Heaven's Light is Our Guide

Course No: CSE 4206

Realistic Activity Recognition using Sensors with Deep Convolutional Neural Network

Presented By
Md. Al Siam
Roll: 1603008
Computer Science and Engineering,
Rajshahi University of Engineering and Technology

Supervised By
Abu Sayeed
Assistant Professor
Computer Science and Engineering,
Rajshahi University of Engineering and Technology

Presentation Outline

- Introduction
- Motivation
- Objectives
- Literature Review
- Dataset Description
- Proposed Methodology
- Results
- Conclusion
- Future Works
- References

What is IoT?

The **Internet of Things** (**IoT**) describes physical objects (or groups of such objects) with **sensors, processing ability, software, and other technologies** that **connect and exchange data** with other devices and systems over the internet or other **communications networks**.

- provides businesses with a real-time look
- enables companies to automate processes
- reduce labor costs
- cuts down on waste and improves service delivery

What is IoT?

IoT is getting of the most important **technologies of everyday life**, and it will continue to pick up steam as more businesses realize the potential of connected devices to keep them competitive.

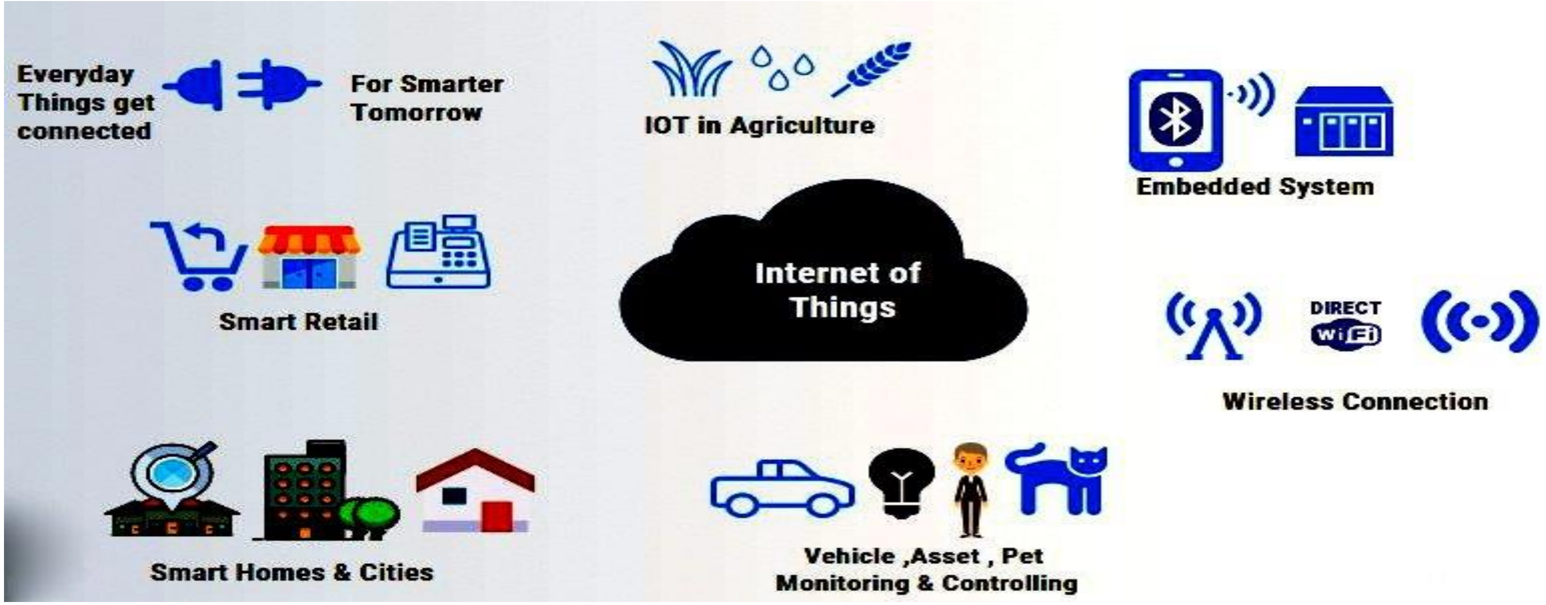


Fig. 1: Examples of IoT in Daily Life

Sensor and Sensor Data

- A sensor is a device, module, machine, or subsystem that **detects events or changes in its environment** and sends the information to other electronics, frequently a computer processor. Sensors are always used with other electronics.

