© November 2022 | IJIRT | Volume 9 Issue 6 | ISSN: 2349-6002

Design and Analysis of Human Activity Recognition using Deep learning Models-A Review

Preethi Salian K[1,2], D. Veerabhadra Babu[3]
[1]*Research Scholar, Srinivas University, Mukka, Mangalore, India*
[2]*Assistant Professor, NMAM Institute of Technology, Nitte, Udupi, India*
[3]*Professor, Srinivas University, Mukka, Mangalore, India*

Abstract-- **Numerous applications, including those in healthcare and smart homes, depend on the detecting human activity- Human Activity Recognition (HAR). In this paper, we present a thorough overview of recent developments and difficulties in deep learning based HAR. In spite of the fact that there has been numerous review of HAR, mostly focused on the arrangement of HAR and examined the most advanced HAR systems that have been deployed using traditional machine learning techniques. Recently, other reviews of research using deep models for HAR have also been conducted, albeit these reviews only cover a small number of deep models and their variations. A thorough and in-depth analysis of HAR using recently created deep learning techniques is still required.**

Index Terms— **Deep learning Model, Human activity recognition, Machine learning.**

I. INTRODUCTION

Numerous applications and services, including healthcare home automation, monitoring, fitness, traffic planning, augmented reality and control, security and accurate advertising, depend on knowledge of human activities [1]. For instance, tracking a people every day activities can be used to determine how many calories he has consumed in a day, which can then be used to recommend a healthy diet for him to stay fit and healthy; similarly, detecting an elderly person's tendency to fall can be utilized to summon emergency support to prevent serious misfortune.

Regular ML techniques can be utilized to identify human behavior. However, traditional machine learning techniques for HAR need that relevant characteristic be designed and chosen. Although this procedure requires time-consuming human labor and specialized knowledge, the developed and chosen features could nonetheless perform below expectations. DL techniques have been put forth recently to lessen the strain of manually engineering features [2]. Deep learning techniques are excellent for HAR and can help HAR in a variety of ways. It first saves the labour of manually developing features, which frequently calls for specialized knowledge. Second, it has demonstrated greater HAR accuracy compared to traditional methods [3]. Thirdly, it can understand from data, which is crucial for HAR because it is impractical to collect a lot of labeled activity data. Fourth, it can handle data related to activity from various users, different device types, and various device positions. It also has the powerful capacity to learn valuable attributes from raw data.

Figure 1 illustrates the connection between AI, ML and DL. A subset of machine learning techniques called "deep learning" comprises several levels of representation. Deep learning networks are also referred to as DNNs since they are ANNs with more than a hidden layer. The authors of Reference [4] divided deep learning models into three categories: deep networks for supervised and unsupervised learning, and hybrid approach. To categorize deep learning models, this work modifies the classification as in the Reference[4] and separate them into deep generative, discriminative, and hybrid models.

Deep generative models seek to either understand the cumulative probability distribution of information and the classes to which they belong [5] or to study usable demonstrations of data through un-supervised learning. Auto-encoders and Restricted Boltzmann Machines (RBMs), [6], generative adversarial networks (GANs), and its variants are common generating models. Discriminative models attempt to comprehend the distribution of conditional probabilities of categories on the data when the labelled data is either directly or indirectly available [7]. Deep discriminative models that are frequently used include recurrent neural networks (RNNs), convolutional neural networks (CNNs), and their variants. In deep hybrid models, which contain a training algorithm and a discriminative model, the output of the generative model is typically used as the input to the due to the different types for regression or classification.

In the literature, there have been a number of HAR surveys to date. Poppe [10] explored several picture representation techniques as well as techniques for classifying actions while reviewing studies on vision-based human action recognition. Aggarwal and Ryoo [11] looked at recognition methods for both high-level and fundamental human behaviors, including sequential approaches, space-time approaches, and hierarchical approaches. Additionally, they suggested a HAR approach-based taxonomy. Chen et al. [12] looked at a number of aspects of sensor-based activity detection with a focus on data, modelling knowledge-driven methods for activity monitoring, and recognition. HAR works based on motion, location, and other contextual information were explored by Incel et al., who also provided a taxonomy of activity recognition on mobile devices. They also discussed how HAR works on phones and its difficulties.

