July 2018

Human Activity Recognition Based on Transfer Learning

Jinyong Pang
University of South Florida, jpang@mail.usf.edu

Human Activity Recognition Based on Transfer Learning

By

Jinyong Pang

A thesis submitted in partial fulfillment
of the requirements for the degree of Master of Art
Department of Mathematics & Statistics
College of Arts & Sciences
University of South Florida

Major Professor: Kandethody Ramachandran, Ph.D.
Examining Committee Member: Gangaram S. Ladde, Ph.D.
Seung-Yeop Lee, Ph.D.

Date of Approval:
Nov. 26, 2018

Keywords: Human Activity Recognition, Transfer Learning, Deep Learning,
Convolutional Neural Networks

Dedication

This master Thesis is dedicated to my parents and my major thesis supervisor.

Table of Contents

List of Tables

List of Figures

Abstract

Human activity recognition (HAR) based on time series data is the problem of classifying various patterns. Its widely applications in health care owns huge commercial benefit. With the increasing spread of smart devices, people have strong desires of customizing services or product adaptive to their features. Deep learning models could handle HAR tasks with a satisfied result. However, training a deep learning model has to consume lots of time and computation resource. Consequently, developing a HAR system effectively becomes a challenging task. In this study, we develop a solid HAR system using Convolutional Neural Network based on transfer learning, which can eliminate those barriers.

The advantage of CNN is its capability of extracting features from data. In this paper, we firstly propose a new CNN architecture to set up a solid HAR system. Trained by WISDM HAR dataset, our proposed CNN model performs well in predicting six human behaviors (sitting, standing, jogging, walking, walking upstairs and walking downstairs) with prediction precision 92.3%. Based on transfer learning theory, learning knowledge from a built pre-trained model is a great starting point to rapid develop a new HAR system with using a new HAR dataset since source dataset and target dataset are relevant. We freeze all parameters in pre-trained HAR system. Then by training CNN model with UCI HAR dataset, parameters and new connections are gradually fixed. A new HAR system based on target dataset is accomplished. Relevant tests in evaluating the performance of prediction precision carry out with common testing indexes.

The weighted prediction precision of recognizing six human activities is 94%. For detecting the new behavior, laying, corresponding prediction precision is the highest one, 99%, among all six test results.

1 Introduction

1.1 Description and Motivation

With many remarkable successes in the development of artificial neural networks technology and the improvement of computational performance, deep learning architectures, such as convolutional neural network (CNN) and recurrent neural network (RNN), demonstrating a powerful ability of extracting features from different types of information, are now playing an indispensable role in many fields of machine learning, including computer version (CV) and natural language processing (NLP). Practically, convenient to the usage of smart devices in daily life are frequently updates of these effective algorithms embedded in the core of each device.

One of these accessible applications in personal intellectual terminals is human activity recognition (HAR), capturing and classifying behavior patterns from time-series data collected by sensors, such as accelerometers and gyroscopes. Combining with deep learning architectures, smart terminals would be able to explorer presentative human-activity patterns and classify them from complex signals mixed with noises, which is very different from the way with the requirement of rich knowledge about different devices and its collection of signals. Therefore, deep learning network would eliminate the traditional dependence on handcrafted statistical features extraction.

In fact, considering various devices measure signals without using the same sensor, it is inevitable to construct many corresponding recognition systems to classify human-activity patterns according to different data collectors, which would be a growing cost with the rapid development and launches of brand-new sensors. The key to solve this issue is constructing a cross-device flexible and extensible human-activity classification model, transferring the original recognition system to a new environment. Transfer learning is a help tool to make this idea come true, using original existing knowledge to set up the solid and versatile human activity recognition architecture.

1.2 Proposed CNN Architecture

In our study, we built up one CNN model for solving human activity recognition as the pre-trained model in phase I which would be used to transfer learning in phase II. Consequently, designing the first CNN model is a key to construct a HAR system with high-quality performance. With many comparisons in experiments using CNN model in classifying imagines, we finally determine the structure suitable to our research. Configurations for CNN models in two phases display in Table 1, Table 1 (a) for phase I and Table 1 (b) for phase II, in which H. means the height of input data, Len. means the length of input data, Num. means the number of channels of input data, strides mean the pace of moving windows (filters), and padding means how many pads would be added to the input data for taking size-unified data in very layers.

Table 1: Structure of proposed CNN architecture in this study

(a) CNN architecture in pre-trained Model (Phase I)

No.	Layers	Activation function	Input			Filter			Strides	Padding
			H.	Len.	Num.	H	Len.	Num.		
1	1d Convo Net	Leak ReLU	1	200	3	1	2	18	1	1
	MaxPooling	-	1	200	18	1	2	18	2	-
2	1d Convo Net	Leak ReLU	1	100	18	1	2	36	1	1
	MaxPooling	-	1	100	36	1	2	36	2	-
3	1d Convo Net	Leak ReLU	1	50	36	1	2	36	1	1
	1d Convo Net	Leak ReLU	1	50	36	1	1	36	1	-
4	1d Convo Net	Leak ReLU	1	50	72	1	2	144	2	-
	MaxPooling	-	1	25	144	1	2	144	2	1
5	Flatten & Dropout	-	1	12	144	Drop_probability=0.5				
6	Prediction	-	-	-	-	Number_of_objects=6				

(b) CNN architecture in Transfer Learning (Phase II)

No.	Layers	Activation function	Input			Filter			Strides	Padding
			H	Len.	Num.	H	Len.	Num.		
1	1d Convo Net	Leak ReLU	1	128	6	1	2	18	1	1
	MaxPooling	-	1	128	18	1	2	18	2	-
2	1d Convo Net	Leak ReLU	1	64	18	1	2	36	1	1
	MaxPooling	-	1	64	36	1	2	36	2	-
3	1d Convo Net	Leak ReLU	1	32	36	1	2	36	1	1
	1d Convo Net	Leak ReLU	1	32	36	1	1	36	1	-
4	1d Convo Net	Leak ReLU	1	32	72	1	2	144	2	-
	MaxPooling	-	1	16	144	1	2	144	2	-
5	Flatten & Dropout	-	1	8	144	Drop_probability=0.5				
6	Prediction	-	-	-	-	Number_of_objects=6				

Our motivation for this proposed CNN architecture is based on the consideration of mimicking the process of handcrafted statistical features extraction. What we should notice is the third part including two layers in our proposed CNN architecture. Introducing 1×1 filters to the middle of the model is for adding more combinations of different features extracted from the first two layers. Additionally, Other parameters and configurations in this proposed CNN model will be introduced in section 5.2.

1.3 Transfer Learning Applied in HAR Task

In this study, we use proposed a 10-layer CNN architecture to construct a solid HAR system by using transfer learning theory. All experiments are divided into two parts: the first part is to build a pre-trained CNN model for dealing with the task of human activities recognition using a large dataset. And then, we freeze inner parameters and configurations in the first six layers of pre-trained model. By training this HAR system with new training data from WISDM dataset, reconstructing the rest parameters and connections from 7^{th} layers to 10^{th} layers and updating parameters in these layers of the CNN architecture, a new HAR system would be developed with the capability of capturing new statistical features from new data.

Relevant studies include technology review with introduction of concepts [1], theoretically transferring from low-level sensor data to high level sensor data [2] and cross-domain HAR system using transfer learning [3]. Previous study on using transfer learning to solve HAR task usually employed traditional machine learning methods, like SVM and kNN algorithm. Most of these study put handcrafted statistical features as a main part of input dataset, which would be a limitation for developing HAR model in commercial application due to its labor cost in extracting features from tons of dataset. According to previous studies in HAR task, both prediction precision and recognition accuracy did not satisfy practical needs. In our study, we managed to solve HAR task from these two aspects. Model training in our experiments only use original time series data as inputs. Finally, a higher prediction precision using our HAR system based on transfer learning demonstrates its solid performance.

1.4 Structure of Thesis

This thesis is organized as follows.

Chapter 1 mainly introduces background knowledge and generally outlines the skeleton of this research.

Chapter 2 discusses and explains relevant technology adopted in our study and model structure proposed in experiments as well as procedures.

Chapter 3 describes the process of building up a pre-trained HAR model using CNN Architecture and constructing HAR system based on transfer learning, and analysis the performance of new HAR system.

Chapter 4 makes a summary of this study and discusses future works in the field of HAR.

2 Literature Review on Human Activity Recognition

The task of human activity recognition (HAR) is to classify body gesture or motion, and then determine or predict states of action or behavior [6]. Its extensive applications, appearing in military health care, physical recovery from disability or injury and clinical deformity correction, are drawing more and more attention on the further development and exploitation from industry and academe. Especially, in public health care, with the pervasion of portable personal digital devices such as smart phones, intelligent watches and multi-media terminals, generating a great number of different types of chronic data, for instance, video recorders, photos streams and spatial-temporal logs, there will be the significant need for personal customization using human activity recognition.

2.1 Categories of Human Activity Recognition

Human activity recognition tasks can be divided in two classes, including space-time approaches in computer vision and sequential approaches in time series analysis.

