EasyChair Preprint

№ 8247

Privacy Preserving Multi-Class Fall Classification Based on Cascaded Learning and Noisy Labels Handling

Leiyu Xie, Yang Sun, Jonathon Chambers and Mohsen Naqvi

June 10, 2022

Privacy Preserving Multi-class Fall Classification Based on Cascaded Learning And Noisy Labels Handling

Leiyu Xie
Intelligent Sensing and Communications Research Group, Newcastle University
Newcastle upon Tyne, UK
l.xie6@newcastle.ac.uk

Yang Sun
Big Data Institute University of Oxford
Oxford, UK
yang.sun@bdi.ox.ac.uk

Jonathon A. Chambers
School of Engineering University of Leicester
Leicester, UK
jonathon.chambers@leicester.ac.uk

Syed Mohsen Naqvi
Intelligent Sensing and Communications Research Group, Newcastle University
Newcastle upon Tyne, UK
mohsen.naqvi@newcastle.ac.uk

***Abstract*—With an increasingly ageing population in the world, fall detection and classification for elderly people becomes an imperative problem that needs to be addressed for assisted living. Currently, most of the fall detection algorithms are based on wearable and non-wearable sensors, such as accelerometer and video camera respectively. In this work, different from previous vision-based methods where the whole images are used, to mitigate the privacy protection problem and detect different types of fall events, we utilize only the skeleton data to achieve the classification of different fall events by using a deep neural network (DNN). The cost of manually labelling and due to varieties of annotators, for a recorded dataset, there always exist errors which will deteriorate the performance. To address this issue, we introduce the confident learning to remove wrong labelled samples and propose a new cascaded learning method to solve the noisy labelled data problem. To confirm the efficacy of the proposed method, we compare different algorithms on the UP-Fall dataset to show that the proposed method performs better than the state-of-the-art.**

***Index Terms*—Fall detection, skeleton features, noisy label, confident learning, cascaded learning**

I. Introduction

At present, falls have surpassed cardiovascular diseases and cancer and become the primary reason for death and health effects in the elderly [1]. Most of the previous works on fall detection focus on binary fall detection i.e. fall event or no-fall event. However, there are different types of fall events in the real world, such as falling forward, falling sideways, falling backwards, which will lead to different types of injuries in the human body parts. Therefore, in this paper, we will focus on the classification of five types of fall events.

Fall detection methods are mainly divided into three types: wearable sensors based, ambient sensors based and vision sensors based [2]–[4]. In fall detection, compared with the wearable sensors based methods, the accuracy of the vision-based methods are slightly lower but there is no requirement for the elderly people to wear the sensors, which they often forget to wear. In vision-based fall detection, one of the main challenges is privacy protection [5], [6]. In order to protect the personal privacy, a skeleton feature extraction model is used to obtain the human skeleton features from the images. By using only the skeleton data, almost all the personal information can be eliminated and the privacy is preserved [7], [8]. Meanwhile, the dynamic lighting condition problem could also be avoided to ensure the robustness of algorithm [9], [10]. Since the size of the skeleton data is much smaller than the whole image data, the requirement of the computational cost is also reduced. In recent years, deep neural network (DNN) is a commonly-used method to address classification problems with promising results [11]. The performance of the DNN is greatly affected by the data quality and labelling errors which occur due to both cognitive errors and model bias errors [12]. It is also crucial to find a solution to address training with noisy labels. Confident learning is an effective method to solve this issue [13]. It is based on the principle of a joint distribution probability density function and focus on label quality by characterizing and identifying noisy labels.

In this paper, we propose a novel cascaded learning DNN to train with noisy labels for fall events classification. The contributions of our work are: (1). A new DNN architecture has been proposed for the fall events classification task. (2). Confident learning is verified to be used to address the noisy label problem on skeleton data. (3). Based on confident learning and the proposed DNN model, finally, we proposed a cascaded learning method to improve the performance of the overall multi-class fall events classification.