- Sensor data is the output of a device **that detects and responds** to some type of **input from the physical environment**.

- Generally, sensor data can be represented in **tabular form** as most of the times they are nothing but some values representing some physical attributes

- In recent years, **smartphones with in-built sensors** have become widely available. Sensor data can be collected and preprocessed as per user requirement by creating a mobile application for the data collection process. Hence, the domain of sensor datasets generated by smartphones is becoming larger day by day.

Motivation

- Rapid Growth of IoT Sensors
- Increasing Availability of Sensor Data
- Using Activity Recognition Methods in Professional Contexts
- Scope of Using Sensor Data for Human Activity Recognition

Objectives

- Processing Sensor Data with Proper Windowing
- Increasing Size of Data with Reasonable Window Overlapping
- Using Deep Convolutional Neural Network on Sensor Data
- Studying about the Usability of Sensor Data from Real Work Environment

Literature Review

Activity Recognition Using Inertial Sensors

- The previous work in HAR based on inertial measurement units (IMUs) has mainly focused on the use of **acceleration signals**.

- **Gupta et al.** used accelerometer signal data and classified activities with **Naïve Bayes and k-nearest neighbor (k-NN)** methods which could demonstrate 98% accuracy [1].

- **Kwapisz et al.** collected labeled accelerometer data from twenty-nine users as they performed daily activities such as **walking, jogging, climbing stairs, sitting, and standing**, and then aggregated this time series data into examples that summarize the user activity over 10- second intervals. They made **a predictive model** for activity recognition [2].

Literature Review

Activity Recognition Using Inertial Sensors

- **Liu et al.** used an **accelerometer signal** which was captured from a **Wii remote** and recognized a predefined set of eight gestures using a DTW-based method [3]

- Several studies have depicted that it is possible to classify human gestures from palm-worn devices [4, 5].

Literature Review

Activity Recognition Using Inertial Sensors

- The **gestures while writing alphabets** is a good example of human activity recognition as there are different gestures while writing different letters. The gestures while writing can be captured with sensor devices. Many experiments on writing alphabets has been conducted using sensor data so far.

- Writing alphabets can be captured by **IMU Sensors** like **accelerometers, gyroscopes**. **Wrist worn devices, mobile phones** which have sensors built-in have been widely used in the experiments.

Literature Review

Activity Recognition Using Inertial Sensors

- **Lin et al.** investigated the **orientations of the surfaces in which users were to write characters**, the stabilization (support) point of the hand, and the rotation injection technique for data augmentation which uses a rotation matrix [6]. They obtained a remarkably high accuracy of 99.99% to recognize 62 characters by 10 subjects with a machine-learning-based approach.
- **Chen et al.** investigated **real-time fingertip detection in frames captured from smart glasses**. They built a synthetic dataset using **Unity3D** and proposed a modified mask regional convolutional neural network. Their method could detect fingertip for air-writing in a minimal length of time for each frame [7].

Literature Review

Activity Recognition Using Inertial Sensors

- **Kim et al.** experimented with the **WiTA dataset**, which contains **air-writing** data for **Korean and English alphabets** collected by RGB cameras [8].

- **Abir et al.** extensively **investigated different interpolation techniques** on seven publicly available sensor based **air-writing datasets** and developed a method **to recognize air-written characters using a 2D-CNN model**. They interpolated sensor data in order to fit the data in deep 2D-CNN model. They obtained remarkable accuracy compared to the latest experiments regarding all of the datasets they used. [9]

Literature Review

Activity Recognition in Professional Contexts

- Recently, very little attention was given to the recognition of activities in **professional contexts**. Special interest seems to be rising for tracking the activities of **medical practitioners, such as doctors and nurses** [10] and, in a smaller scale, activities such as **cooking** [11].
- **Joshua and Varghese** [12] studied **masonry activities using accelerometers in a laboratory** and small-scale setting. Their study showed up to a 80% classification accuracy in relatively unconstrained environments.