In space-time approaches, the essential for recognizing human activities is to measure the similarity between two volumes in images. [101] proposed an approach the changes of shapes on a series of images, which is corresponding to a moving human being. Comparing the patches of volumes was also proposed at the same time [102]. A more solid approach of extracting features of volumes and matching them effectively is [103]. Latter, with the development of neural network applied in processing images, deep convolutional neural network was applied in action recognition based on learning semantic trajectory-pooled data from raw video [104]. Long-term recurrent convolutional networks have a significant advantage in solving visual recognition by its memory elements in each network layer [105]. The most helpful study on human activity recognition is interpreting activity from video to natural language by using deep leaning architectures [106]. All spatio-temporal feature-based approaches are with limitation that they cannot recognizing complex activities.

In sequential approaches, traditional statistical techniques are initially proposed for handling human activity recognition tasks. Using principle component analysis (PCA) based on singular value decompositions (SVD), Yacoob processed one input as a signals, with sequential statistical features extractions [107], indicating all behaviors are the linear combination with different weighted statistical features. Lublineraman et al. [108] also proposed a linear time invariant (LTI) system based on Fourier descriptors for learning features of dynamic changes. LTI model can also classify a new input with similar features, such as slow walk and fast walk. Hidden Markov models (HMMs) used to recognize human activity was presented by Yamato et al [109], who is inspired by its application in recognizing speech recognition with only two features, point and wave. Oliver et al. [110] introduced the coupled HMM (CHMM) for modeling HAR systems as an improvement for HMMs approach, in which this CHMM model is able to recognize complex human behaviors. Moreover, dynamic Bayesian networks (DBN) is a successful extension of HMMs [111], which could recognize the behavior of two interacting persons. DBN used more features from the orientation of each body parts for determining which features associated with aimed activity. Hierarchy approaches were designed for understanding human activity in a reasonable way. Nevatia et al. [112] created a features representation language to describe human activities, making the process of recognition in three levels of hierarchy. Other sequential approaches in solving HAR problems include a heuristic algorithm [113], Boltzman Machine [24] and Markov logic networks [68].

2.2 Human Activity Recognition with Visual Sensors

Although, there are many different sensors embedded in various equipment producing diverse data types or data structures, when analyzing human activity via all kinds of data, they are generally employed by two approaches, vision-based activity recognition and sensor based activity recognition. For vision-based activity recognition approaches, data sources usually are those facilities, such as monitors or infrared thermal cameras, which can capture image-based or video-based information of human actions and 3D-movement tracks in changing environment [5]. Therefore, many popular algorithms in computer vision perform effectively, for example, Deep Residual Network [7] and Deep Convolutional Neural Networks [4], in this scenario. Accordingly, with DCNN being proposed in ILSVRC2012, deep learning algorithms sparked the research of neural network applied in computer vision, and fueled more brand-new powerful computational tools launched, contributing to artificial intelligence.

2.3 Human Activity Recognition with Wearable Sensors

For sensor-based human activity recognition, information from sensors attached to humans called wearable sensors, are time series data [9]. Traditional models were built by machine learning approaches from extracting features to classifying and prediction activity patterns, in which Hidden Markov Model (HMM) [11] and Support Vector Machine (SVM) [10] are always more popular previously. Recently, deep learning algorithms, like Convolutional Neural Networks (CNNs), play an essential role in constructing human activity recognition models since its powerful learning ability would automatically have a comprehensive grasp of features from collected-data, completely different from previous procedures of data-processing with handcrafted features. Another important advantage of modeling with CNN is processing high-throughput sequences simultaneously, regardless of noises and different lengths in data. The performance for human activity recognition system using CNN is significantly successful in three public datasets with great robustness and high accuracy [8]. Statistically, human activity recognition task is a problem of classification. As mentioned before, CNN is an effective approach to train and construct HAR system based on the data collected from wearable sensors. However, it is unavoidable to meet the problem of high computation cost, time-consuming cost and large-scale labeling processes. The low-cost technology for new large datasets coming from different population is to construct a bridge from existing HAR system to new tasks instead of setting up a new system with much more resource, improving accuracy and adaption of original HAR system in performing new tasks. Transfer learning is becoming a valid access to this goal. The superiorities of transfer learning are obvious, including less time-cost in learning new tasks, less information required, and more versatile situations being handled effectively [12].

3 Deep Learning Algorithm

3.1 Overview and Architecture

This section covers an outline of Deep Learning and detailed literature review of Convolutional Neural Network, the basic algorithm on the first step in constructing HAR system

Machine learning is an important branch of Artificial Intelligence, in which Artificial neural network (ANN) is now a potential algorithm. Deep learning is a kind of artificial neural network models with large and deep architectures, yet different from traditional neural network in specific computation in each layers which consists of many neurons, computational units. These neurons make a summation of data or information from previous neurons via an operation of a non-linear function, simultaneously processing inputs and generating outputs sent to next neurons in the same layer. With a series of complicated computation in many layers in the middle of the neural network, the final layer will carry out a classification, regression or fitting.

In deep learning, the process of learning is an assignment of searching proper powers or weights making the neural network reach desired proposes. In order to learning with much more accuracy, deep learning is constructed based on plenty of neurons and layers as well as special connection fashions according to various practical problems.

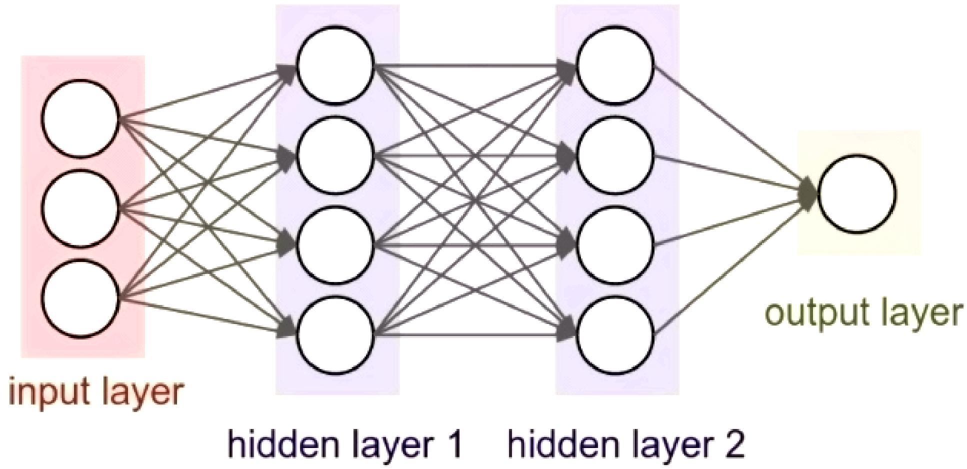


Figure 1: Artificial Neural Network Structure

In the 1940s, McCulloch & Pitts proposed the basic concepts on neural networks, which works for binary outputs from one neuron [13]. Latter, in the 1960s, inspired by cat's vision system, complicated cell was found in this system and stimulated the initiative in development of deep neural network [15]. In the late 1960s, the fact that one-layer network can solve the problems of classification with considerably limitations was discussed [14]. During the 1970s, there is no significant development of neural network.

With known approaches automatic differentiation [17] and, based on it, back propagation rules [20] being proposed, researchers can train multiple neural networks by gradient of a loss function according to the weights obtained from previous neutrons. Moreover, a new organization of neural network was created, including Hopfield networks, the cornerstone of Convolutional neural networks (CNN) [19,18]. However, other methods in machine learning such as Support Vector Machine (SVM) were introduced in solving the same problem in 1995 [22]. In 2006, vanishing gradient problem was solved by a fast learning algorithm for deep belief nets [23]. This research paper is a milestone in the development of artificial neural network making deeper and deeper networks training faster than before using a new technology, Restricted Boltzman Machine [24], on every layer. This excellent discovery open up new vistas of research for modern neural networks and deep learning algorithms, by which the performance in both supervised learning and unsupervised learning are robust in many fields, such as Alpha Go robot, Image classification competitions and autonomous-cars technologies.

3.2 Convolutional Neural Network (CNN)

Convolutional neural network (CNN), one of well-known deep learning structures, was an innovation inspired by cat's visual cortex system, overcoming the vanishing gradient problem and the problem of unconnected weights in each layer of neural networks [15]. The overall structure of CNN will be specified as following.

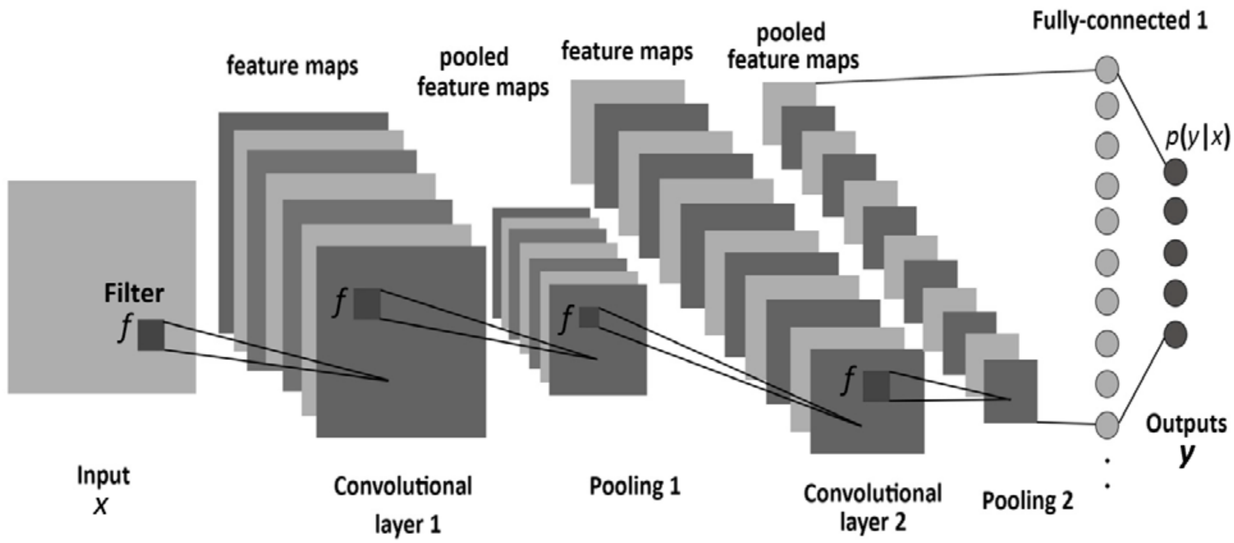


Figure 2: Convolutional Neural Network Architecture

The first layer of CNNs is a Convolutional layer. If matrix f is a convolutional filter or kernel and matrix X represents input data, the processing of convolutional computation is that filter *f* will be sliding along input data x with fixed stride, in which the operation of dot product is computed at each step and the output of each slide called feature map will be sent to next layer as input. As Figure 2 shows, each layer owns totally different filters with the same functions, lowering dimensions and extracting essential information.