II. Related work

In terms of fall detection, many datasets were collected during the last decade. We will give only two examples: the UP-Fall dataset (latest) [14] and the UR-Fall (widely-used) [15]. The UR-Fall dataset contains 70 video sequences, 30 of them are fall sequences while the rest are for daily living. The UP-Fall dataset is a larger dataset in fall detection compared with the UR-Fall dataset, which contains 5 fall events and 6 daily activities and 561 video sequences.

With the UR-Fall dataset, transfer learning and the convolutional neural network (CNN) are applied to achieve the fall detection and earn competitive performance compared with the wearable sensor-based method [16]. Recently, a fall detection method is proposed that uses saliency maps to train a two-stream CNN at the image-level [17]. Then, with the UP-Fall dataset, the method in [14] evaluates the dataset with four different classifiers: support vector machine (SVM); K-nearest Neighbour (KNN); Random Forest (RF); and Multi-Layer Perceptron (MLP). Recently, [18] used human skeleton data from the UP-Fall dataset to address the fall detection problem and obtain a promising performance in both fall detection and activity recognition.

However, the data labelling problem is not addressed, and the automatic annotations always have errors to degrade the performance. In general, to address noisy label issues, [19] proposed an algorithm called Co-teaching to reduce the influence of noisy labels when addressing the image classification problems. Then, [20] added weights into the loss function which is generated by MentorNet. In [21], a meta-learning algorithm is introduced to overcome the noisy label in the training stage. Recently, confident learning is proposed to calculate the joint distribution probability density function between the true labels and noisy labels for pruning the noisy labels without requiring hyperparameters [13].

III. Proposed Method

A. Confident Learning for Noisy Label Pruning

According to [13], confident learning could identify the label errors in the datasets and improve the classification performance by calculating the joint distribution between noisy label $\widetilde{y}$ and the true label y^*. Assume training dataset $D = (x, \widetilde{y})_n^m$, which denotes n samples with m categories with noisy label $\widetilde{y}$ for samples x. We calculate the possibility threshold as:

$$\tau_f = \frac{\sum_{x \in D_{\widetilde{y}=f}} \widehat{p}_f(x)}{|D_{\widetilde{y}=f}|} \tag{1}$$

where τ_f is the possibility threshold for all samples labelled as $\widetilde{y} = f$. If the predicted probability for $\widetilde{y} = d$ has $\widehat{p}_f(x) \geq \tau_f$, then it will be suspected as a wrong annotation.

$$C_{\widetilde{y}_d, y_f^*} := \left| \widehat{D}_{\widetilde{y}=d, y^*=f} \right| \tag{2}$$

According to equation (2), the confusion matrix could be obtained by counting $C_{\widetilde{y}_d, y_f^*}$ in x, and the predicted label is d but true label is f.

$$\widehat{J}_{\widetilde{y}=d, y^*=f} = \frac{\frac{C_{\widetilde{y}=d, y^*=f}}{\sum_{f \in [m]} C_{\widetilde{y}=d, y^*=f}} \cdot |D_{\widetilde{y}=d}|}{\sum_{d \in [m], f \in [m]} \left(\frac{C_{\widetilde{y}=d, y^*=f}}{\sum_{f \in [m]} C_{\widetilde{y}=d, y^*=f}} \cdot |D_{\widetilde{y}=d}| \right)} \tag{3}$$

where $(\cdot)$ is the multiplication operator. After the joint distribution $\widehat{J}_{\widetilde{y}=d, y^*=f}$ between noisy labels $\widetilde{y}$ and true labels y^* is obtained. We could use the following methods to identify the suspicious labels:

Confusion: The noisy labels are selected by using off-diagonal elements of the confusion matrix.

PBC: For all the categories of the samples in m, $d \in [m]$, $n \cdot \sum_{f \in [m]: f \neq d} (\widehat{J}_{\widetilde{y}=d, y^*=f^{[d]}})$ samples for filtering with lowest confidence will be identified as noisy labels.

PBNR: $n \cdot \widehat{J}_{\widetilde{y}=d, y^*=f}$ samples in off-diagonal will be selected $x \in X_{\widetilde{y}=d}$ with max margin $\widehat{p}_{x, \widetilde{y}=f} - \widehat{p}_{x, \widetilde{y}=d}$.

C+NR: Indicates the prune and operator to the sample if both PBC and PBNR are true.