Literature Review

Activity Recognition in Professional Contexts

- **Akhavian and Behzadan** [13], simulated, using only two subjects, a three-class construction activity setup that included **sawing, hammering and turning a wrench** and loading and unloading activities. Their three-class model obtained accuracies close to 90%, with high variability on users and activities.
- The rest of the studies related to construction focused on complementary cases without human focus, such as **tracking the activity of particular equipment** [14], or **machinery** [15].

Dataset Description

Used Dataset

VTT-ConIot Dataset: A Realistic Dataset for Activity Recognition of Construction Workers Using IMU Devices

- ❖ VTT-ConIoT is a publicly available dataset for the evaluation of **human activity recognition (HAR)** from **inertial sensors** in professional construction settings.

- ❖ The dataset contains data from **13 users** and **16 different activities**, which is collected from three different wearable sensor locations.

Dataset Description

VTT-ConIot Dataset

- ❖ The participants of working age in a range from **25 to 55 years old**
- ❖ Data was collected from 13 users for 16 different activities performed for a duration of **one minute each** in an "in-lab" setup that mimics the activities observed at a construction site.
- ❖ Among the 13 participants, **10 were men** and **3 were women**, a gender distribution similar to what can be seen in typical construction sites in Europe.
- ❖ The subjects were wearing **sensorized clothes** that incorporated three (Inertial Measurement Unit) IMU sensors
- ❖ One sensor located in the **hip**, and two other ones located near the **shoulder** of the nondominant hand of each participant

Dataset Description

VTT-ConIot Dataset

Fig. 2. Depiction of the sensor locations. From left to right: planned locations, depiction of the actual sensors and example setup in the working clothes.

Dataset Description

VTT-ConIot Dataset

The sensors provide:

- ❖ 3-axis accelerometer data
- ❖ 3-axis gyroscope data data
- ❖ 3-axis magnetometer data

Dataset Description

VTT-ConIot Dataset

Sampling Rates of the Used Devices

- Accelerometer: 103 Hz
- Gyroscope: 97 Hz
- Magnetometer: 97 Hz
- Barometer: 0.033 Hz

After collecting the raw data with the mentioned devices, the gyroscope and magnetometer data was resampled using linear interpolation to synchronise with the accelerometer data. This Interpolated version of data is published and publicly available.

Dataset Description

VTT-ConIot Dataset

- There are around **6000 signal samples** for each user and each activity in the dataset.
- There is lacking of 1 to 2 signal samples in samples per user and activity, which is very negligible
- A user refused to perform a activity. This data is missing. This is related to floorwork.

Dataset Description

VTT-ConIot Dataset

Name of the Activities

The activities are classified into **6 major classes** and **16 granular classes**. They are:

- **Painting:**

  Roll-Painting

  Spraying-Paint

  Leveling-Paint

Dataset Description

VTT-ConIot Dataset

Name of the Activities

- **Cleaning:**
  - **Vacuum-cleaning**
  - **Picking-objects**

- **Climbing**
  - **Climbing-stairs**
  - **Jumping-down**
  - **Stairs-Up-Down**

Dataset Description

VTT-ConIot Dataset

Name of the Activities

- **HandsUp**
  - **Laying-back**
  - **HandsUp-high**
  - **HandsUp-low**
- **FloorWork**
  - **Crouch-floor**
  - **Kneel-floor**

