 269–275.

[23] A. Benmansour, A. Bouchachia, and M. Feham, "Multioccupant activity recognition in pervasive smart home environments," *ACM Computing Surveys (CSUR)*, vol. 48, no. 3, pp. 1–36, 2015.

[24] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE communications surveys & tutorials*, vol. 15, no. 3, pp. 1192–1209, 2012.

[25] H. Fu, Z. Zheng, S. Zhu, and P. Mohapatra, "Keeping context in mind: Automating mobile app access control with user interface inspection," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*. IEEE, 2019, pp. 2089–2097.

[26] K. Olejnik, I. Dacosta, J. S. Machado, K. Huguenin, M. E. Khan, and J.-P. Hubaux, "Smarper: Context-aware and automatic runtime-permissions for mobile devices," in *2017 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2017, pp. 1058–1076.

[27] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, "Random-data perturbation techniques and privacy-preserving data mining," *Knowledge and Information Systems (KBS)*, vol. 7, no. 4, pp. 387–414, 2005.

[28] H. Wang and Z. Xu, "Cts-dp: Publishing correlated time-series data via differential privacy," *Knowledge-Based Systems*, vol. 122, pp. 167–179, 2017.

[29] Y.-S. Moon, H.-S. Kim, S.-P. Kim, and E. Bertino, "Publishing time-series data under preservation of privacy and distance orders," in *International Conference on Database and Expert Systems Applications (DEXA)*. Springer, 2010, pp. 17–31.

[30] J. Cong and B. Xiao, "Minimizing computation in convolutional neural networks," in *International Conference on Artificial Neural Networks (ICANN)*. Springer, 2014, pp. 281–290.

[31] W. Luo, Y. Li, R. Urtasun, and R. Zemel, "Understanding the effective receptive field in deep convolutional neural networks," in *Advances in neural information processing systems (NIPS)*, 2016, pp. 4898–4906.

[32] T. Brezmes, J.-L. Gorricho, and J. Cotrina, "Activity recognition from accelerometer data on a mobile phone," in *International Work-Conference on Artificial Neural Networks (IWANN)*. Springer, 2009, pp. 796–799.

[33] Y. Li, N. Wang, J. Liu, and X. Hou, "Demystifying neural style transfer," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, 2017, pp. 2230–2236.

[34] G. Vavoulas., C. Chatzaki., T. Malliotakis., M. Pediaditis., and M. Tsiknakis., "The mobiact dataset: Recognition of activities of daily living using smartphones," in *Proceedings of the International Conference on Information and Communication Technologies for Ageing Well and e-Health - ICT4AWE, (ICT4AGEINGWELL 2016)*, INSTICC. SciTePress, 2016, pp. 143–151.

[35] X. Yu, Z. Zhou, M. Xu, X. You, and X.-Y. Li, "Thumbup: Identification and authentication by smartwatch using simple hand gestures," in *2020 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2020, pp. 1–10.

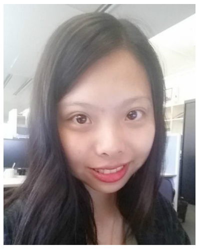
[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations (ICLR)*, 2015, pp. 1–15.

[37] A. Achille and S. Soatto, "Emergence of invariance and disentanglement in deep representations," *The Journal of Machine Learning Research*, vol. 19, no. 1, pp. 1947–1980, 2018.

Dalin Zhang (Member, IEEE) is currently an Assistant Professor at the Department of Computer Science, Aalborg University, Denmark. He is also a faculty member in the Center for Data-Intensive Systems (Daisy). Before joining Aalborg University in 2020, he was pursuing his Ph.D. degree at the School of Computer Science and Engineering, the University of New South Wales Sydney (UNSW Sydney), Australia, between 2017-2020. He was at Spreadtrum Communications, Inc, China, working as a Digital Integrated Circuit Design Engineer from 2015 to 2017. He received his Master degree from the University of Chinese Academy of Sciences in 2015, and Bachelor Degree from Jilin University in 2012, both majoring in Microelectronics. His research interest includes brain-computer interface (BCI), human activity recognition and Internet of Things (IoT).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3078086, IEEE Transactions on Mobile Computing

