Volume 3, Issue , 2024
PP 176-187

International Journal of Futuristic Innovation in Arts, Humanities and Management (IJFIAHM)

DEEP LEARNING HUMAN ACTIVITY RECOGNITION

1st Miss. Kasturi Harbade
kasturiharbade@gmail.com
Student, Department of Computer Science & Engineering,
Shri Sai College of Engineering & Technology, Chandrapur, India.

2nd Mr. Lowlesh Yadav
lowlesh.yadav@gmail.com
Assistant Professor, Department of Computer Science & Engineering,
Shri Sai College of Engineering & Technology, Chandrapur, India.

3rd Mr. Ashish Deharkar
ashish.deharkar@gmail.com
Assistant Professor, Department of Computer Science & Engineering,
Shri Sai College of Engineering & Technology, Chandrapur, India.

ABSTRACT

Human activity recognition is an area of interest in colorful disciplines similar as senior and health care, smart- structures and eavesdrop shaft, with multiple approaches to working the problem directly and efficiently. For numerous times hand- drafted features were manually uprooted from raw data signals, and conditioning were classified using support vector machines and hidden Markov models. To further ameliorate on this system and to prize applicable features in an automated fashion, deep literacy styles have been used. The most common of these styles are Long Short- Term Memory models (LSTM), which can take the successional nature of the data into consideration and outperform being ways, but which have two main risks; longer training times and loss of distant pass memory. A relevantly new type of network, the Temporal Convolutional Network (TCN), overcomes these risks, as it takes significantly lower time to train than LSTMs and also has a lesser capability to capture further of the long-term dependences than LSTMs. When paired with a Convolutional Auto-Encoder (CAE) to remove noise and reduce the complexity of the problem, our results show that both models perform inversely well, the results also show, for assiduity operations, the TCN can directly be used for fall discovery or analogous events within a smart structure terrain.

Keyword: Deep Learning, Human Activity Recognition,Detection.

1. INTRODUCTION:

In recent times, Human activity recognition (HAR) and bracket have gained instigation in both assiduity and academic exploration due to a vast number of operations associated with them. One area in particular where this exploration has huge interest is smart homes and the

Internet of effects (14,19). Other areas include crowd counting, health and senior care, in particular fall discovery, (20), (3). Fall discovery has been a popular area of exploration to enable further independent living for both the senior and impaired within their own accommodation, but also within surroundings where cameras cannot be used due to data protection. There are two main approaches used for HAR invasive Andon invasive. Invasive HAR involves wearing detectors to track humans to produce a rich dataset for models to learn from, whilenon-invasive HAR allows humans to be covered without any attached bias(21). One way to do this is using WIFI signals, which are extensively available in utmost structures.

In HAR, the main conditioning in the bracket task is sitting, standing, walking, lying down, falling down and mortal absence. All of these conditioning is of interest in the area of smart homes, while the falling down exertion is of particular interest in health and senior care, where cameras cannot be installed in private apartments but there's a need to cover cases. Thisnon-invasive, data sensitive system to warn staff to a case falling is of great interest to the assiduity.

2.METHODOLOGY:

Preliminarily, HAR was performed using spatial features and SVMs to classify the point representations, which could be moreover thick or meager spatial points on histograms (4, 8, 19). latterly CNNs started to gain instigation, showing that DL could find applicable features, and hence outperform these styles relatively significantly (7, 9, 23). Ronao and Cho (16) argued that HAR has several real- world operations, in particular Ambient supported Living, due to the rising cost that a geriatric population has on the economy. If further of the senior could be given the occasion to safely sustain themselves at home, also the pressure on the healthcare system would be reduced. They and others have shown how DL can take raw data signal from detectors for illustration CNNs can be used to prize important features from accelerometer raw data (10, 12, 16).