Dataset Description

VTT-ConIot Dataset

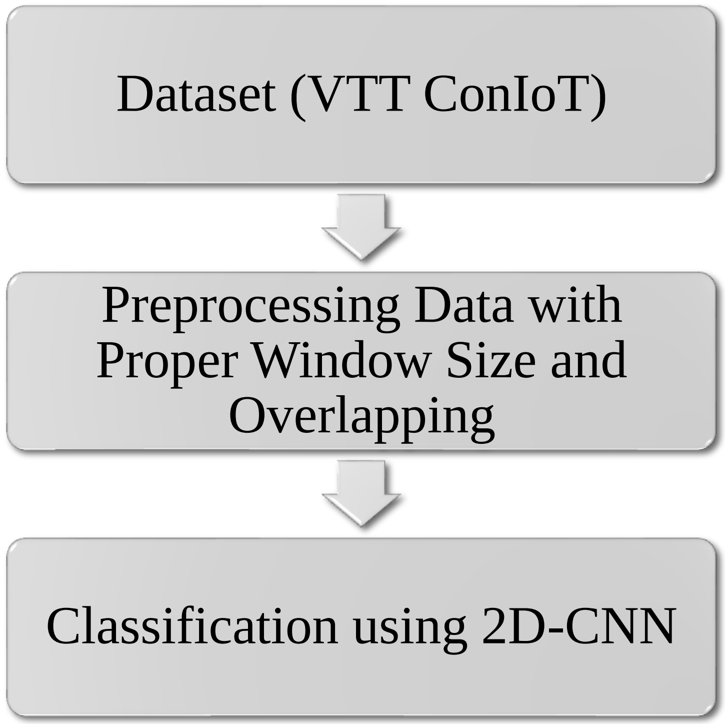
<u>Name of the Activities</u>

- **Walking Displace**
  - **Walk-straight**
  - **Walk-winding**
  - **Pushing-cart**

Challenges

- Increasing samples for small size signal data with window segmentation
- Selecting proper window size for making samples
- Using proper features that is generally representative to the labels
- Usability of CNN for classifying signal data

Proposed Methodology



Proposed Methodology

Preprocessing Data with Proper Window Size and Overlapping

We have a **6000-length signal** data for each user and each activity. As the activities are repetitive, we can use segmentation and **windowing**.

The length of segmentation window (frame size) and overlapping area (hop size) has to be defined considering the repeatitive nature of the corresponding dataset.

Proposed Methodology

Preprocessing Data with Proper Window Size and Overlapping

Then we get a window which contains an activity level.

As every window has its length (frame size) and a width (number of features taken), it can be regarded as a 2-dimensional image.

Here 2D CNN comes in picture. We will experiment considering each data frame and classify it with 2D CNN.

Proposed Model

Operation Group	Layer Name	Filter Size	No. of Filters	Stride Size	Padding Size	Activation Function	Output Size*	No. of Parameters*
-	Input	-	-	-	-	-	$3000 \times 7 \times 1$	0
Group1	Conv1-1	2×2	32	1×1	1×1	ReLU	$3000 \times 7 \times 32$	160
	Conv1-2	2×2	32	1×1	1×1	ReLU	$3000 \times 7 \times 32$	4,128
	MaxPool1	2×2	1	2×2	0	-	$3000 \times 7 \times 32$	0
	Dropout			$p = 10\%$			$1500 \times 4 \times 32$	0
Group2	Conv2-1	2×2	64	1×1	1×1	ReLU	$1500 \times 4 \times 64$	8,256
	Conv2-2	2×2	64	1×1	1×1	ReLU	$1500 \times 4 \times 64$	16,448
	MaxPool2	2×2	1	2×2	0	-	$750 \times 2 \times 64$	0
	Dropout			$p = 20\%$			$750 \times 2 \times 64$	0
Group3	Conv3-1	2×2	128	1×1	1×1	ReLU	$750 \times 2 \times 128$	32,896
	Conv3-2	2×2	128	1×1	1×1	ReLU	$750 \times 2 \times 128$	65,664
	MaxPool3	2×2	1	2×2	0	-	$375 \times 1 \times 128$	0
	Dropout			$p = 20\%$			$375 \times 1 \times 128$	0
Group4	Flatten	-	-	-	-	-	48000	0
	Dense	-	-	-	-	ReLU	512	1,638,912
	Dropout			$p = 50\%$			512	0
	Dense	-	-	-	-	Softmax	16	13,338
							Total	1,779,802