More specifically, convolution operation originally generated from signals processing. Looking at the formula (1), convolution is a sum*s(t)* of a series of weighted values with weighting function *w(a)* in which weights changes with the variation of the value of (*t*-*a*) at the point of *t*.

$$s(t) = \int x(a)w(t-a)da. \tag{1}$$

A discrete convolution which is used in deep learning is as define:

$$s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a)w(t-a). \tag{2}$$

Generally, multiplication by a matrix is the essence in discrete convolution applied in computation of deep learning. The detailed process of this operation is shown as Figure 3, which is also the computation rule for a filter in convolutional layer. Filter *f* is a weight matrix sliding with fixed stride on the input matrix *X*, every time taking convolution operation, an element would be computed in output matrix *X'*.

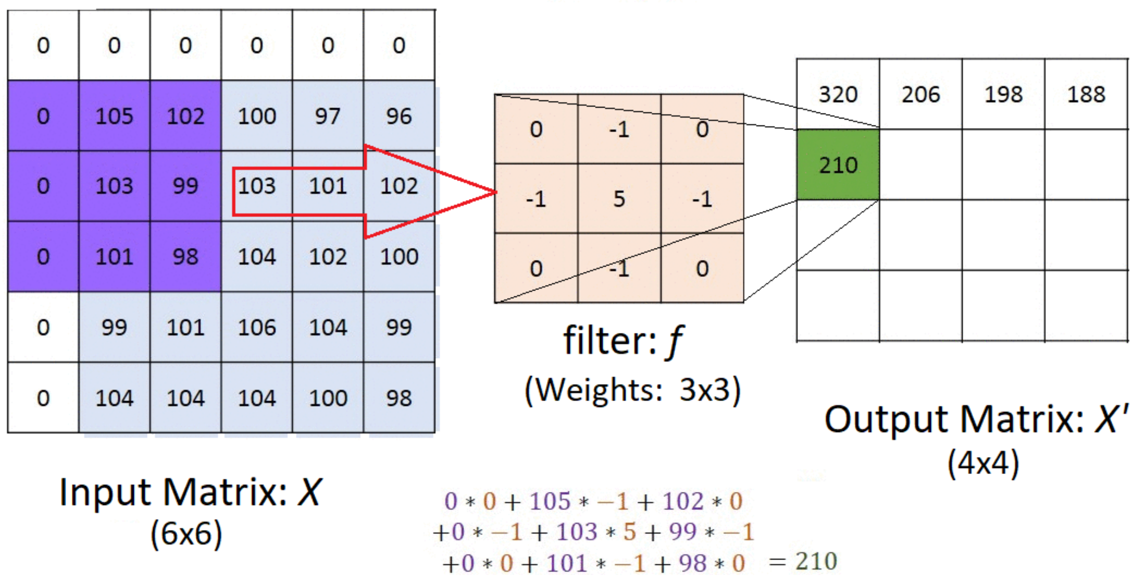


Figure 3: Operation of convolutional computation

Applying the filter onto the input matrix would change the size of the input. It is necessary to make a summary about this regularity. Suppose a volume of size $W \times H \times D$, other hyper-parameters are about filters including the number of filters *K*, the spatial extent *F*, the stride *S* and the amount of zero padding *P*. When this filter operating on that volume, the output volume is of size $W' \times H' \times D'$, where

$$W' = [(W - F + 2P)/S] + 1,\ \ H' = [(H - F - 2P)/S] + 1,\ \ D' = K\ .$$

When parameter sharing, the number of weights would be $(F \times F \times D) \times K$, the number of bias is *K*.

The aim for filter is to extract information and amplify these content with weights as well as eliminating useless information in the forward propagation. At the beginning of training model, the weights in each filter are initialized randomly. According the errors between target and output, there is a procedure adjusting these weights in different filters in each layers, called the back propagation which is aim to shrink regions of parameters from soft-max layer to hidden layer or ranges of weights in filters, in which all computations obey the chain rule for computing the derivative of the composition of two or more functions. That is one reason why training a deep learning model would waste too much time as well as computation resource.

The function of pooling layer is reducing the computation and the numbers of parameters in whole network, in another word, reducing dimensions. The general rule for pooling layer is to keep the maximum or compute the average in each sliding windows.

Generally, behind convolutional layer, the next is the activation layer (Rectified Linear layer), in which there are Rectified Linear Units with a nonlinear activation function in CNN structure. The most commonly used nonlinear activation function is ReLu, a simple thresholding operation. If ReLu function does not work well, Leaky ReLu and ELU function are better recommendations. This layer is indispensable since it can accelerate the convergence of whole neural network. Therefore, a good choice of nonlinear activation function would influence the performance of training neural networks. Figure 4 shows six popular activation functions and their function plots.

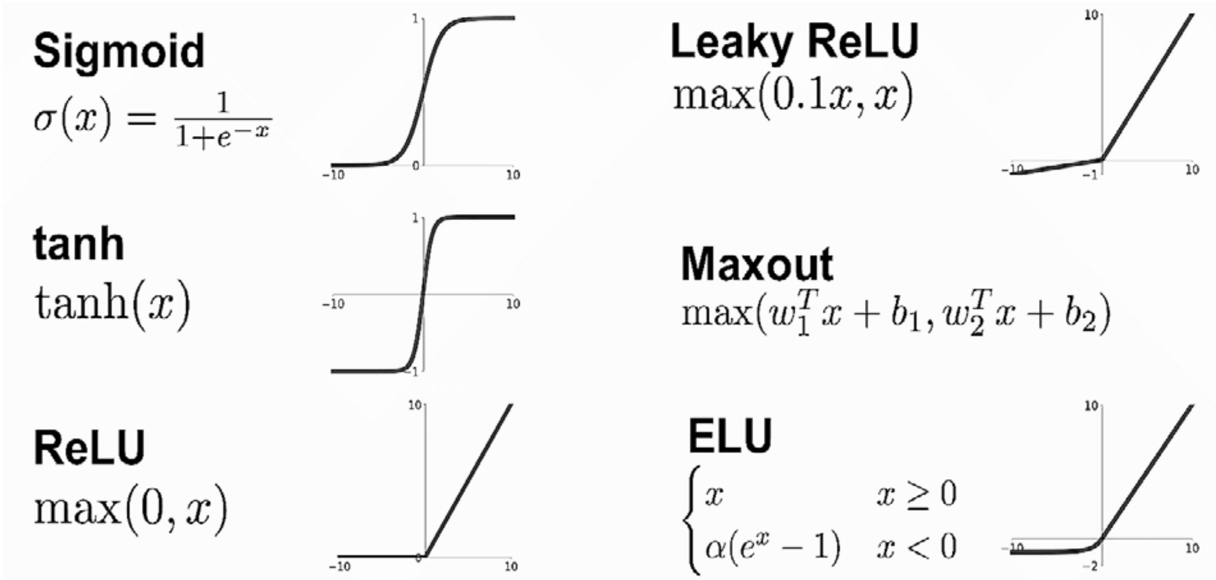


Figure 4: 6 Commonly used nonlinear activation functions in CNNs

Next layer in CNN is poling layer whose goal is reduce dimensions, and summarize or refine representative information and features. There are usually two approaches to achieve this step. The first one is to select the maximum from each sliding blocks along input data, another one is averages.

Before entering the final layer of CNN, it is a fully connected layer. All outputs from previous layers will be flattened into a one-dimensional vector *y* for classification as Figure 2 shows. And then, the final layer, computing probability *p(y|x)* corresponding to each class for predicting, is soft-max layer, which maps a length-p vector of real values to a length-*K* vector of values using a logistic function.

4 Transfer Leaning Technique

Transfer learning is a very helpful tool in this study, improving the performance of HAR system. In this section, we make a short but detailed introduction of transfer learning and its application in HAR.

4.1 Introduction

Learning new knowledge and skills is one of the most important capabilities for human beings. Based on personal own studying experience and pervious knowledge stored in brain, we are able to learn similar knowledge in a simplified way, without studying it from the beginning. For instance, learning how to ride a bicycle would help us learn how to ride a motor cycle, and learning how to use assembly language would accelerate us to learn scripting languages, like Ruby, Python and Lua. This is the study on how human beings learn new knowledge by individual way to transfer information preprocessed before to learn similar new information [8].