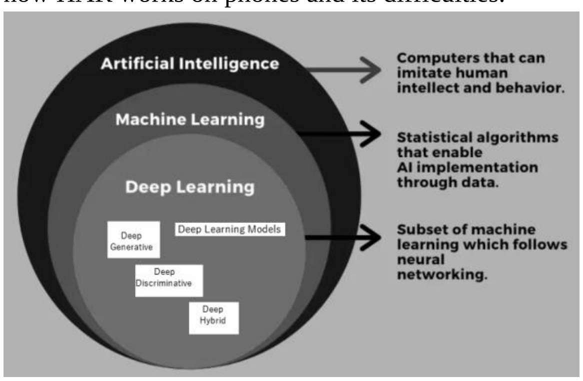

Figure 1: Relationship between AI, Ml and DL

II. LITERATURE REVIEW

A.Review Stage

Despite their apparent similarity, Action denotes to some work that is carried out; Activity shows the condition of being vigorously occupied; Behavior denotes how one conducts themselves in the exterior interactions of life. Each item has a distinct synonym and is the source of several jargons. The definitions of the three concepts can be expanded upon in sociology and philosophy. Action and social action are linked, according to Max Weber [13]. He holds that an actor's sentiments or experiences serve as the driving force behind intentional action. The only way to think about conduct is as a reflection of cues or impulses. Additionally, activities can be divided into levels based on their complexity, ranging from low to high [14]. In this essay, do not distinguish between activities of different levels for the sake of clarity. Instead, broaden Reference's taxonomy of activity [15] and divide human activities into many categories in accordance with the application domains. Locomotion, method of transportation, usage of phone, leisure, daily activities, health-related activities, security, and security are the main types of activities that are investigated in the survey.

B. Sensor used for Human Activity Recognition(HAR)

a) Ambient Sensors: Installing ambient sensors at fixed places is necessary to detect activities, which often include a server and a client. This refers to both clients and servers used as sensors in order to keep things simple. Cellular, Global Navigation Satellite System (GNSS), Zigbee, Wi-Fi, RFID and FM (Frequency Modulation), are the most common ambient sensors used for HAR**.**

b) Wearable Sensors: These are portable and frequently integrated into smart devices. The accelerometer, gyroscope, camera, light sensor, magnetometer, light, acoustic and Bio sensor, magnetometer, and barometer are typical sensors used for HAR. Wearable sensors are more accurate than ambient sensors because the user typically takes the sensors about. In addition, they do not experience the coverage problem that environmental sensors do. Wearable sensors, however, can only identify the activities of the person wearing them; they do not offer various person detection. Along with ambient sensors and portable sensors, other sensors, including event camera, can be used for HAR.

HAR with wearable sensors in survey, Lara and Labrador [1] covered various activity kinds, design challenges, and HAR detection techniques. They specifically assessed 28 systems for recognition precision, energy usage, intrusiveness, and flexibility. Radio-based HAR techniques, primarily those based on Wi-Fi, Zigbee, and Radio Frequency Identification (RFID), were surveyed by Wang and Zhou [17]. Shoaib et al[18] .'s assessment of smartphone-based HAR systems that merely make use of the device's internal sensors. Yousefi et al. steered a survey on HAR via Wi-Fi Channel State Information (CSI)[19].

Mukhopadhyay [20] analyzed pertinent technology and procedures for using wearable sensors to track human activities.

In this tutorial, systematically go over the HAR utilizing deep learning techniques and approaches. This paper begins by outlining a taxonomy of human activity before introducing widely used deep model construction strategies, sensors preprocessing methods, and evaluation strategies. Though numerous recent publications have examined various works employing deep models for action recognition, they only cover a small number of deep models and exclude both preprocessing approaches and assessment measures. The primary driving force behind this effort is continued requirement for a thorough, deeper investigation on HAR using newly emerging DL techniques.

III. DEEP LEARNING FOR HAR

The summary of applying deep learning techniques for HAR is given. First, information is gathered from various sensors. This information can include pictures, Wi-Fi CSI, gyros accelerations, barometer readings, gyroscope values, readings from biosensors, sound, and more. Second, preprocessing methods like scaling, Third, different deep models (such as RBM, auto-encoder, and RNN) can be selected in the model construction component to learn valuable features. At the top layer, a classifier comes next. After making a model, use the input data to train it. The network's parameters, including its masses, are adjusted throughout training. Finally, forecast future data behavior using the trained model.

Preprocessing, Model building and Evaluation will be discussed in the following sections. Figure 2 shows the involvement of deep learning model in HAR.