Lina Yao (Member, IEEE) is currently an Associate Professor at School of Computer Science and Engineering, the University of New South Wales (UNSW Sydney), Australia. She received her Ph.D. degree and Master degree both from the University of Adelaide (UoA) in 2014 and 2010, respectively, and her Bachelor degree from Shandong University (SDU). She was ARC Discovery Early Career Researcher Award (DECRA) Fellow between 2016-2018 (awarded in 2015). Before she joined UNSW in 2016, she was a lecturer and an ARC research associate at UoA. Her research interest lies in Data Mining and Machine Learning applications with the focuses on Internet of Things Analytics, recommender systems, human activity recognition and Brain Computer Interface. She is a member of the IEEE and the ACM.



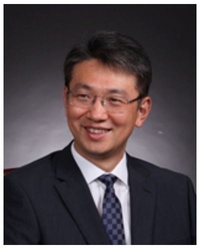
Kaixuan Chen (Member, IEEE) is an Assistant Professor at the Department of Computer Science, Aalborg University, Denmark. She is also a faculty member in the Center for Data-Intensive Systems (Daisy). Before joining Aalborg University, she received her Ph.D. degree from the School of Computer Science and Engineering, University of New South Wales (UNSW Sydney) in 2020. She earned her Bachelor degree from Xi'an Jiaotong University (XJTU) in 2015, majoring in Telecommunication Engineering. Currently, her scientific research interest is in Data Mining, Deep Learning and Internet of Things (IoT).

Zheng Yang (Senior Member, IEEE) received the BE degree in computer science from Tsinghua University, in 2006 and the Ph.D. degree in computer science from the Hong Kong University of Science and Technology, in 2010. He is currently an associate professor at the Institute of Trustworthy Network and System, School of Software, Tsinghua University. His main research interests include wireless ad-hoc/sensor networks, and mobile computing. He is a senior member of the IEEE.

Xin Gao (Member, IEEE) is a professor of computer science at King Abdullah University of Science and Technology (KAUST), Saudi Arabia. He is also the Associate Director of the Computational Bioscience Research Center (CBRC), Deputy Director of the Smart Health Initiative (SHI), and the Lead of the Structural and Functional Bioinformatics (SFB) Group at KAUST. Prior to joining KAUST, he was a Lane Fellow at Lane Center for Computational Biology in School of Computer Science at Carnegie Mellon University. He earned his bachelor degree in Computer Science in 2004 from Tsinghua University and his Ph.D. degree in Computer Science in 2009 from University of Waterloo.

Dr. Gao's research interest lies at the intersection between machine learning and biology. In the field of computer science, he is interested in developing theories and methodologies related to deep learning, probabilistic graphical models, kernel methods and matrix factorization. In the field of bioinformatics, his group works on building computational models, developing machine learning techniques, and designing efficient and effective algorithms to tackle key open problems along the path from biological sequence analysis, to 3D structure determination, to function annotation, to understanding and controlling molecular behaviors in complex biological networks, and, recently, to biomedicine and healthcare.

Dr. Gao has published more than 240 papers in the fields of bioinformatics and machine learning. He is the associate editor of Genomics, Proteomics & Bioinformatics, BMC Bioinformatics, Journal of Bioinformatics and Computational Biology, and Quantitative Biology, and the guest editor-in-chief of IEEE/ACM Transactions on Computational Biology and Bioinformatics, Methods, and Frontiers in Molecular Bioscience.



Yunhao Liu (Fellow, IEEE) received the B.S. degree from the Automation Department, Tsinghua University, Beijing, China, the M.A. degree from Beijing Foreign Studies University, Beijing, and the M.S. and Ph.D. degrees in computer science and engineering from Michigan State University (MSU), East Lansing, MI, USA. He is currently a Chang Jiang Chair Professorship with Tsinghua University. His current research interests include wireless sensor network, pervasive computing, peer-to-peer computing, and Internet of Things.

Dr. Liu currently serves as the Editor-in-Chief for ACM Transactions on Sensor Networks. He was an Associate Editor of the IEEE/ACM Transactions on Networking from 2012 to 2016, and an Associate Editor-in-Chief of the IEEE Transactions on Parallel and Distribution Systems from 2011 to 2015. He is a fellow of the IEEE and the ACM.