In the field of machine learning, introducing transfer learning algorithm would make a breakthrough on the common presumption that a training dataset must be of the same source as a future testing dataset, indicating two datasets are identically distribution. For those latter datasets collected from different distributions with various features in similar tasks, transfer learning could prepare traditional machine learning algorithms to have a great grasp of new knowledge from future dataset from another distribution by reusing previous preprocessed information. This is an essential function for machine learning based on transfer learning, reducing the cost of labelling new data, retraining new model and computational resource.

In history, transfer learning started from the study of multiple tasks learning models [9], focusing on learning common or latent statistical features from both source and target tasks in multitask. A clearer definition of transfer learning was from The Defense Advanced Research Projects Agency's Information Processing Technology Office that absorbing the knowledge from single or multiple informative source tasks and employing the valid information to an aimed target task is called transfer learning. Different from former concepts, the new definition concentrated more on target tasks without limitation of discovering common features from multiple tasks.

4.2 Definition

Before mathematically defining transfer learning, we firstly introduce the concepts of domain, task and dataset given by Pan and Yang in 2010 [10].

Definition 1 (**Domain** [10]) *A domain is defined as* $D = \{X, P(x)\}$*, where X is feature space and P(x) is marginal probability distribution,* $x \in X$.

Definition 2 (**Task** [10]) *Given a specific domain, a task is* $T = \{Y, f(x)\}$*, where y is a label space and f(x) is a predictive function, the conditional distribution of P(y|X),* $y \in Y$.

Definition 3 (**Dataset** [11]) *A dataset is defined as* $S = \{X, P(x), Y, f(x)\}$*, a set of data from a specific domain with a specific task.*

Accordingly, the definition of transfer learning is defined as following.

Definition 4 (**Transfer learning** [10]) *Given a source domain D_S and learning task T_S, a target domain D_T and learning task T_T, transfer learning improve the target predictive function $f_T(\cdot)$ in a target domain D_T and learning task T_S, where $D_S \neq D_T$, or $T_S \neq T_T$.*

4.3 Categories of Transfer Learning

In which situation should we use transfer learning? As mentioned in this section, transfer learning can employ the knowledge from a small labeled dataset to a new unlabeled dataset, in which two datasets are related in similar fields. There is no need to make a transfer learning if no relationships between learnt knowledge and unrelated fields. If making a transfer between them, negative transfer learning would be carried out with a bad efficiency, which is a task without any practical benefit. Following Table clarify the differences between traditional machine learning and transfer learning.

Table 2: Differences of application situations (Machine Learning & Transfer Learning)

<table>
<tr><th colspan="2">Learning Type</th><th>Source & Target Domains</th><th>Source & Target Tasks</th></tr>
<tr><td colspan="2">Traditional Machine Learning</td><td>The same</td><td>The same</td></tr>
<tr><td rowspan="3">Transfer Learning</td><td>Inductive Transfer Learning</td><td>The same</td><td>Different but related</td></tr>
<tr><td>Unsupervised Transfer Learning</td><td>Different but related</td><td>Different but related</td></tr>
<tr><td>Transductive Transfer Learning</td><td>Different but related</td><td>The same</td></tr>
</table>

More specifically, if there are relationship between feature spaces of two domains, that means the source and target domains are relevant. Generally, according to the different situation between source and target domains and tasks, transfer leaning would be categorized in three types, including inductive transfer learning, transductive transfer learning and unsupervised transfer learning [74].

In the inductive transfer learning, target tasks and source tasks are different regardless of the relationship between source and target domains. Predictive function would be constructed by labeled data in the target domain. Further categories in this situation are two classes. The first class of inductive transfer learning are of similar functions of multitask learning, with labeled data in source domain, learning features from source and target dataset at same time. Another class is without labeled data in source domain, which means between source and target domains the label spaces could be different, lead marginal information of source domain to be useless [75]. In transductive transfer learning, source and target tasks are the same, but source and target domains are different. In this case, lots of labeled data are used in source domain but no labeled data are employed in target domain. Two classes of transductive transfer learning are categorized in aspect of feature spaces. The first class is source and target domains have different feature spaces. Another one is they have the same spaces of feature but different marginal distribution of the input as training dataset, which is mostly applied in the fields of nature language processing using transfer learning in solving covariate shift problem [76, 77]. The unsupervised transfer learning, source and target task are similar but not totally the same, which focuses on handling the problem of dimensionality reduction and density estimation [78, 79]. Certainly, in this situation, labeled data are not available in both source and target domain as training data.

Based on the approaches applied in transfer learning models, they can be briefly classified in four types. The first one is instance transfer, which reweighted source domain would be used in target domain by importance sampling and instance reweighting [80, 78, 82, 83, 84, 85, 86, 87]. The second approach is feature representations' transferring, which transfers knowledge from across domains encoded into the learned feature representation and applies them on the new feature representation of target task [81, 88, 89, 90, 91, 92]. The third approach applied in transfer learning is called parameters' transfer learning, apparently, which regard those parameters or hyper-parameters in pre-trained model as transferable across tasks [94, 95, 96, 97, 98]. The final approach is based on the relationship transfer learning [99], transferring the relationship between two datasets, where statistical relational learning techniques is the main tool in machine learning [100]. Transfer learning right now has been applied in small-scale tasks in classification, such as image transfer learning, video classification and text clustering.

The goal of transfer learning is to transfer knowledge between related source and target domains [26]. In other words, transfer learning can employ knowledge leaned from original source material to another similar material in new environment, extending models' adaptation and application in various connected data as well as tasks with the same desires.

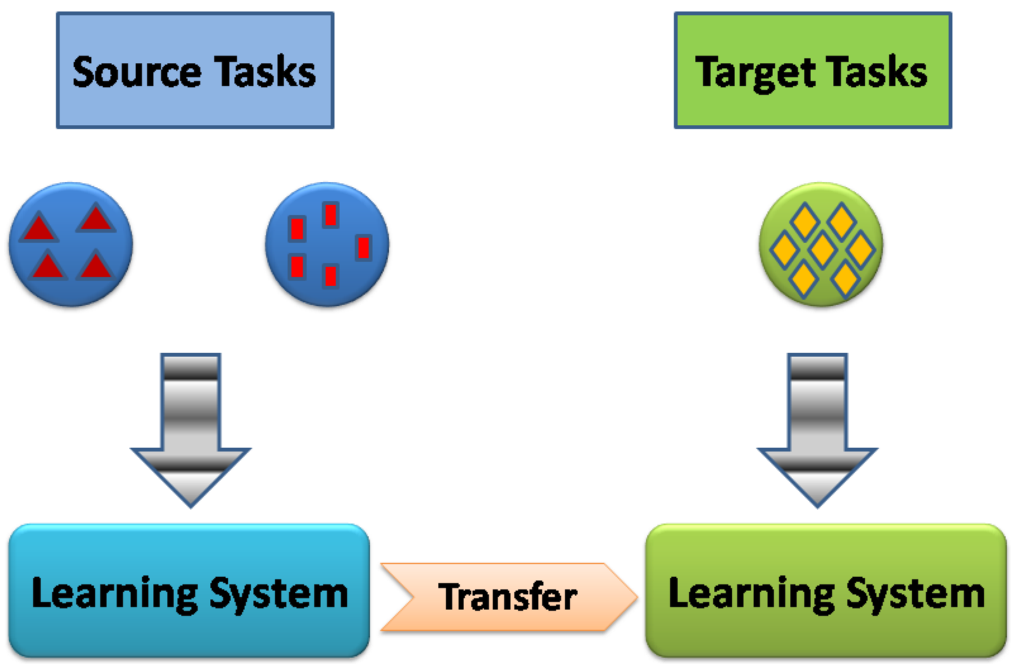


Figure 5: Transfer Learning Architecture

4.4 Contribution

The main contributions in this study are the following:

1. We present HAR system using CNN architecture based on transfer learning: a deep learning framework that is able to capture features of six human behaviors (WALKING, WALKING UPSTAIRS, WALKING DOWNSTAIRS, SITTING, STANDING & LYING) from time series data. Compared with other traditional algorithms in model evaluation, our HAR system is more dependable.
2. We demonstrate Convolutional Neural Network is more suitable to analyze sequence data like time series data in HAR task and it is more adaptive for modeling based on Transfer learning
3. We also prove that, in HAR task, transfer learning could personalize the assignment of recognizing behaviors in a short time with a higher prediction precision. Based on transfer learning, a new HAR system would be developed without too much cost of time, computation resource and additional data processing.

5 Experiment and Results Analysis

In this section, a Human activity recognition model using proposed Convolutional Neural Network is built up as an original pre-trained model. Comparing the performance with other three machine learning algorithms on the same dataset displays advantages of our model. Based on transfer learning theories, the experiments of the new HAR system demonstrates its strengths with different evaluation indices.

5.1 Datasets

In this paper, we construct the pre-trained HAR model using WISDM dataset [30] and develop the HAR system in transfer learning with UCI HAR dataset [28].

5.1.1 WISDM Dataset

WISDM dataset [30], collected from Wireless Sensor Data Mining laboratory (WISDM), also measured tri-axial acceleration time series data in three different directions by using smart phone. 46 volunteers contributed 1098209 samples with sampling frequency of 20 Hz, recording totally around 915 minutes. This dataset contains six various behaviors, including WALKING, WALKING UPSTAIRS, WALKING DOWNSTAIRS, SITTING, STANDING and JOGGING. The corresponding proportions of six different activities are shown in Figure 6(a).