Table 1. Network architecture of the 2D-CNN model for human activity recognition

Proposed Model

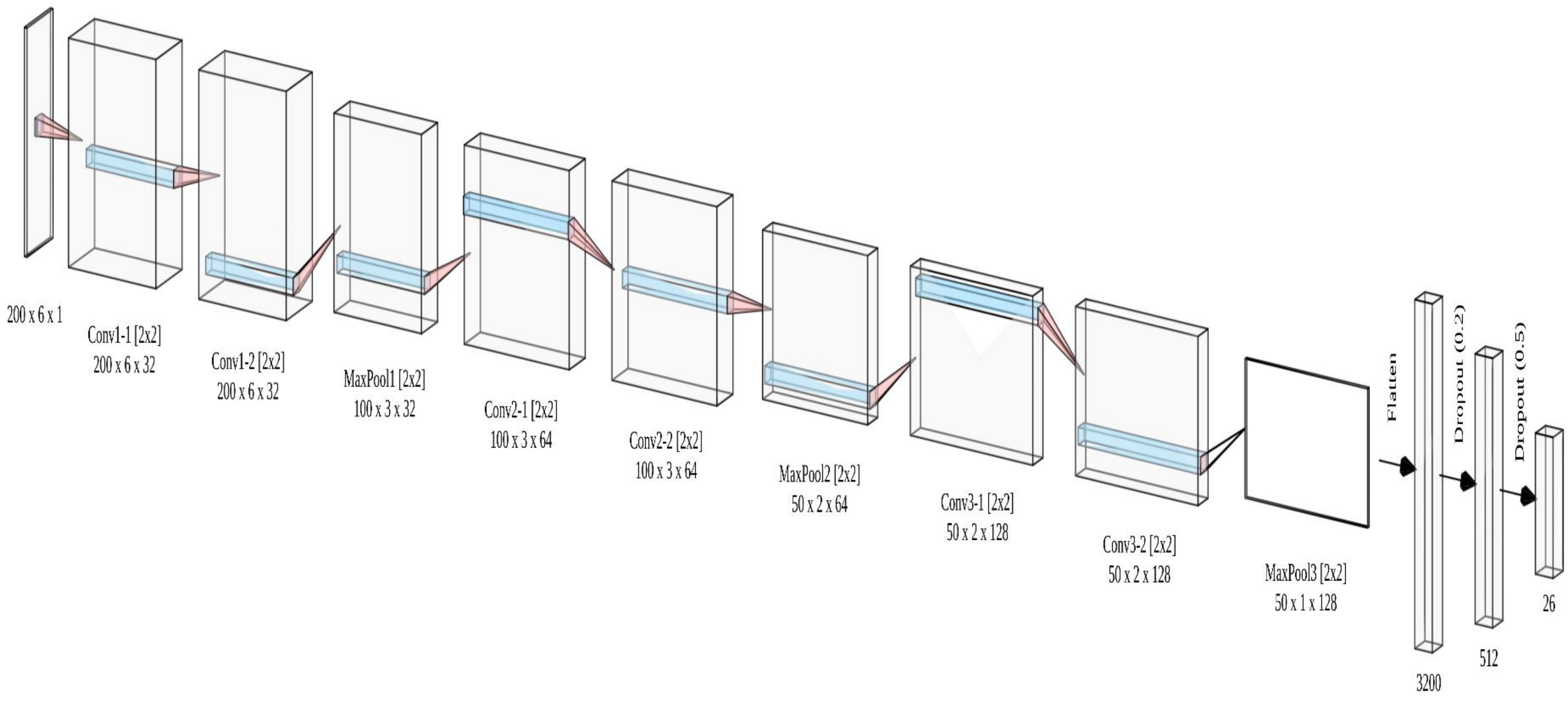


Figure 1. Network architecture of the 2D-CNN model for human activity recognition

Methodology

- This is a widely used convention in CNN how we made the groups in our architecture. That is**, putting some convolutional layer, followed by some pooling layer, and dropout layer.**

- In the context of a convolutional neural network, a convolution is a linear operation that involves the **multiplication of a set of weights**. The CNN technique was designed for two-dimensional input, the multiplication is performed between an array of input data and a two-dimensional array of weights, called a filter or a kernel

- **The Dropout layer is a mask that nullifies the contribution of some neurons towards the next layer and leaves unmodified all others**. Dropout layers are important in training CNNs because they prevent overfitting on the training data. If they aren't present, the first batch of training samples influences the learning in a disproportionately high manner.