B. Proposed DNN

Since the size of skeleton data obtained from the video image is much less than the original video image data, the skeleton data may have trade-off between privacy protection and desired information. To overcome the possible information loss, we propose a new DNN model with inner concatenation between different layers to reuse the information. The final output of the model is a weighted ensemble from sub-outputs.

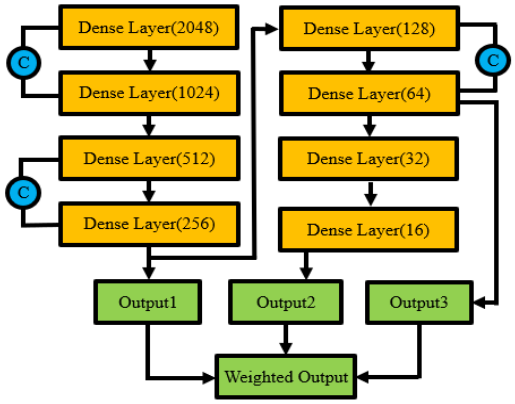

Fig. 1. The proposed DNN architecture for privacy protection and mitigating the information loss.

The proposed DNN architecture is shown in Fig. 1. It has 8 dense layers and 3 concatenation operations between the layers in order to reuse the skeleton information. Meanwhile, 3 sub-outputs from different layers are extracted to generate the final weighted output, we call it as inner-ensemble. It is believed that the sub-outputs from different layers will have different sensitivity and precision for different activities. In each dense layer, batch normalization and Relu function are also applied.

The loss function used in the proposed DNN model:

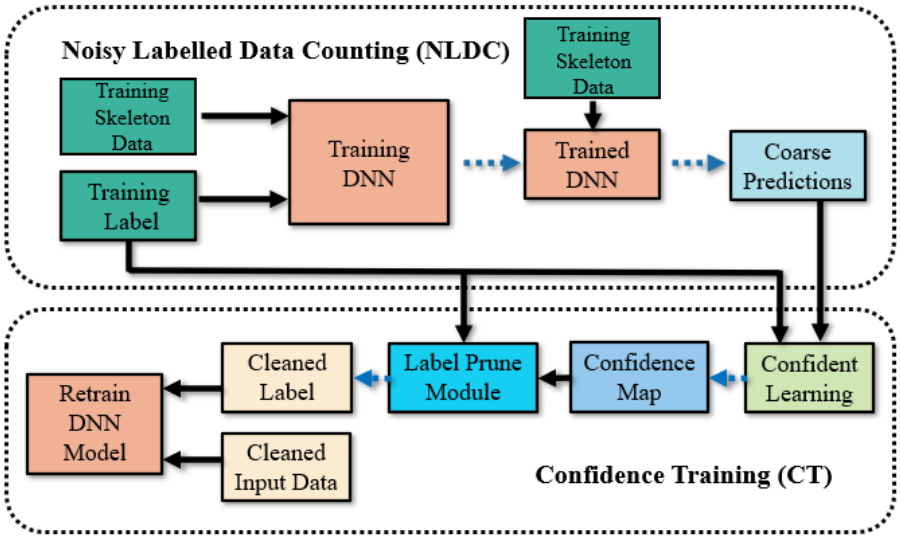

Fig. 2. The framework of the proposed cascaded learning method. In the noisy labelled data counting (NLDC) stage, the DNN model will be trained with noisy labelled data to perform the coarse predictions, which will be used as the input of the confidence training (CT) stage, and clean the data by using different pruning techniques. The cleaned data will be used for retraining the proposed DNN model shown in Fig. 1.

$$loss = -\sum_{i=1}^{3} \omega_i (y_i \cdot log(\widehat{y}_i)) \tag{4}$$

where ω_i represents the loss weight of the sub-output i. And $y_i, \widehat{y}_i$ indicates the target and prediction for the sample in the ith sub-output, respectively. The final loss is the weighted combination of three sub-losses, therefore the number of sub-loss is 3, and the weight ratio of the sub-output is 1:1:2.