The default input in WISDM dataset is one sequence with 200 sampling points. Another part of this dataset are transformed from input space (with size 1097009×3) to feature space, the characteristic matrix (with size 5418×43). Based on our proposed structure of Convolutional Neural Network model, the pre-trained HAR system was created by training with 80% of sequences as the training set and testing model's performance with 20% of the sequences as the testing set. Original data as the input are not transformed into feature space. Furthermore, three traditional machine learning algorithms are employed in solving this classification problem as comparable models.

5.1.2 UCI HAR Dataset

UCI HAR dataset, offered by University of California Irvine, is one of the most popular open testing datasets in the research of Human activity recognition based on wearable sensors. 30 volunteers wearing the smart phone (Samsung Galaxy S II) assembled with its accelerometer and gyroscope participated the data collection. The original data contain 3-axial linear acceleration time series data and 3-axial angular velocity time series data at a constant frequency of 50 Hz, including six types of activities: WALKING, WALKING_UPSTAIRS, WALKING-DOWNSTAIRS, SITTING, STANDING and LAYING. More details of the proportion of six activities is shown in Figure 6 (b).

All time series data labeled with activities are divided into segments by a sliding window with a fixed length of 128 sample points corresponding to a size of 2.56 seconds, and preprocessed with 561engineered features (eigenvalues). Based on our proposed construction of this Convolutional Neural Network, for final run, the data are randomly divided into 70% and 30% of totally 10299 labeled samples, respectively as a training set and a test set. This dataset is employed in constructing new HAR system based on transfer learning, by training the parameters in latter layers in pre-trained HAR system and extracting highly abstractive features of new objects. We also report the accuracy and classification errors on the test set.

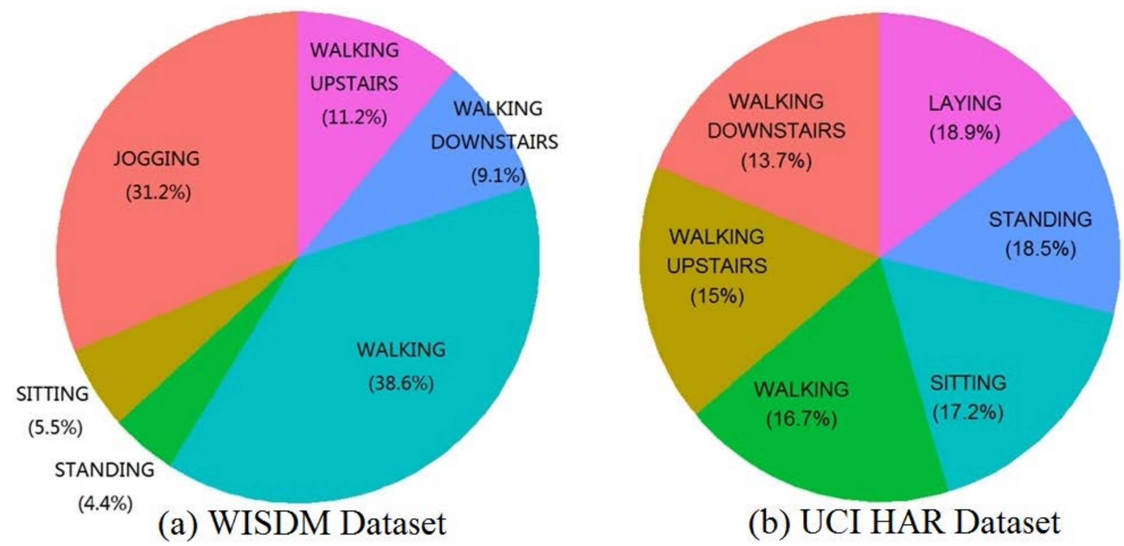


Figure 6: Percentage of activities in two datasets

5.1.3 Similarity between WISDM dataset and UCI HAR dataset

One of essential assumptions in machine learning that training datasets and other future datasets come from the same feature space with one identical distribution is accented in transfer learning, that source and target domains are also from the similar distribution. Consequently, before performing the task of transfer learning, examining the similarity between source and target domains is a key step to guarantee the transfer learning model works well. In this section, an experiment for measuring similarity in time series data is carried out between two HAR datasets.

In time series data mining, one of the most efficient algorithms in examining similarity of two pieces of time series data is dynamic time warping (DTW), initially proposed by Berndt and Clifford (1994) [67], which is widely applied in word speech recognition and temporal signal processing. Using this DTW, we are able to measure similarities from those temporal sequences covering information of person's walking in different paces by warping time axis.

Suppose there are two temporal sequence W and U from two HAR dataset of default size as inputs, n and m respectively [66, 69, 70].

$$W = w_1, w_2, ..., w_i, ..., w_n$$

$$U = u_1, u_2, ..., u_j, ..., u_m$$

Then arranging these two time series on the sides of a grid or a matrix by the size of n× m. Here, using Euclidean distance as the measurement of distance $d(w_i, u_j)$ in this matrix between any two points w_i and u_j, we define

$$P = p_1, p_2, ..., p_k, ..., p_K, \max(n, m) \le K \le m + n - 1$$

as warping path, a mapping between two temporal sequences, meeting the following conditions.

(i) Boundary Conditions: $p_1=(1,1)$ and $p_k=(n, m)$ are the star and the end of this warping path aligned on the diagonal opposite corner elements of the grid or the matrix.

(ii) Monotonicity Condition: $n_1 \le n_2 \le ... \le n_L$ and $m_1 \le m_2 \le ... \le m_L$ requires points or elements in the warping path P to monotonic temporal space.

(iii) Continuity Condition: $w_k=(a, b)$ then $w_{k-1}=(a', b')$, where $0 \le a-a'$ and $0 \le b-b'$, requiring the steps of the warping path should be neighboring elements.

Based on conditions mentioned above, the goal for us is to get an optimal warping path to evaluate the cumulative sum of distance $D(i, j)$ and minimize it with neighboring blocks: $D(i,j) = d(w_i, u_j) + \min\{D(i-1, j-1), D(i-1, j), D(i, j-1)\}$

The DTW algorithm could be summarized as following Table 3:

Table 3: A summary of DTW algorithm

Algorithm: Optimal Warping Path
Input: Accumulated cost matrix D **Output:** Optimal warping path P **Procedure:** The optimal path $P = p_1, p_2, ..., p_k, ..., p_K$, $\max(n,m) \le K \le m+n-1$ is computed in reverse order of the indices starting with $p_K=(N, M)$. Suppose $p_k=(n, m)$ is computed. In case $(n,m)=(1,1)$, one must have $k=1$ and we are done. Otherwise, $p_{K-1} := \begin{cases} (1, m-1), & if \;\; n=1 \\ (n-1, 1) & if \;\; m=1 \\ \arg\min\{D(n-1,m-1), D(n-1,m), D(n,m-1)\}, & otherwise \end{cases}$ where we take the smallest pair in case "argmin" is not unique.

The smaller cumulative distance means the strong similarity of two different time series data.

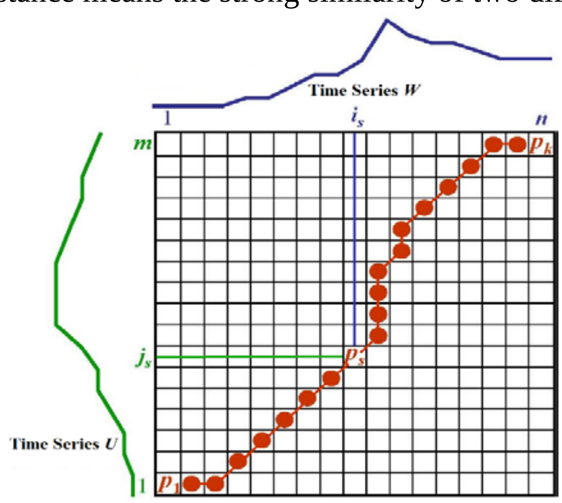
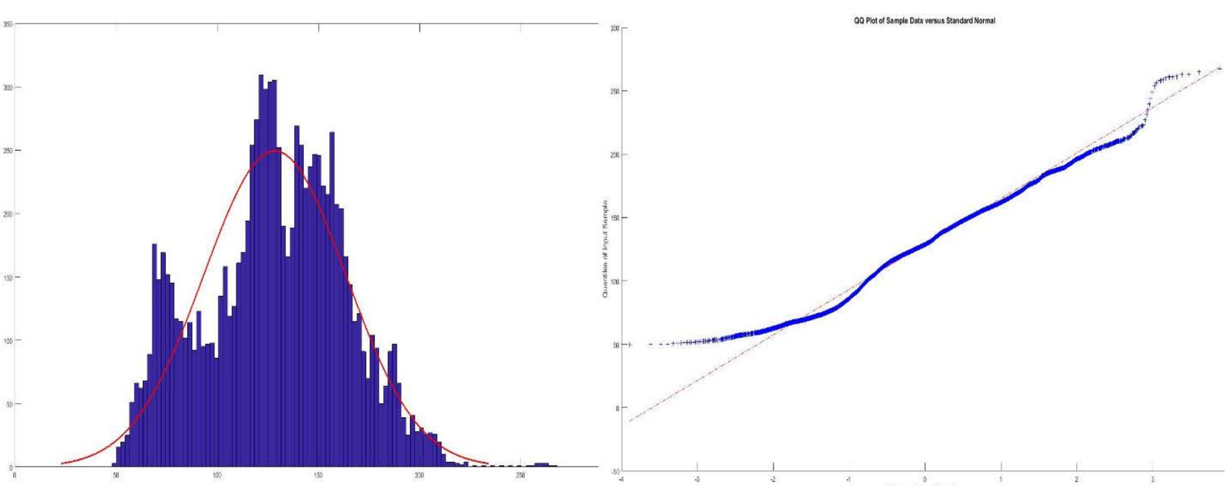