Yousefi etal.(22)pre-processed channel state information CSI by means of principle element analysis tode-noise the signal, and also used a short- time Fourier transfigure to prize features. These features were also used as inputs to arbitrary timber (RF) and hidden Markov model (HMM) classifiers. The results were compared to a DL approach using a LSTM, which didn't bearde-noising or point birth as these are performed within the model. The RF and HMM models achieved64.67 and73.33 independently. The LSTM model scored90.5

delicacy over 17 better than the HMM model and roughly 26 better than the RF. These results show that DL models can eclipse classically styles, though the author noted that LSTMs were important slower to train. This is where our proposed model helps as it's significantly faster to train, as mentioned below.

Zou etal. (24) introduced a model called Auto- Encoder Long- term Recur rent Convolution Network (AE- LRCN), which was a DL approach to HAR. An autoencoder was used for representation literacy and to remove essential noise. The raw CSI time window was converted into an idle space of 256 features, whereas our proposed system used a CAE not only to remove noise, but to compress the CSI time window into an idle space of 12 features. Next Zou et (24) used a convolutional network for point birth, which had 2 convolution layers followed by 2 completely connected layers. The proposed model does not bear this step as the autoencoder had formerly performed aggressive point selection. Eventually, for successional literacy Zou etal. (24) enforced the popular LSTM, which has been shown to perform veritably well on this type of data. We introduce the TCN model, which is new to the area of HAR using CSI data, to learn the successional nature of the data. We show that both the TCN and LSTM achieve state- of- the- art results in HAR, but the new proposed system is much more effective. Wang etal. (18) introduced a DL- grounded channel picky exertion recognition system called CSAR. This system requires considerable pre-processing, starting with channel quality evaluation and selection. They elect the channels with a breadth over a threshold, and neglect the others under the supposition that they're uninformative. Next, they use channel hopping, where CSAR circularly hops through these named channels, combining adja cent channels into an extended channel with advanced bandwidth. Wang etal. (18) denoise the data by using a low- pass sludge with a cut- off frequency of 100Hz, andPCA for data reduction andde-noising. Eventually, the DL model enforced is the LSTM. In our proposed model the CAE denoises the signal, while the TCN — which is significantly faster than the LSTM is used to learn the successional nature of the CSI data. Wang etal. (18) used analogous conditioning to our work, achieving on average over 95 delicacies, which compares well to our results.

Li etal. (11) used CSI data collected from multiple access points to classify four different conditioning. A DL model, conforming of multiple inputs into a convolutional neural network (CNN) combined into a completely connected subcaste also into a classifier, is compared to a SVM approach. The results show that the DL system learns to model the data more, achieving lesser delicacy. Li et (11) transfigure the CSI data into spectrographs which

are also divided into time windows. Bolomanetal. (15) compared a LSTM to one of the utmost com mon univariate models for time series, the bus-Accumulative Integrated Moving Average (ARIMA) model, using WIFI data. LSTM prognostications were significantly more, reducing the root mean square error (RSME) by between80.9 and, showing formerly again that the DL approach is more promising in this area.

Trascauetal. (17) used detector successional data to descry conduct of colorful types. The authors compared their model, which comported of a TCN, to different setups of LSTMs. They noted that the TCN was briskly to train than the LSTMs, and their results show that the TCN was about 5 further accurate. Another intriguing result showed that the TCN was more robust on cross-subject and cross-view than LSTMs, which is also true in our work. Nair etal. (13) and Lea etal. (9) have successfully used TCNs for exertion recognition. Both com pruned their TCN models to LSTMs, showing lesser scores in the evaluation criteria they used. Nair etal. reflected that the training time of the TCNs were in orders of magnitude hastily than that of LSTMs due to their Parallelizable armature. Their work also showed that TCNs were more robust and generalized better, possible due to their longer memory retention. Hou etal. (6) also used a TCN model on two hand gesture recognition datasets. The TCN was compared to CNNs, LSTMs and RNNs and outperformed them all. The results were also compared to approaches similar as handwrought features and histograms of slants, but again showed a far superior capability to automatically prize applicable features. Eventually, Hou etal. stated that the TCN could be trained in an extremely short time, attesting Nair (13).