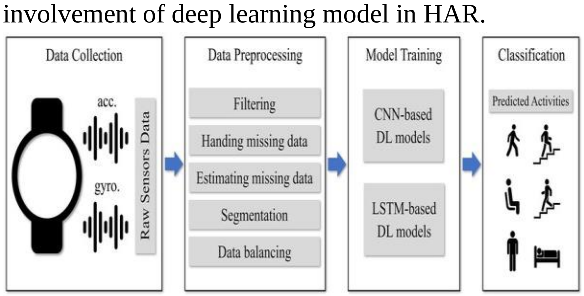

Figure 2: Overview of deep learning for HAR.

A. Preprocessing

Segmentation, Filtering , handling missing data and data balancing are some of the primary preprocessing methods. In the following sections, it is discussed in detail.

Filtering: Data filtering typically entails removing information that the reader won't find useful or that might be confusing. Large and complicated data sets are frequently produced as a result of reports and database queries from database tools. Data that is redundant or unreliable can befuddle or disorient a user. Results can also be more effective by filtering the data. Data filters can also be used in other situations to limit access to private information.

Segmentation: Before feeding the data to a deep model, segmentation may be required depending on the type of sensor data being used. It is possible to identify actions like walking and gesturing from a single picture using image-based HAR. However, we unable to distinguish between different activities based on an accelerometer measurement, Wi-Fi CSI, barometer reading, etc. It is due to the fact that no data sample, with the exception of an image, can fully represent an action. It should be noted that because separate sensors may have different sampling frequencies, while merging data from various sensors to conduct HAR, must bring into line the data to the same time window. Subsequently the sampling frequency for the similar sensor might not be constant, some deep models (such as the auto-encoder) also frequently need stabilizing the input samples within a temporal window by utterance [21]. To handle data with erratic time intervals, certain time-aware models have been put forth recently.

Dealing with Missing Data: Missing data might result in decreased efficiency, skewed outcomes, and increased complexity [22]. Different approaches can be taken to cope with missing data. If any of the samples have incomplete data in their attributes, a straightforward solution is to ignore those samples or even to erase the entire attribute. If there are a lot of samples with missing values, this might considerably limit the amount of training data, which would lower the classifier performance developed using that data. The mean or average of the non-missing data in the same property may be used to fill in the missing values as an alternative. It is quick and simple, but it does not take into account how different properties are correlated.

B. Deep Learning Models

Convolution Neural network: Contrary to AE's, CNNs and RBM-based methods are racist and discriminatory models that substitute the operation for general multiplications in at smallest one of the layers [23]. Two of the techniques utilized in CNNs are pooling and convolution. The three key ideas in convolution are sparse connection, broader public, and equal variant representations. The output of pooling is a statistic of the inputs. Different pooling functions include max pooling, pooling layer, L2-norm accumulating, and tree pooling.

The four types of layers, Convolution layer, pooling layer detector layer, and fully connected layer that are often present in basic CNNs. A deep CNN can be formed by piling these layers. Numerous CNN versions have been projected as a consequence of the great performance of CNNs in various areas, particularly in classification of images [24]. The term "tiled CNN" refers to one of the common variations [25]. Convolution's weight-sharing approach can greatly decrease the number of parameters, but it also prevents the product from learning additional invariant characteristics.

The tiled CNN offers a solution to this issue because it can learn different feature maps by restricting weights functions that are k steps apart to be identical. The parameter k in this case is termed the tile size, and when the tile size is equal to 1, the tiled CNN is equivalent to the basic CNN. As illustrated in Figures 3 tiled CNNs are superior than basic CNNs in that they not only minimize the number of parameters that must be trained, but also enable the learning of additional invariances (b). Tiled CNNs have been shown to perform better than standard CNNs in the references [25]. There are numerous different variations of the standard CNN in addition to tiled CNNs. The authors of Reference [26] provided numerous CNN variations that are enhanced in various ways.