Figure 7: The warping path constructed by two temporal sequences

In our experiment, we measure the similarities of six different human activities (WALKING, WALKING UPSTAIRS, WALKING DOWNSTAIRS, SITTING, STANDING, JOGGING) in three axis x, y, z between two HAR datasets with 10,000 trials of randomly selected samples for each axis of each label with the default sequence lengths, 200 points for WISDM dataset and 128 points for UCI dataset. After getting 10,000 cumulative distances for each experiment, implying the difference between each group of two sequences with same label from separate dataset, we centralized 10,000 measurements of sequence of differences each time from their group averages to 0, approximately regarding them as a large sample from the standard normal distributions with a mean of 0 and a standard deviation of 1. Then two tailed Z-tests are employed under the null hypothesis that $\mu = \mu_0 = 0$ at critical point $\alpha = 0.01$, assuming each two subgroups of time series data with the same label from two datasets are from the same normal distribution. The following table is the table of Z scores of all experiments for measuring possible similarities between five labels.



(a) Normality fitting (b) QQ plot of 10,000 sampling of sitting label

Figure 8: similarity tests of sitting behavior (x-axis) between two datasets

Table 4: Z-scores for the distributions of similarity measurements

W \ U	Walking	Upstairs	Downstairs	Sitting	Standing
Walking	1.5632 1.5284 1.4276	2.1720 2.1324 2.1964	2.1892 2.1664 2.1636	2.6256 2.6148 2.6140	2.6232 2.6084 2.6180
Upstairs		1.7852 1.7640 1.8392	2.2776 2.2284 2.2260	2.5984 2.5908 2.6060	2.6708 2.6740 2.6784
Downstairs			1.8508 1.8380 1.9092	3.0924 3.0912 3.0896	2.6764 2.6976 2.6952
Sitting				1.1588 1.1476 1.1528	3.3676 3.0932 3.1372
Standing					1.1708 1.1660 1.1696

The yellow part implies that at the critical point $\alpha = 0.01$, the differences between each two labels are significant. In other part, the similarities for each two of labels are significant. For each comparison of two different datasets with the same label, they are of similarities statistically. Because of these conclusions of this experiments, it is reasonable to conclude that two dataset are from similar features space or distribution.

5.2 Experiment Configuration

We firstly set a series of parameters shown in Table 5, constructing and training proposed CNN model as a pre-trained HAR system, and then building a new HAR system by using transfer learning.

Table 5: Parameters setting in training Convolutional Neural Network models

Parameters	***Pre-trained CNN Model***	***Transfer Learning Model***
Batch size	800	600
Sequence length	200	128
Number of channels	3	6
Number of labels	6	6 (with 1 new label)
Learning rate	0.00025	0.0001
Training batch (epochs)	1000	800
Optimizer	Adam algorithm	

In the first phase, SVM, kNN (k=5) and Decision tree algorithm are employed in comparing the performance by building various HAR system with CNN model. For these three algorithms, the percentage of training set is 70% of original data and the rest 30% of the raw data consist of test set.

Moreover, in this paper all experiments are carried out by using Tensorflow online platform, CoLab, a free product for improving Deep Learning algorithm. Linux Ubuntu Operation system is the running environment for coding, and Python 3.6.6 as well Shell is the programming language in implementing deep learning algorithm and developing HAR system based on transfer learning.

5.3 Experiment Evaluation

In order to evaluate the performance of HAR models using different algorithms, there are generally four indices including Total Test Accuracy (Acc.), Precision (P), Recall Value (Rec.) and F1 Score (F) used for examining the performance of machine learning algorithms. Here in this paper, Total Test Accuracy (Acc.), Recall Value (Rec.) and F1 Score (F) are adopted in evaluating the performance of HAR models.

First, it is necessary to introduce several simple concepts since they are useful to make those evaluating indices measurable. The positive case is the class of our interest, so an opposite one is negative. The positive one called True Positive (TP), is predicted as positive, while the positive one predicted as negative is called False Negative (FN). The negative case predicted as negative is called True Negative (TN), and the negative one predicted as positive is called False Positive (FP). All four indices are computed by the four simple concepts as shown in Table 4.

Table 6: Formulas of four indices for evaluation

Index Name (abbr.)	**Formula**
Total Test Accuracy (***Acc***.)	***Acc*** = (TP+TN)/(TP+FN+FP+TN)
Recall Value (***Rec.***)	***Rec*** = TP/(TP+FN)
Precision (***P***)	***P*** = TP/(TP+FP)
F1 Score (***F***)	***F*** = 2P*Rec/(P+Rec)

5.3.1 Phase I: HAR System using CNN Architecture

During constructing CNN model with WIDSM dataset for solving the problem of Human activity recognition, we simultaneously model with the same dataset using other solid algorithms including SVM, kNN (k=5) and Decision tree algorithm. Finally, we make a summary, considering their performance in modeling as shown in Table 7.

Table 7: The Overall Performance of four algorithms on WIDSM Dataset

Algorithm	Precision	Recall	F1_score
SVM	0.835	0.847	0.841
kNN	0.782	0.779	0.780
Decision Tree	0.846	0.863	0.854
CNN	**0.923**	**0.931**	**0.927**

According to the result of model-evaluation, it is obvious to find that CNN algorithm applied in constructing Human Activity Recognition system is of best performance through three indices. Furthermore, CNN model obtain high scores not only on overall evaluations, but also on each of six predictable objects as shown in Figure 9 compared with other three algorithms with the comprehensive examining index *F* score. Comparing with traditional machine learning algorithms, CNN algorithm is a solid method to construct a data-driven model of classification and prediction due to its higher performance and potential adaptation in transfer learning.

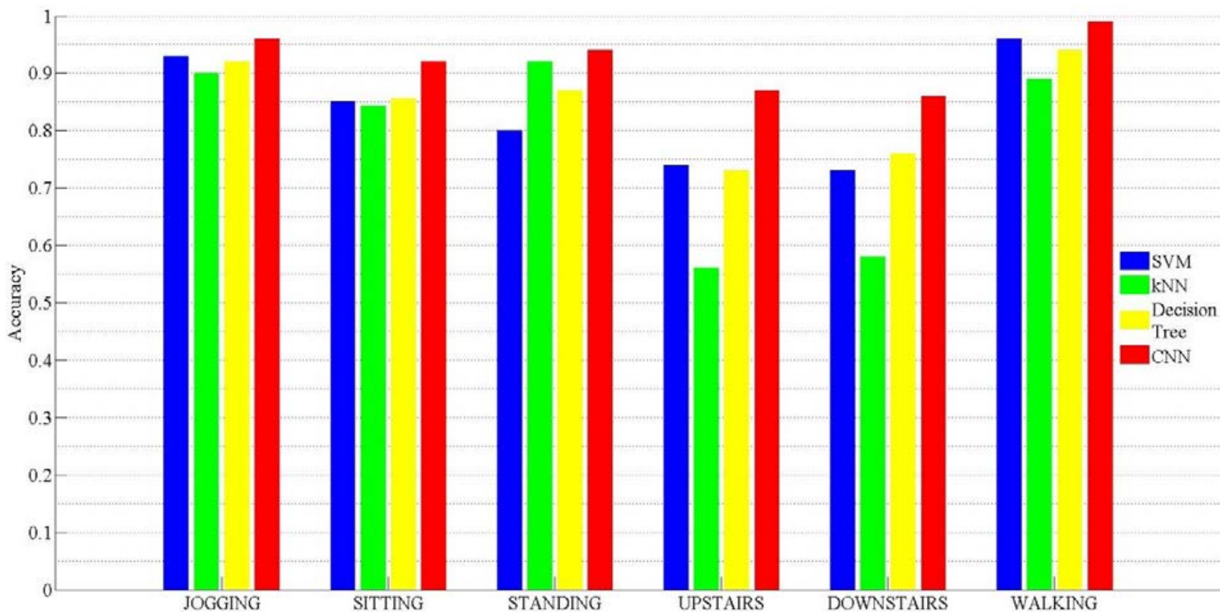


Figure 9: F1 score of four algorithms on different human activities recognition

According to the comparison, CNN gains the highest score among all algorithms on each activity prediction. For more details about CNN model's performance, the classification confusion matrix based on test dataset is shown as Table 8, providing us information on how well the HAR system did for each class in prediction.

Table 8: HAR classification Confusion Matrix

Tr \ P	JOG	SIT	STAND	UPSTAIRS	DOWNSTAIRS	WALK
JOG	**65590**	0	0	1164	472	1164
SIT	169	**10915**	350	410	57	0
STAND	73	73	**8887**	144	222	73
UPSTAIRS	713	130	63	**21205**	1500	713
DOWNSTAIRS	189	125	0	1579	**16991**	879
WALK	68	0	0	847	271	**83735**

Different from previous studies in which test dataset used in evaluating the performance of models is of a small amount, in this study, a large number of test cases could comprehensively measure the quality of operating HAR system. Although there exist prediction biases in each class, the overall prediction precision of HAR system based on CNN algorithm is up to 0.923.