C. Cascaded Learning

Based on the DNN model described above, we propose a cascaded learning based method as shown in Fig. 2. The cascaded learning pipeline contains two stages, noisy labelled data counting (NLDC) and confidence training (CT). In the NLDC stage, firstly, the DNN model is trained with uncleaned training data, i.e., labels which contain labelling errors. Then, secondly, by using the trained DNN, the coarse predictions corresponding to the uncleaned training data are provided as an input to the CT stage.

In the CT stage, by using the coarse predictions and the noisy labels from the NLDC stage, a confidence map is obtained which indicate the suspicious noisy labels. Then, according to the confidence map, in the label prune module, the training samples with suspicious noisy labels are removed to generate a new cleaned training dataset. Finally, we re-train the proposed DNN model as shown in Fig. 1. In order to keep up with the referenced work, we conducted the experiments and obtained the performance on the original data rather than the cleaned data. The performance is provided in the following section.

IV. Experiments

A. Dataset and Preprocessing

In the UP-Fall dataset, there are 5 fall events and 6 normal daily activities in both 2 perspective cameras. The CAM1, is named as a side way camera in the proposed work. The work in [18] is the baseline, they used RF, SVM, KNN, MLP to perform the fall detection. To avoid any confusion, we make the dataset into 12 classes with one more class called unknown activity. In the proposed work, we focus on the five fall events classification and the performance will be shown in this section. By using Alphapose [22], the human skeleton will be obtained which contains 17 joint points. Each feature point contains 3 dimensions which are joints scores and 2-D coordinates. Therefore, 51 attributes are used as features for each image for the model to classify the five fall events.

Fig. 3. An example of forward falling using hands. (a) is the original RGB image, (b) is the skeleton data extracted from AlphaPose [22].

In the preprocessing step, all the blank images without a subject in the frames are removed. The ratio of the number of falls and no-fall in the data set is approximately 3:97. After the preprocessing step, in total there are 220,660 groups of skeleton data, 154,462 in training and 66,198 testing sets, respectively. Besides, we set parts of the training set as our validation set in order to prevent the over fitting issue. Fig. 3 shows the skeleton data example which is extracted by using AlphaPose. The experiments are conducted on a work station

with 4 GeForce GTX 1080Ti GPUs, and 16GB of RAM. And the framework is implemented based on Keras.

TABLE I
ABBREVIATIONS OF FIVE FALL EVENTS IN THE PROPOSED METHOD

HF	Hands forward Falling
KF	Knees forward Falling
BF	Backward Falling
SF	Sideways Falling
SDF	Sit Down Falling

Table 1 shows the details of the five activities in UP-Fall dataset. It is highlighted that different from the other fall events, the subjects are facing to the camera when they are recording the sideways fall. In the other 4 fall events, subjects are side-way in the camera field of view.

B. Results and Discussion

To evaluate the overall classification accuracy, F1-score is selected as the performance measure. Fig. 4 shows the performance of the four classifiers in the baseline work and the proposed DNN with noisy labelled data.

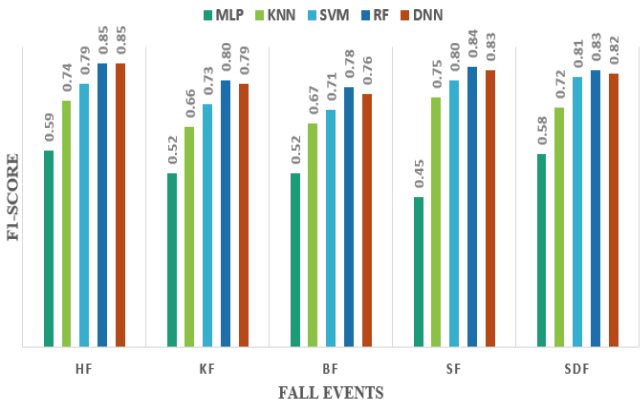

Fig. 4. Shows the classification performance evaluated by using F1-score with noisy labelled data. The X-axis shows the fall events abbreviated in Table 1.

It can be seen in Fig. 4 that among the four classification methods, RF has the highest performance but the DNN also achieves the competitive performance. According to Table 2, although MLP has the lowest inference time, it gives the lowest classification performance. Meanwhile, the proposed CD-DNN has the second shorter inference time and achieves the competitive performance compared with the best classification performance of all falls events.