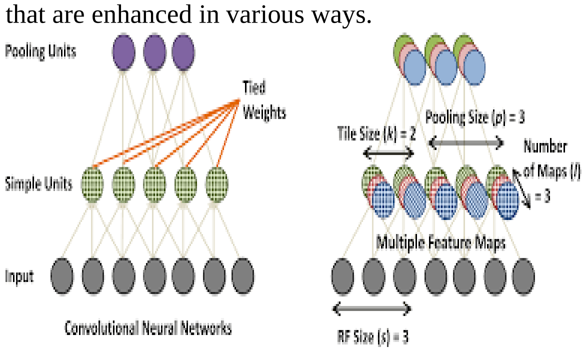

Figure 3: Basic CNN and Tiled CNN

The term "tiled CNN" refers to one of the common variations [25]. Convolution's weight-sharing approach can greatly decrease the number of parameters, but it also prevents additional invariant characteristics. The tiled CNN is a remedy for this issue, by requiring weights functions to be k steps apart and equal, can learn different feature maps. The parameter k in this case is termed the tile size, and when the tile size is equal to 1, the tiled CNN is equivalent to the basic CNN. As illustrated in Figures 3 tiled CNNs are superior than basic CNNs in that they not only minimize the number of parameters that must be trained, but also enable the learning of additional invariance. Tiled CNNs have been shown to perform better than standard CNNs in the references. There are numerous different variations of the standard CNN in addition to tiled CNNs. The authors of Reference provided numerous CNN versions that are enhanced in terms of layer architecture, loss function, activation function, computation, regularization, and optimization.

CNNs have also been extensively utilized for HAR as one of the early effective deep learning techniques. Using a CNN, Ronao and Cho [27] identified six different types of locomotor behaviors.

LSTM (Long Short-term Memory) – A Recurrent Neural network (RNN): Due to concerns with vanishing or inflating gradients, the techniques used to train the RNN frequently have a restriction on the memory generated from the recurrent connections. Utilizing LSTM RNN is a well-liked method to lessen the possessions of waning and discharge gradients [28]. The LSTM RNN differs from traditional RNNs in that memory cells are used in place of hidden units [29]. The LSTM was first proposed in Reference [30], then it was modified in Reference [66], which is when it started to gain popularity.

Although the typical LSTM has shown promise in a number of applications, it might not be able to comprehend input patterns that are more complex than a sequence. a tree-structured LSTM network, S-LSTM, is suggested in the survey [31] as a solution to this problem. S-LSTM memory blocks, which include two forget gates, input gate, an output gate and a cell gate, make up the S-LSTM network. Although the S-LSTM network has a larger computational cost than the normal LSTM network, it performs better in challenging sequential modelling situations. Gated

Recurrent Units (GRUs) can dynamically detect sequential dependences, are a noteworthy variation of the fundamental RNN. It is more computationally effective than the LSTM while having a gated structure that is comparable to that of the LSTM. The LSTM and the GRU have some similar traits [32]. They differ from conventional recurrent units in that they have an extra module in the updating method from time ts to time ts + 1. They can remember already-existing traits and skip over several temporal steps thanks to this trait. They do, however, differ in some ways. For instance, the output gate of the LSTM unit will control memory content exposure. Without exercising any oversight, the GRU merely discloses all content. Reference [32] lists more precise parallels and divergences between the LSTM and the GRU.
RNNs that are discriminative are employed in the context of HAR. Its training is carried out in a supervised manner, which reduces the cost function of the output of the network and the associated label. They showed that the unidirectional DRNN performs better than the cascaded DRNN and the bidirectional DRNN. Similar to this, Inoue et al. recognized human actions from acceleration signals using a DRNN made of LSTMs. Multiple LSTM networks are combined using a HAR approach described by Guan and Plötz, who demonstrated that the combined LSTM networks outperform the individual LSTM networks. A spatial-temporal attention and semantic graph-based structural-RNN technique for identifying group activities from films was put out by Qi et al. A binarized BLSTM RNN was created by Edel and Köppe for the purpose of identifying routine activities and motion patterns. The binarized BLSTM is more computationally efficient than the normal LSTM because it may drastically reduce memory size and accesses by substituting bit-wise operations for arithmetic operations.

IV. PERFORMANCE ANALYSIS IN HAR

The following two methods can be used for evaluating the performance of the deep neural model used in HAR, cross validation and holdout.

A. Cross Validation

Cross validation is a common evaluation method. The most popular kind of cross validation is n-fold crossing validation, which divides the data into k sets and utilizes one set as testing data and the other n sets as training data. In this procedure, which is performed k times, each of the sets is used as test data just once. K often has a value of 10 or 15. The model's performance is calculated as the mean across all n trials. There are various cross validation techniques besides n-fold cross validation, like leave-one-out and shuffle-split. When using data, validation is more reliable and effective than the holdout method. The biggest disadvantage of cross validation is that it makes calculations more expensive. Cross validation is useful for small datasets. While assessing deep models, it is also advisable to do the cross validation process several times with a different random seed.