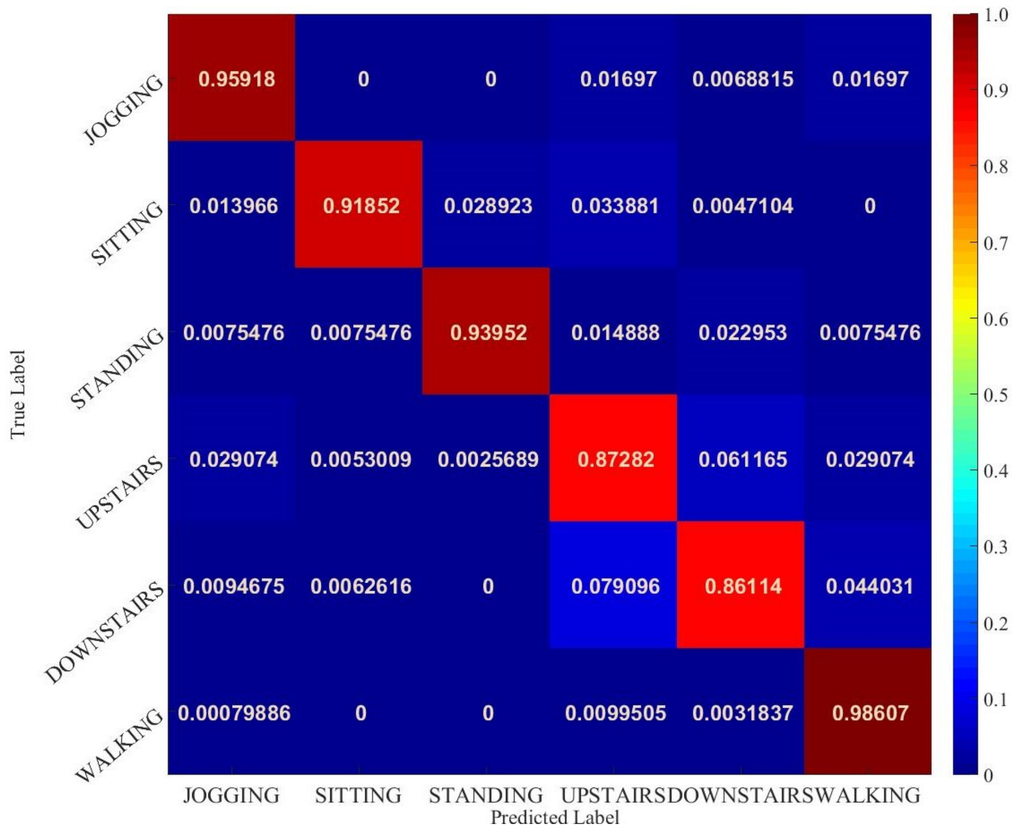


Figure 10: Heat map of HAR classification Confusion Matrix on test dataset

For getting an intuitive sense of testing precision, the heat map of tests' result shown in Figure 10 demonstrates several characteristics of HAR system. The prediction biases in each class of this model is significantly low. The regions of light red and light blue, including UPSTAIRS and DOWNSTAIRS, reminds us that the classification and prediction of these two activity in this model is not highly clear-cut. In another word, the classification of UPSTAIRS and DOWNSTAIRS produce most prediction biases (more than 0.05), impacting the overall performance of this HAR system. However, predictions in other classes are of higher quality, especially, JOGGING and WALKING, with prediction accuracy higher than 0.95.

The rest of tiny prediction biases directly perform the robustness of this model due to the impossibility of eliminating irregular noise existing in original time series data. Consequently, comparing previous studies using lots of engineering features extracted from original data as inputs for reducing the impact of noise, we trained our HAR system directly using raw time series data with fixed segmentations so as to improve the tolerance of confused signal and complicated noise, as well as adjusting configurations of CNN architecture step by step.

In phase I, stable performance of HAR system using CNN algorithm satisfies our expectation of a pre-trained model. Based on this well-performed CNN architecture, in phase II, a new HAR system using transfer learning was carried out with UCI HAR dataset. The process of transfer learning is to store knowledge gained from solving a problem and apply this information to a different but similar task. In our study, pre-trained CNN architecture is regarded as an integration of experience and knowledge in solving HAR task. Updating parameters and weights during fine-tuning the CNN with a new training dataset, we get a new HAR system developed from pre-trained HAR system without a bunch of consumption in preprocessing dataset, computation resources and time.

5.3.2 Phase II: HAR System based on Transfer Learning

After developing a HAR system based on Transfer learning using CNN architecture, we summarize and analyze the performance of this brand-new model from the results of 2974 tests. According to the evaluation of HAR system using transfer learning as shown in Table 9, the overall prediction precision is up to 0.936. The index Recall, reflecting intuitively the ability of the classifier to find all the positive samples, is also higher the HAR system in phase I. A weighted harmonic mean of the precision and recall is F1 score (0.935), revealing a comprehensive accuracy in prediction.

Table 9: The prediction performance of six objects using Transfer Learning

	Precision	Recall	F1_Score	Support
WALKING	0.972	0.909	0.940	496
UPSTAIRS	0.946	0.892	0.918	471
DOWNSTAIRS	0.831	0.948	0.885	420
SITTING	0.973	0.892	0.931	491
STANDING	0.901	0.976	0.937	532
LAYING	0.998	0.993	0.995	537
Ave/Total	0.936	0.933	0.935	2947

For investigating more details of the transfer-learning model, confusion matrix of prediction on test dataset delivers a solid support to the overall accuracy that the performance of new HAR in learning and capturing the characteristics of new dataset becomes stronger without too much impact from tiny prediction biases and errors. Confusing classifications significantly reduced, comparing to the same result in pre-trained model which is shown in Table 10.

Table 10: HAR classification Confusion Matrix

	WALKING	UPSTAIRS	DOWNSTAIRS	SITTING	STANDING	LAYING
WALKING	451	8	37	0	0	0
UPSTAIRS	7	420	44	0	0	0
DOWNSTAIRS	5	16	398	0	1	0
SITTING	0	0	0	438	52	1
STANDING	1	0	0	12	519	0
LAYING	0	0	0	0	4	533

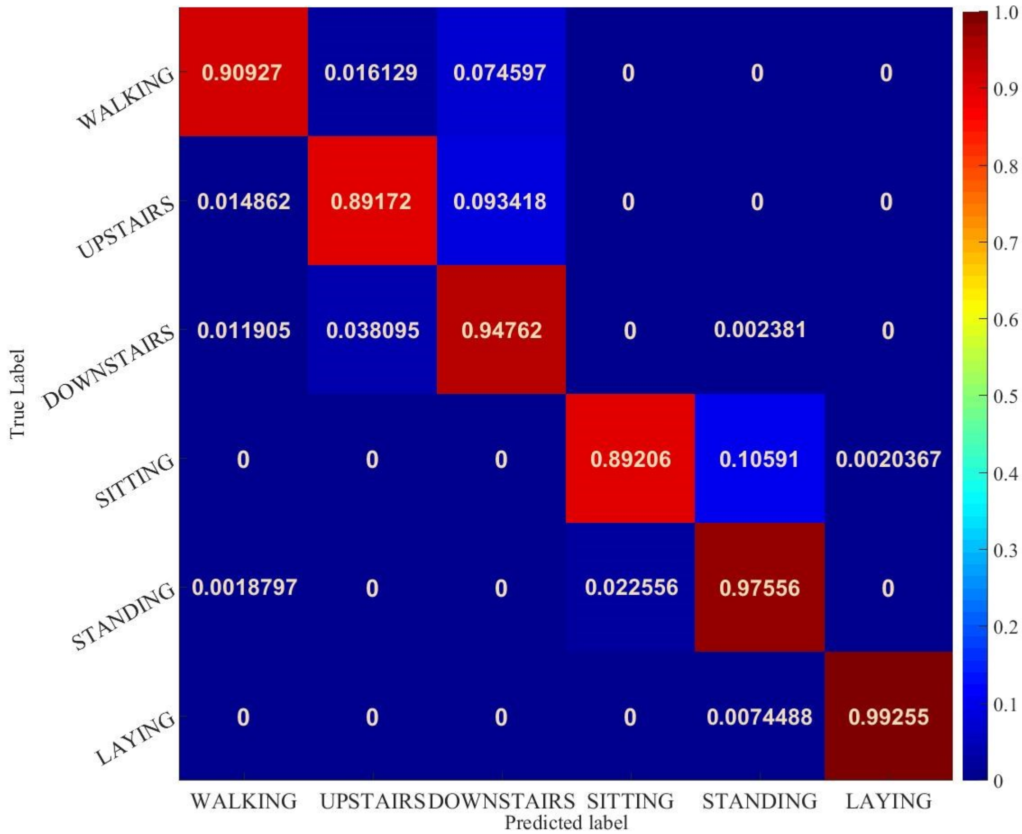


Figure 11: Heat map of new HAR classification Confusion Matrix

A data visualization of the performance of new HAR system relies on the confusion matrix with prediction precision shown in Figure 11. For the dataset with a new Label, *LAYING*, transfer learning model has a great grasp of new knowledge about its features. Accordingly, the prediction precision for *LAYING* becomes the highest one among all six objects, indicating that new system obtains strong capability of learning new similar knowledge by transferring original source information absorbed from pre-trained HAR system into the new system.