TABLE II
INFERENCE TIME OF OUR PROPOSED METHOD AND OTHER METHODS BY USING CLEANED DATA.

Methods	MLP	*CD-DNN*	RF	KNN	SVM
Time (s)	0.34	*1.23*	2.13	146.87	157.64

The RF earns the first place in all events classification, however, from Table 2, the inference time of the proposed CD-DNN (1.23 seconds) is much shorter than RF (2.13 seconds), which has 73.17 % improvement in the inference time. Therefore, the RF- and DNN-based fall detection algorithm have trade-off between the classification performance and its inference time. Since fall detection always requires low inference time, the proposed CD-DNN will be the better choice.

Fig. 5. Examples of different types of noisy labels.

We provided some visualization results in Fig. 5, to show the examples of the noisy labels found by confident learning in the dataset. According to the observation, in terms of fall events, in Fig. 5 (a), the true label is falling by using hands, but the given label is falling by using knees. While in Fig. 5 (b), the true label is falling by using knees but the given label is falling by using hands. Moreover, in daily activities, Fig. 5 (c) shows the case where the given label is laying but the true label is side-way falling.

TABLE III
CLASSIFICATION PERFORMANCE USING F1-SCORE WITH PROPOSED CASCADED LEARNING.

Methods	HF	KF	BF	SF	SDF
RF [18]	0.85	0.80	0.78	0.84	0.83
RF-PBC	0.83	0.81	0.78	0.83	0.83
RF-C+NR	0.84	0.82	0.80	0.86	0.86
RF-PBNR	0.86	0.83	0.79	0.87	0.84
RF-Confusion	**0.88**	**0.85**	**0.82**	**0.87**	**0.87**
DNN	0.85	0.79	0.76	0.83	0.82
CD-DNN-PBC	0.83	0.82	0.79	0.84	0.84
CD-DNN-C+NR	0.86	0.80	0.79	0.84	0.83
CD-DNN-PBNR	0.82	0.81	0.81	0.87	0.84
CD-DNN-Confusion	**0.85**	**0.83**	**0.83**	**0.88**	**0.88**

In Table 3, we give the comparison results between the baseline method shown in [18] and the proposed methods. It can be observed that after using the proposed label pruning, almost all the fall events classification performance have improved. These results confirm the importance of confident learning in label pruning methods when using skeleton data. Meanwhile, within label pruning using confident learning, the *Confusion* method can achieve the best performance. After comparing the results between *RF-Confusion* and *DNN-Confusion* in Table 3, *RF-Confusion* earns the best performance in 2 events (HF and KF) while *DNN-Confusion* gives the best performance in 3 events (BF, SF and SDF).

The results confirm that different from previous work, where confident learning is applied at image-level, the noisy label issue with skeleton data can also be addressed. Moreover, after introducing confident learning, some labelling errors of fall events could be removed, which is beneficial to classification models being trained under correct supervision. According

to the classification performance and computational cost, the proposed CD-DNN will be the best choice for fall events classification based on the privacy preserving skeleton dataset.

V. Conclusion

In this paper, we proposed a novel cascaded DNN architecture to achieve the fall events classification by using skeleton data. Meanwhile, in order to address the noisy label issue in the training, we introduced the confident learning to build a cascaded learning method, which helped to improve the reliability. From all the experimental results, it can be confirmed that with skeleton data, the overall performance of fall detection was significantly improved by using the proposed method. Compared with RF, the proposed method can achieve competitive performance with less computational cost which could better meet the requirements of fall detection. The proposed CD-DNN models can be modified as an advanced architectures for further performance improvement.

References

[1] R. Igual, C. Medrano, and I. Plaza, "Challenges, issues and trends in fall detection systems," *Biomedical engineering online*, vol. 12, no. 1, pp. 1–24, 2013.

[2] M. Mubashir, L. Shao, and L. Seed, "A survey on fall detection: Principles and approaches," *Neurocomputing*, vol. 100, pp. 144–152, 2013.