3.DEEP LEARNING ACTIVITY DETECTION CHAIN:

In this section we describe CAE- TCN (our new HAR fashion), starting with thepre-processing of the CSI data, followed by the noise- junking and complexity reduction using the CAE, and finishing with the TCN model to learn the successional latent- space representation of the conditioning, which will classify the chances as a particular exertion. Traditionally, approaches to this HAR problem were dived using pollutants, similar as the standard or Butterworth pollutants to smoothing data, remove noise by discard ing the first principal element, followed by a point birth phase generally involving some sphere expert knowledge. The final stage comported of a classifier, generally an SVM, to learn to collude the named features onto an n- dimensional space which could also be used to prognosticate unseen samples. We propose a new DL approach that uses a CAE to find an embedding space

for the time windows of the array created using the CSI data. The part of the CAE is twofold (1) find this minimal embedding space, and (2) remove unwanted noise from the signal. The final part of the DL approach feeds this embedding space into a DL successional model. generally, for this stage an LSTM could be used, but we propose a different model the lately-developed TCN. The TCN learns the essential temporal dependences demanded to distinguish each of the conditioning that will be vital for the conclusion stage on new data. We will also show that although the TCN yields veritably analogous results to the LSTM, the TCN is computationally more effective.

3.1. Pre-processing:

The CSI data is entered in complex form, the real number representing the breadth and the imaginary number representing the phase. As we're dealing with only a single transmitter and receiver setup, the phase is of little value and is discarded. An array of the breadth of the 90 subcarriers, 30 within each frequency band, is saved to an array. This array comported of 20 seconds of an exertion, where the CSI data was tried 5 times per second. This 90 × 100 array

Component	Description

was also resolve into 4-alternate windows with a 50 imbrication. Each 20-alternate exertion array was converted into a sequence of nine 90 × 20-time windows, which formed the dataset that the CAE was trained and validated on. The usual 80-10-10 training, confirmation and testing split was used to insure no leakage into the training phases of the CAE or TCN.

Data Collection	Sensors like accelerometers and gyroscopes record motion data.
Data Preprocessing	Clean and prepare data, extract features, and segment the data.
Deep Learning Model	Utilizes neural networks for activity recognition.
Training and Validation	Training the model with labeled data and optimizing parameters.
Inference/Prediction	Deployed model for real-time activity recognition in applications.

3.2. Convolutional bus- Encoder:

The CAE is the first link in the proposed DL chain, and was fed with arbitrary samples, in each batch, of the 4-alternate time window arrays from the training dataset. An input image isof size 90 × 20, representing the 90 subcarriers and time way which have been tried at a rate of 5 per second. The CAE is a dimensional reduction model with an encoder and a decoder. The encoder reduces the size of the input while maintaining enough information to reconstruct the array. It maps the input array into a n- dimensional space where n is the latent embedding at the center of the CAE. The decoder also learns the remapping of this latent vector back into the input.

The CAE is trained by trying to reduce the loss between the original array and the repairedarray, while being forced through the tailbackis squeezed, the more efficiently the TCN will perform, but like all models it has to be guided to find the idle size too important reduction and the model will fail to meet to a reasonable state. For these trials, the encoder had 5 convolutional layers with 128 pollutants in the first 4 layers and 4 pollutants in the 5th embedding) subcaste.

The decoder was a glass image of the first 4 layers. Each of the convolutional layers was followed by batch normalization, a dense remedied direct unit, and powerhouse at a 25, except for the embedding subcaste which had only batch normalization. A kernel size of 3 × 3 was named throughout the network. The array was reduced from 90 × 20 to a size of 3 × 1, and since the bedding subcaste had 4 pollutants the idle subcaste had a vector size of 12.

A CAE is a unique DL model the target affair is the input but, as mentioned over, is squeezed through a narrow subcaste known as the embedding subcaste. This also helps to remove unwanted signal/ noise in the array. Since the labors x 0 are trained to match the inputs x, the loss function used was the root mean squared error (RMSE). To avoid overfitting, an early stopping criterion was used to elect the time in which the model was supposed to be trained.