Methodology

Leaving One Subject Out Method

- We evaluate all our models using a cross-validation scheme based on **LOSO**, where for N subjects, N different models are trained. Each model is trained using data from all the subjects except one, which is then used for testing and computing the performance of the model.

- Thirteen models per classifier and modality will be evaluated.

Experimental Setup

- ❖ Experimented on Google Colaboratory
- ❖ 12 GB RAM, GPU Accelereation Facility
- ❖ Epochs: 50
- ❖ Learning Rate: 0.001
- ❖ Early stopping was used to halt the training process when no improvement occurs in the performance

Results

Subject	Makela et al.	Our Deep CNN Model
S1	49.50%	61.94%
S2	46.10%	68.39%
S3	41.50%	80.00%
S4	46.60%	64.52 %
S5	49.30%	71.61%
S6	51.80%	65.52 %
S7	62.30%	67.74 %
S8	56.70%	49.03 %
S9	39.20%	73.55%
S10	58.90%	83.87%
S11	56.00%	59.35%
S12	58.20%	61.94%
S13	62.10%	64.52%
Mean	**52.10%**	**67.08%**

Table: LOSO (per-subject) validation results

Results

Subject	Classification Accuracy
S1	61.94%
S2	69.38%
S3	83.23%
S4	74.19%
S5	69.38%
S6	89.66%
S7	93.55%
S8	65.16%
S9	74.84%
S10	83.87%
S11	66.45%
S12	71.61%
S13	77.42%
Mean	**75.43%**

Table: LOSO (per-subject) classification results with the left-out subject data

Performance Analysis

- They accuracies of the work of Makela et al. [16] and our work is pretty similar.

- This means that machine deep learning based approaches for real time activity recognition are reliable as well as machine learning based methods.

Performance Analysis

- ❖ The dataset we used to test our model is small. There are 13 users who performed 16 activities each. There are around 6000 signals per subject per activity. Even after using the hopping technique, i.e, overlapping the signals with a move forward of beginning of the corresponding signals of the previous time frame, there was around 2060 samples per subject per activity.

Conclusion

This presentation mainly focused on –

- ❖ Human Activity Recognition with IoT Sensor Data
- ❖ Segmentation of IoT Sensor Data
- ❖ Using CNN on Sensor Data
- ❖ LOSO approach for evaluation of model

Conclusion

- Data from many sensor placements and modalities are included in the dataset. Despite the relative simplicity of the method, our baseline results, which are based on a LOSO evaluation scheme, show that generalization to various people is feasible.
- We achieved a classification accuracy of 75.43% with our deep learning based approach, which is very close to the machine learning based approach.
- This proves the prospects of deep learning based approaches in construction site based activity recognition problems.

Future Works

Experimentation with Larger Datasets

- The dataset we used to test our model is small.
- Deep learning models generally learn well when large dataset is fed to the network. If large dataset is fed to the network, there are good number of variant data which helps the model to generalise the characteristics of the classes.
- We have mentioned that there are scarcity of dataset about activities which are performed in the working sites, let alone the construction sites. So attempts can be made to make large dataset in working sites or constructions sites.

Future Works

Experimentation with Deeper Models

- The model we have experimented with has three groups of sequential convolutional-convolutional-maxpooling-dropout layers. Then there is flatten layer, dense layer, dropout layer and another output dense layer, respectively. The model is, however, not big enough.
- Hence, deeper and larger models have always the possibility to perform better. There can be used deeper models, also with another conventions of making deep learning model can be followed rather than construction of our model.

Future Works

Combination of Machine Learning and Deep Learning Based Approaches

- Statistical features can be used to feed the model.
- Machine Learning based classification and deep learning based classification can be applied and there should be ensembled decision for better classification.

Published Papers

- Abir, F.A., **Siam, M.**, Sayeed, A., Hasan, M., Mehedi, A. and Shin, J., 2021. Deep Learning Based Air-Writing Recognition with the Choice of Proper Interpolation Technique. Sensors, 21(24), p.8407.
- Hasan, M.A.M., Al Abir, F., **Al Siam**, M. and Shin, J., 2022. Gait Recognition with Wearable Sensors using Modified Residual Block-based Lightweight CNN. IEEE Access.