[3] M. Yu, A. Rhuma, S. M. Naqvi, L. Wang, and J. A. Chambers, "A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 6, pp. 1274–1286, 2012.

[4] M. Yu, Y. Yu, A. Rhuma, S. M. Naqvi, L. Wang, and J. A. Chambers, "An online one class support vector machine-based person-specific fall detection system for monitoring an elderly individual in a room environment," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 6, pp. 1002–1014, 2013.

[5] T. Xu, Y. Zhou, and J. Zhu, "New advances and challenges of fall detection systems: A survey," *Applied Sciences*, vol. 8, no. 3, pp. 418–428, 2018.

[6] F. Angelini, Z. Fu, Y. Long, L. Shao, and S. M. Naqvi, "2d pose-based real-time human action recognition with occlusion-handling," *IEEE Transactions on Multimedia*, vol. 22, no. 6, pp. 1433–1446, 2019.

[7] C.-B. Lin, Z. Dong, W.-K. Kuan, and Y.-F. Huang, "A framework for fall detection based on OpenPose skeleton and LSTM/GRU models," *Applied Sciences*, vol. 11, no. 1, pp. 329–348, 2021.

[8] F. Angelini, J. Yan, and S. M. Naqvi, "Privacy-preserving online human behaviour anomaly detection based on body movements and objects positions," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019.

[9] W. Chen, Z. Jiang, H. Guo, and X. Ni, "Fall detection based on key points of human-skeleton using OpenPose," *Symmetry*, vol. 12, no. 5, pp. 744–760, 2020.

[10] F. Angelini and S. M. Naqvi, "Joint RGB-pose based human action recognition for anomaly detection applications," in *IEEE International Conference on Information Fusion (FUSION)*, 2019.

[11] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, M. Hasan, C. Brian E. Van, A. A. Awwal, and V. K. Asari, "A state-of-the-art survey on deep learning theory and architectures," *Electronics*, vol. 8, no. 3, pp. 292–357, 2019.

[12] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, "Efficient processing of deep neural networks: A tutorial and survey," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2295–2329, 2017.

[13] C. Northcutt, L. Jiang, and I. Chuang, "Confident learning: Estimating uncertainty in dataset labels," *Journal of Artificial Intelligence Research*, vol. 70, pp. 1373–1411, 2021.

[14] L. Martínez-Villaseñor, H. Ponce, J. Brieva, E. Moya-Albor, J. Núñez-Martínez, and C.Peñafort-Asturiano, "Up-fall detection dataset: A multimodal approach," *Sensors*, vol. 19, no. 9, pp. 1988–2103, 2019.

[15] B. Kwolek and M. Kepski, "Human fall detection on embedded platform using depth maps and wireless accelerometer," *Computer Methods and Programs in Biomedicine*, vol. 117, no. 3, pp. 489–501, 2014.

[16] A. Núñez-Marcos, G. Azkune, and I. Arganda-Carreras, "Vision-based fall detection with convolutional neural networks," *Wireless Communications and Mobile Computing*, vol. 2017, pp. 1–16, 2017.

[17] H. Li, C. Li, and Y. Ding, "Fall detection based on fused saliency maps," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 1883–1900, 2021.

[18] H. Ramirez, S. A. Velastin, I. Meza, E. Fabregas, D. Makris, and G. Farias, "Fall detection and activity recognition using human skeleton features," *IEEE Access*, vol. 9, pp. 33532–33542, 2021.

[19] B. Han, Q. Yao, X. Yu, G. Niu, M. Xu, W. Hu, I. Tsang, and M. Sugiyama, "Co-teaching: Robust training of deep neural networks with extremely noisy labels," *arXiv preprint arXiv:1804.06872*, 2018.

[20] L. Jiang, Z. Zhou, T. Leung, L.-J. Li, and F.-F. Li, "Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels," in *International Conference on Machine Learning, ICML*, 2018.

[21] M. Ren, W. Zeng, B. Yang, and R. Urtasun, "Learning to reweight examples for robust deep learning," in *International Conference on Machine Learning, ICML*. PMLR, 2018.

[22] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "RMPE: Regional multi-person pose estimation," in *International Conference on Computer Vision, ICCV*, 2017.