The optimizer used to train the CAE parameters was Stochastic grade Descent with a reducing literacy rate. The original literacy rate was0.1, and when no drop in the loss within a tolerance value of 20 ages passed, it was reduced by a factor of 10. The minimal value of the literacy rate was 10 −5 and at this value, formerly the tolerance value was reached without enhancement, training stopped. It should also be noted that the network was allowed to explore the loss space for 100 ages before any drop in the literacy rate started. This gave a redundant insurance that the network didn't get stuck in an original minimum at the launch of the training process. The sequences of nine 4 alternate 90 × 20 window arrays for an exertion has been counterplotted onto a sequence of nine vectors of length 12, which was also be used to train and test the TCN.

3.3. Temporal Convolution Network:

The dereliction choice when dealing with sequence problems were LSTMs because of their important capability to find patterns in temporal dependences in sequential data.

A recent advancement from the CNN, known for image recognition achievements, is the TCN which uses a 1D completely- connected structure, in which each retired subcaste is the same length as the input subcaste. To ensure that layers have the same size, zero- padding of length sludge size minus boneis added. Next, unproductive complications (defined over) are used to insure no information leakage from future to history.

Dilated complications are used to reduce the model’s look reverse history complexity which is size direct in network depth and sludge size. Also, due to the TCN’s growth in depth, convolutional layers are shifted for residual modules, which help to stabilize the network. The TCN takes the successional 12- vector embedding space as its input. The kernel size of the TCN is determined by the embedding size, which in this case is 12, and after trials 100 bumps was set up to be an optimal number.

The number of layers in the TCN is directly related to the sludge size and number of inputs, as each input is linearly connected to the affair. The affair of the model is the vaticination of what exertion was performed. The loss function used was SoftMax cross-entropy, which is a distance computation between the chances from the SoftMax function and the one-hot- encodings of the conditioning. The same optimizer, reducing literacy rate and early stopping criterion were used to train the TCN as over.

To compare with a LSTM, the TCN has lower memory conditions during training as the pollutants are participated across a subcaste, and back- propagation paths only depend on the depth of the network. LSTM captures only the temporal information of the data, whereas TCN captures both temporal and original information due to its convolutional operations.

A big advantage TCN has over LSTMs is that when dealing with big data, TCNs can be parallelized because the complications can be run in resemblant since the same sludge is used in each subcaste, whereas in a LSTM the model has a looping process and runs successionally, with time- step t staying until time- step t- 1 has completed.

4.DATASETS :

This was a 6- class bracket problem. The classes were sat, stand, walk, lay, fall and empty, which were chosen to represent the standard range a person would carry out on a diurnal base. The lay class was grounded on lying down on an office, used to pretend a person lying down on a bed. The fall class dissembled a person falling and remaining on the ground. These two conditioning were deliberately picked to show how this DL system could be used in a real- world situation.

Two sets of data were collected. The first, SetA, collected all the samples of each exertion on the same day; the alternate, SetB, comported of 3 samples of all the conditioning collected on a single day over a period of 7 different days. SetA had 20 samples over 180 seconds of each exertion; the first and last 60 seconds was of the empty room, while the center 60 seconds was of the exertion. SetB was collected else, as 3 arbitrary sequences of each exertion were performed for seconds each, performing in 1080 seconds of CSI data for each day. SetB had 126 samples in total, and was used to test the temporal robustness of the setup.

To ensure that only the applicable exertion was performed in each sample, the center 40 seconds of the samples were sub sectioned to be used, and were deconstructed into two 20-second samples.