B. Holdout

A simple evaluation method called the holdout randomly divides the data into training, test, and validation accuracy. The training data set is used to fit the model. Utilizing the validation data, the fit model's efficacy is assessed in order to determine the ideal parameter values. To assess the model's generalization error, the trained model then is tested on the testing data. The holdout technique, which has the advantages of simplicity and speed, is frequently used when a large dataset is available or the training of the system is taking a long time. However, it has a high degree of variability because the classification accuracy might fluctuate significantly across training and test sets of data. Due to the stochastic nature of deep models and the fact that they incorporate several randomness sources, such as random initial weights, it is usual practice to repeat the evaluation several times (for example, 30 times) [26]. In other words, the same model is assessed repeatedly on the same data with only the seed used to produce random numbers changing. The model's performance can therefore be interpreted as the average performance.
There are other evaluation criteria used to assess the effectiveness of the deep model in addition to the evaluation techniques mentioned above. Deep models' success in classifying data is frequently evaluated using a variety of metrics, including accuracy, recall, error rate, F-measure, precision, ROC curve, and area under the curve. However, these measures are equally relevant to multi-class classification problems. To be clear, introduce such metrics using confusion matrix of a binary task to forecast if a human is present in a picture.

Table I: Confusion matrix of Binary classification task

	Positive – Actual	Negative – Actual
Positive Predicted	TP	FP
Negative Predicted	TN	TN

The confusion matrix, which is used to explain various metrics, is displayed in Table 1. A true positive (tp) is when the predicted label of the data example is 1, and the actual label is also 1, and a true negative (tn) is when the predicted label of the data example is 0, and the actual label is also 0.

VII. CONCLUSION

A thorough explanation on DL techniques for HAR is explained in this paper. In particular, a thorough introduction has been given from preprocessing, model development through evaluation. Although HAR is the primary focus of this study, similar techniques can be used to other tasks like speech recognition and image processing. In the following, open challenges has been given that remain in deep learning or HAR. There are several levels of activity. While basic activities (like walking or jogging) can often be recognized by a single sensor (such as an accelerometer), complicated activities (like participating in a meeting) necessitate the use of numerous sensors in order to achieve reliable recognition. Recent research has shown that the use of several sensors increases the classification accuracy of locomotors activities. Further research must be done on techniques for classifying more universal activities based on multimodal information from both stationary and mobile sensors using Deep models with less labeled data**.** Unsupervised data can be used by generative deep models (like AEs and GANs), but they are not immediately suitable for HAR. It is necessary to create new deep models that can be trained using little labeled input. Promising are hybrid models, which mix discriminative and generative models. Even though they are still in their infancy, numerous studies have been conducted. New semi-supervised deep models and active deep models that require less labeled data are also promising.

REFERENCE

[1] Oscar D. Lara and Miguel A. Labrador. 2012. A survey on human activity recognition using wearable sensors. IEEE Commun. Surv. Tutor. 15, 3 (2012), 1192–1209.

[2] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. Nature 521, 7553 (2015), 436.

[3] Fuqiang Gu, Kourosh Khoshelham, and Shahrokh Valaee. 2017. Locomotion activity recognition: A deep learning approach. In Proceedings of the IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC'17). IEEE, 1–5.

[4] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. Nature 521, 7553 (2015), 436 4. Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder-decoder approaches. In Proceedings of Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation. 103–111.

[5] Ruslan Salakhutdinov. 2015. Learning deep generative models. Annu. Rev. Stat. Appl. 2 (2015), 361–385

[6] Michael Tschannen, Olivier Bachem, and Mario Lucic. 2018. Recent advances in autoencoder-based representation learning. In Third workshop on Bayesian Deep Learning (NeurIPS'18).

[7] Yanyang Yan, Wenqi Ren, and Xiaochun Cao. 2018. Recolored image detection via a deep discriminative model. IEEE Trans. Inf. Forens. Secur. 14, 1 (2018), 5–17.

[8] Ross Girshick. 2015. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision. 1440–1448.

[9] Yoshua Bengio, Nicolas Boulanger-Lewandowski, and Razvan Pascanu. 2013. Advances in optimizing recurrent networks. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 8624–8628.