References

[1] Gupta, P.; Dallas, T. Feature Selection and Activity Recognition System Using a Single Triaxial Accelerometer. IEEE Trans. Biomed. Eng. 2014, 61, 1780–1786.

[2] Kwapisz, J.; Weiss, G.; Moore, S. Activity Recognition Using Cell Phone Accelerometers. SIGKDD Explor. Newsl. 2011, 12, 74–82

[3Liu, J.; Zhong, L.; Wickramasuriya, J.; Vasudevan, V. uWave: Accelerometer-based personalized gesture recognition and its applications. Pervasive Mob. Comput. 2009, 5, 657–675 .

[4] Cheng, W.; Jhan, D. Triaxial Accelerometer-Based Fall Detection Method Using a Self-Constructing Cascade-AdaBoost-SVM Classifier. IEEE J. Biomed. Health Inform. 2013, 17, 411–419.

[5] Matic, A.; Osmani, V.; Mayora, O. Speech activity detection using accelerometer. In Proceedings of the 2012 IEEE Annual International Conference on Engineering in Medicine and Biology Society, San Diego, CA, USA, 28 August–1 September 2012; pp. 2112–2115.

References

[6] Lin, X.; Chen, Y.; Chang, X.W.; Liu, X.; Wang, X. Show: Smart handwriting on watches. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2018, 1, 1–23.

[7] Chen, Y.H.; Huang, C.H.; Syu, S.W.; Kuo, T.Y.; Su, P.C. Egocentric-View Fingertip Detection for Air Writing Based on Convolutional Neural Networks. Sensors 2021, 21, 4382.

[8] Kim, U.H.; Hwang, Y.; Lee, S.K.; Kim, J.H. Writing in The Air: Unconstrained Text Recognition from Finger Movement Using Spatio-Temporal Convolution. arXiv 2021, arXiv:2104.09021.

[9] Abir, F.A., Siam, M.A., Sayeed, A., Hasan, M.A.M. and Shin, J., 2021. Deep Learning Based Air-Writing Recognition with the Choice of Proper Interpolation Technique. *Sensors*, *21*(24), p.8407.

[10] Inoue, S.; Ueda, N.; Nohara, Y.; Nakashima, N. Mobile activity recognition for a whole day: Recognizing real nursing activities with big dataset. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Umeda, Japan, 7–11 September 2015; pp. 1269–1280.

References

[11] Alia, S.S.; Lago, P.; Takeda, S.; Adachi, K.; Benaissa, B.; Ahad, M.A.R.; Inoue, S. Summary of the cooking activity recognition challenge. In Human Activity Recognition Challenge; Springer: Berlin, Germany, 2021; pp. 1–13.

[12] Joshua, L.; Varghese, K. Accelerometer-based activity recognition in construction. J. Comput. Civ. Eng. 2011, 25, 370–379.

[13] Akhavian, R.; Behzadan, A.H. Smartphone-based construction workers' activity recognition and classification. Autom. Constr. 2016, 71, 198–209.

[14] Rashid, K.M.; Louis, J. Times-series data augmentation and deep learning for construction equipment activity recognition. Adv. Eng. Inform. 2019, 42, 100944.

[15] Sherafat, B.; Ahn, C.R.; Akhavian, R.; Behzadan, A.H.; Golparvar-Fard, M.; Kim, H.; Lee, Y.C.; Rashidi, A.; Azar, E.R. Automated methods for activity recognition of construction workers and equipment: State-of-the-art review. J. Constr. Eng. Manag. 2020, 146, 03120002.

References

[16] S.-M. Mäkela, A. Lämsä, J. S. Keränen, J. Liikka, J. Ronkainen, J. Peltola, J. Häikiö, S. Järvinen, and M. Bordallo López, "Introducing vtt-coniot: A realistic dataset for activity recognition of construction workers using imu devices," Sustainability, vol. 14, no. 1, p. 220, 2021.

Any
Query?