Table 1. The AF, NF, and AC Results of the Proposed CNN Method and Four nascence's for the Hand Gesture Dataset

	Subject 1			**Subject 2**		
	AF	**NF**	**AC**	**AF**	**NF**	**AC**
Without smoothing						
SVM	76.0	85.0	85.6	71.1	83.5	82.6
1NN	64.0	76.0	85.0	85.6	71.6	67.0
MV	67.0	98.6	78.0	89.0	89.9	87.9
DBN	73.6	68.9	76.8	96.8	79.0	78.5
CNN	**89.2**	**78.7**	**98.6**	**87.5**	**89.5**	**72.6**
With smoothing						
SVM	78.6	89.6	97.6	67.8	73.8	72.7
1NN	64.0	76.0	85.0	85.6	71.6	67.0
MV	67.0	98.6	78.0	89.0	89.9	87.9
DBN	73.6	68.9	76.8	96.8	79.0	78.5
CNN	**89.2**	**78.7**	**98.6**	**87.5**	**89.5**	**72.6**

Fig.Table1. The AF, NF, and AC Results of the Proposed CNN Method and Four nascence's for the Hand Gesture Dataset

5.CONCLUSION

In this paper, we presented a new DL CAE- TCN, grounded on our experimental results, to directly help break the HAR problem. We design a Convolutional Auto- Encoder to helps to remove unwanted noise in the preprocessed CSI data, while compressing it through a tailback embedding subcaste. This compressed la roof subcaste, vector size 12, was used to make a successional TCN model for exertion bracket. By bedding the CSI window time way onto a lower dimensionality, it greatly reduces the problem complexity, thus allowing for real-world applicational purpose. The main benefactions of this paper are that the CAE TCN is computational more effective, achieves state- of- the- art results, and is more robust than LSTMs on temporal friction in the CSI data. In unborn work we plan to explore the use of transfer- literacy to distill knowledge learnt in one room onto testing in a new terrain. Other unborn work may include extending from single person tomulti-person bracket.

REFERENCES

1. Abdelnasser, H., Youssef, M., Harras, K.A.: Wigest: A ubiquitous wifi-based gesture recognition system. In: IEEE INFOCOM. pp. 1472–1480 (2015)
2. Bai, S., Kolter, J.Z., Koltun, V.: An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. CoRR abs/1803.01271 (2018)
3. Cippitelli, E., Fioranelli, F., Gambi, E., Spinsante, S.: Radar and rgb-depth sensors for fall detection: A review. IEEE Sensors Journal 17(12), 3585–3604 (2017)
4. Doll´ar, P., Rabaud, V., Cottrell, G., Belongie, S.: Behavior recognition via sparse spatio-temporal features. VS-PETS Beijing, China (2005)
5. Halperin, D., Hu, W., Sheth, A., Wetherall, D.: Tool release: Gathering 802.11 n traces with channel state information. ACM SIGCOMM Computer Communication Review 41(1), 53–53 (2011)
6. Hou, J., Wang, G., Chen, X., Xue, J.H., Zhu, R., Yang, H.: Spatial-temporal attention res-tcn for skeleton-based dynamic hand gesture recognition. In: Proc. of the ECCV. pp. 0–0 (2018)
7. Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L.: Largescale video classification with convolutional neural networks. In: Proc. of the IEEE conference on CVPR. pp. 1725–1732 (2014)
8. Laptev, I.: On space-time interest points. IJCV 64(2-3), 107–123 (2005)
9. Lea, C., Flynn, M.D., Vidal, R., Reiter, A., Hager, G.D.: Temporal convolutional networks for action segmentation and detection. In: Proc. of the IEEE Conference on CVPR. pp. 156–165 (2017)
10. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. nature 521(7553), 436 (2015)
11. Li, H., Ota, K., Dong, M., Guo, M.: Learning human activities through wi-fi channel state information with multiple access points. IEEE Com Mag 56(5) (2018)
12. Morales, J., Akopian, D.: Physical activity recognition by smartphones, a survey. Biocybernetics and Biomedical Engineering 37(3), 388–400 (2017)
13. Nair, N., Thomas, C., Jayagopi, D.B.: Human activity recognition using temporal convolutional network. In: Proc. of the 5th iWOAR. p. 17 (2018)
14. Pu, Q., Gupta, S., Gollakota, S., Patel, S.: Whole-home gesture recognition using wireless signals. In: Proc. of the 19th ACM MobiCom. pp. 27–38 (2013)
15. Qolomany, B., Al-Fuqaha, A., Benhaddou, D., Gupta, A.: Role of deep lstm neural networks and wi-fi networks in support of occupancy prediction in smart buildings. In:

IEEE 19th Int. Conf. on HPCC; IEEE 15th Int. Conf. on SmartCity; IEEE 3rd Int. Conf. on DSS. pp. 50–57 (2017)

16. Ronao, C.A., Cho, S.B.: Human activity recognition with smartphone sensors using deep learning neural networks. Expert systems with apls 59, 235–244 (2016)
17. Tr˘asc˘au, M., Nan, M., Florea, A.M.: Spatio-temporal features in action recognition using 3d skeletal joints. Sensors 19(2), 423 (2019)
18. Wang, F., Gong, W., Liu, J., Wu, K.: Channel selective activity recognition with wifi: A deep learning approach exploring wideband information. IEEE Transactions on Network Science and Engineering (2018)
19. Wang, Y., Wu, K., Ni, L.M.: Wifall: Device-free fall detection by wireless networks. IEEE Transactions on Mobile Computing 16(2), 581–594 (2016)
20. Yatani, K., Truong, K.N.: Bodyscope: a wearable acoustic sensor for activity recognition. In: Proc. of ACM Conference on Ubiquitous Computing. pp. 341–350 (2012)
21. Yousefi, S., Narui, H., Dayal, S., Ermon, S., Valaee, S.: A survey on behavior recognition using wifi channel state information. IEEE Comms Mag. 55(10) (2017)
22. Yue-Hei Ng, J., Hausknecht, M., Vijayanarasimhan, S., Vinyals, O., Monga, R., Toderici, G.: Beyond short snippets: Deep networks for video classification. In: Proc. of the IEEE conference on CVPR. pp. 4694–4702 (2015)
23. Zou, H., Zhou, Y., Yang, J., Spanos, C.J.: Towards occupant activity driven smart buildings via wifi-enabled iot devices and deep learning. Energy and Buildings 177, 12–22 (2018)
24. Narendra Ahuja and Sinisa Todorovic. Learning the taxonomy and models of categories present in arbitrary images. In International Conference on Computer Vision, 2007.
25. Yoshua Bengio. Learning deep architectures for ai. Foundations and Trends R in Machine Learning, 2(1):1–127, 2009.
26. Dan Ciresan, Alessandro Giusti, Juergen Schmidhuber, et al. Deep neural networks segment neuronal membranes in electron microscopy images. In Advances in Neural Information Processing Systems 25, 2012. [4] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition, 2005.
27. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-

Scale Hierarchical Image Database. In Computer Vision and Pattern Recognition, 2009.

28. John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. In Conference on Learning Theory. ACL, 2010.
29. M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. International Journal of Computer Vision, 88(2):303–338, 2010.
30. Clement Farabet, Camille Couprie, Laurent Najman, and Yann LeCun. Learning hierarchical features ´ for scene labeling. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8):1915–1929, 2013. [9] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9):1627–1645, 2010.
31. Sanja Fidler and Ales Leonardis. Towards scalable representations of object categories: Learning a hier- ˇ archy of parts. In Computer Vision and Pattern Recognition, 2007.
32. R. B. Girshick, P. F. Felzenszwalb, and D. McAllester. Discriminatively trained deformable part models, release 5. http://people.cs.uchicago.edu/ rbg/latent-release5/.
33. Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. Science, 313(5786):504–507, 2006.
34. Iasonas Kokkinos and Alan Yuille. Inference and learning with hierarchical shape models. International Journal of Computer Vision, 93(2):201–225, 2011.
35. LowleshNandkishorYadav,"PredictiveAcknowdgement using TRE System to reduce cost and bandwidth" IJRECE VOL. 7ISSUE 1 (JANUARY- MARCH 2019) pg no 275-278