[10] Ronald Poppe. 2010. A survey on vision-based human action recognition. Image Vis. Comput. 28, 6 (2010), 976–990.

[11] Jake K. Aggarwal and Michael S. Ryoo. 2011. Human activity analysis: A review. ACM Comput. Surv. 43, 3 (2011), 16.

[12] Liming Chen, Jesse Hoey, Chris D. Nugent, Diane J. Cook, and Zhiwen Yu. 2012. Sensor-

based activity recognition. IEEE Trans. Syst. Man Cyberneti. C 42, 6 (2012), 790–808.

[13] Tom Campbell. 1981. Seven Theories of Human Society. Clarendon Press, 169–229

[14] Thomas B. Moeslund, Adrian Hilton, and Volker Krüger. 2006. A survey of advances in vision-based human motion capture and analysis. Comput. Vis. Image Understand. 104, 2–3 (2006), 90–126.

[15] Ozlem Durmaz Incel, Mustafa Kose, and Cem Ersoy. 2013. A review and taxonomy of activity recognition on mobile phones. BioNanoScience 3, 2 (2013), 145–171.

[16] Arnon Amir, Brian Taba, David Berg, Timothy Melano, Jeffrey McKinstry, Carmelo Di Nolfo, Tapan Nayak, Alexander Andreopoulos, Guillaume Garreau, Marcela Mendoza, et al. 2017. A low power, fully event-based gesture recognition system. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 7243–7252.

[17] Shuangquan Wang and Gang Zhou. 2015. A review on radio based activity recognition. Dig. Commun. Netw. 1, 1 (2015), 20–29.

[18] Muhammad Shoaib, Stephan Bosch, Ozlem Incel, Hans Scholten, and Paul Havinga. 2015. A survey of online activity recognition using mobile phones. Sensors 15, 1 (2015), 2059–2085.

[19] Siamak Yousefi, Hirokazu Narui, Sankalp Dayal, Stefano Ermon, and Shahrokh Valaee. 2017. A survey on behavior recognition using wifi channel state information. IEEE Commun. Mag. 55, 10 (2017), 98–104. [2] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. Nature 521, 7553 (2015), 436.

[20] Subhas Chandra Mukhopadhyay. 2014. Wearable sensors for human activity monitoring: A review. IEEE Sens. J. 15, 3 (2014), 1321–1330.

[21] Fuqiang Gu, Kourosh Khoshelham, Shahrokh Valaee, Jianga Shang, and Rui Zhang. 2018. Locomotion activity recognition using stacked denoising autoencoders. IEEE IoT J. 5, 3 (2018), 2085–2093.

[22] Salvador García, Julián Luengo, and Francisco Herrera. 2015. Data Preprocessing in Data Mining. Springer.

[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 770–778.

[24] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. 2018. Recent advances in convolutional neural networks. Pattern Recogn. 77 (2018), 354–377.

[25] Quoc V. Le, Jiquan Ngiam, Zhenghao Chen, Daniel Chia, Pang W. Koh, and Andrew Y. Ng. 2010. Tiled convolutional neural networks. In Advances in Neural Information Processing Systems. 1279–1287.

[26] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. 2018. Recent advances in convolutional neural networks. Pattern Recogn. 77 (2018), 354–377.

[27] Charissa Ann Ronao and Sung-Bae Cho. 2016. Human activity recognition with smartphone sensors using deep learning neural networks. Expert Syst. Appl. 59 (2016), 235–244.

[28] 28. Fuqiang Gu, Xuke Hu, Milad Ramezani, Debaditya Acharya, Kourosh Khoshelham, Shahrokh Valaee, and Jianga Shang. 2019. Indoor localization improved by spatial context—A survey. Comput. Surv. 52, 3, Article 64 (Jul. 2019), 35 pages. https://doi.org/10.1145/3322241.

[29] Hojjat Salehinejad, Sharan Sankar, Joseph Barfett, Errol Colak, and Shahrokh Valaee. 2017. Recent advances in recurrent neural networks. arXiv:1801.01078. Retrieved from https://arxiv.org/abs/1801.01078.

[30] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. Neural Comput. 9, 8 (1997), 1735–1780.

[31] Xiaodan Zhu, Parinaz Sobihani, and Hongyu Guo. 2015. Long short-term memory over recursive structures. In Proceedings of the International Conference on Machine Learning. 1604–1612.

[32] JungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural

networks on sequence modeling. In Workshop on Deep Learning (NIPS'14).