Another advantages in new system, as heat map told us, is that transfer learning clearly divide all objects into two abstract classes with a clear boundary between active behaviors (WALING, WALKING UPDTAIRS, WALKING DOWNSTAIRS) and sedentary behaviors (SITTING, STANDING, LAYING), automatically gaining a function in detecting if a human move or not. Especially, for recognizing the sedentary behaviors, the prediction precision is up to 0.95339 in average. Consequently, depending on this characteristics, it is reasonable to conclude that transfer of knowledge from source to target was carried out in training the new HAR system.

However, even though the prediction precision of six targets in UCI HAR dataset are higher than previous model, the lower accuracy of detecting the behavior, *WALKING UPSTAIRS*, still exists in new HAR system which means the confusion about recognizing WALKING UPSTAIRS and WALKING DOWNSTAIRS need to be controlled and then gradually reduced.

During the process of training CNN architectures, the time assumption in training new CNN architecture with UCI dataset is around 28 minutes. The overall and individual prediction precisions in new HAR system using transfer learning are better than those in pre-trained HAR system with WISDM dataset. The transfer learning reduced time assumption, furthermore improving model's performance in recognizing human activities by learning new relevant knowledge and information.

Comparing to the prediction precision in previous studies, in Kaggle competition of data science with the problem of human activity recognition, the best performance at the 1st rank (update on Jul.2018) with Precision, Recall and F1 score are 0.95, 0.94 and 0.94 respectively. The precisions of our HAR system based on transfer learning are 0.94, 0.93 and 0.94 correspondingly. Base on the balanced index F1 score, the weighted harmonic mean of the precision and recall, the performance of new HAR system in our transfer learning study is a competitive rival with three advantages including its lower consumption of time and computation, adaptation of new knowledge and robustness of its extensibility.

5.4 Importance and Contribution

In this study, based on our proposed CNN architecture, we accomplished solving the challenge of human activity recognition using transfer learning. With WISDM dataset and UCI HAR dataset collected from smartphones, a pre-trained CNN model are developed in phases I with satisfied performance in classification, and then a new HAR system, transferring knowledge from WISDM to UCI datasets during training pre-trained CNN architecture, operates with overall prediction precision 0.94.

Among the previous works that handle HAR tasks, the best model with an outstanding overall prediction precision of 0.9759 is [65], which using four-layer CNN architecture with SVM for feature selecting and reducing computational cost [65], called DCNN+. Removing the SVM part, the model called DCNN is of similar architecture as our CNN model. For illustrating the importance of our research, we compare the performance of HAR system based on transfer learning with the best model with two criteria. The first one is computation cost, which is defined as the number of parameters computed in one second. The second one is prediction accuracy. The performance comparison is shown in Table 11.

Table 11: Performance Comparison

	Computation Cost (*p/s*)	Accuracy
DCNN [65]	1.56	0.9518
DCNN+ [65]	3.85	0.9759
Transfer leaning Model	0.86	0.94

Although the accuracy in our HAR system is 96% of DCNN+ model, in aspect of efficiency, the computation cost is significantly lower than other methods mentioned in [65], which is satisfied our proposed expectation that reduce the cost of computation with a higher prediction accuracy.

Without using any preprocessed engineering features of data as input, the performance of HAR systems demonstrated that CNN is capable of extracting and learning useful information from time series data of six objects (with labels including WALING, WALKING UPDTAIRS, WALKING DOWNSTAIRS SITTING, STANDING, LAYING). By employing transfer learning in CNN model, the test result compared with previous studies is robust in improving prediction precision (from 0.923 to 0.94), and lowering biases of recognition with fewer tiny prediction errors.

6 Conclusion and Future Work

In this paper, we present a Convolutional Neural Network based on transfer learning in solving the problem of human activity recognition with using time series dataset. We firstly design a CNN architecture to simulating the process of statistical features extraction, and then validate this idea by developing a HAR system based on this proposed CNN model with WISDM dataset. All datasets in our research are collected from smart phone with fixed segmentations. This HAR system is capable to distinguish six different human activities (WALKING, WALKING UPSTAIRS, WALKING DOWNSTAIRS, SITTING, STANDING and JOGGING) Comparing the performance with three traditional algorithms in recognizing human behaviors with the same dataset, CNN architecture display its solid performance with an overall prediction precision 92.3%. Then, we develop the new HAR system based on transfer learning. Previous HAR system plays an essential role as a pre-trained model, in which all parameters and configurations are frozen. Transferring knowledge from WISDM dataset to UCI HAR dataset by frozen parameters in first six layers and re-constructing new connections as well as new configurations from the 7th to the 10th layers, the new HAR system captures new knowledge of features from new data, recognizes six human activities with a higher overall recognition accuracy 94% and also takes tolerance to biases from original data, reducing the prediction errors. It is reasonable to inference from experiment results that transfer learning is indispensable for a deep learning model to be versatile in similar tasks, for a comprehensive understanding of different patterns during model training is the key to solve the problem of classification. Admittedly, although the performance of our model is reliable in classifying human activities, there is always the prediction with errors between similar behaviors like walking downstairs and walking upstairs. Another challenge of transfer learning applied in HAR task is how to automatically recognize and predict point changes between different human behaviors, in another word, putting continuous time series data as input is more close to real situations. Consequently, finding a better structure of deep learning model would be a bright way to solve this new problem in HAR tasks

In future work, how to reduce the size of deep learning model, the training time and the number of redundant information or parameters is one of research tracks in human activity recognition. Designing a proper deep learning architecture for the usage of transfer learning and setting up a good evaluation system to examine the performance of HAR system is also inevitable, for right now, current deep learning researches focus on how to solving the problem in computer vision. Actually, with the smart devices quickly spreading among people, there would be numerous data created, like path tracking data and GPS data. Adopting those kinds of data in dealing with HAR problem may be helpful. Based on the aspect of deep learning applied in HAR tasks, adaptively learning from new coming data as input to fine-tuning built HAR system ensemble in smart device instantly could be a more practical challenge in the field of human activity recognition and transfer learning, since this task requires HAR system possessing a stronger ability of evaluating the quality of data features.

References

[1] Cook, D., Feuz, K. D., & Krishnan, N. C. (2013). Transfer learning for activity recognition: A survey. *Knowledge and information systems*, *36*(3), 537-556.

[2] Yang, Q. (2009, July). Activity recognition: linking low-level sensors to high-level intelligence. In *IJCAI* (Vol. 9, pp. 20-25).

[3] Hu, D. H., Zheng, V. W., & Yang, Q. (2011). Cross-domain activity recognition via transfer learning. *Pervasive and Mobile Computing*, *7*(3), 344-358.

[4] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[5] Ranasinghe, S., Al Machot, F., & Mayr, H. C. (2016). A review on applications of activity recognition systems with regard to performance and evaluation. International Journal of Distributed Sensor Networks, 12(8), 1550147716665520.

[6] Chen, Y., & Xue, Y. (2015, October). A deep learning approach to human activity recognition based on single accelerometer. In *Systems, man, and cybernetics (smc), 2015 ieee international conference on* (pp. 1488-1492). IEEE.

[7] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[8] Jiang, W., & Yin, Z. (2015, October). Human activity recognition using wearable sensors by deep convolutional neural networks. In Proceedings of the 23rd ACM international conference on Multimedia (pp. 1307-1310). ACM.

[9] Sunny, J. T., George, S. M., Kizhakkethottam, J. J., Sunny, J. T., George, S. M., & Kizhakkethottam, J. J. (2015). Applications and challenges of human activity recognition using sensors in a smart environment. IJIRST Int. J. Innov. Res. Sci. Technol, 2, 50-57.

[10] Cao, H., Nguyen, M. N., Phua, C., Krishnaswamy, S., & Li, X. (2012, September). An integrated framework for human activity classification. In UbiComp (pp. 331-340).

[11] Wang, J., Chen, Y., Hu, L., Peng, X., & Philip, S. Y. (2018, March). Stratified transfer learning for cross-domain activity recognition. In 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom) (pp. 1-10). IEEE.

[12] Cook, D., Feuz, K. D., & Krishnan, N. C. (2013). Transfer learning for activity recognition: A survey. *Knowledge and information systems*, *36*(3), 537-556.

[13] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. science, 313(5786), 504-507.

[14] Minsky, M. (1974). A framework for representing knowledge.

[15] Bengio, Y., Lamblin, P., Popovici, D., & Larochelle, H. (2007). Greedy layer-wise training of deep networks. In Advances in neural information processing systems (pp. 153-160).

[17] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural networks, 61, 85-117.

[18] Fukushima, K. (1988). Neocognitron: A hierarchical neural network capable of visual pattern recognition. Neural networks, 1(2), 119-130.

[19] Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proceedings of the national academy of sciences, 79(8), 2554-2558.

[20] Werbos, P. J. (1982). Applications of advances in nonlinear sensitivity analysis. In System modeling and optimization (pp. 762-770). Springer, Berlin, Heidelberg.

[22] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.

[23] Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. Neural computation, 18(7), 1527-1554.

[24] Larochelle, H., Mandel, M., Pascanu, R., & Bengio, Y. (2012). Learning algorithms for the classification restricted boltzmann machine. *Journal of Machine Learning Research*, *13*(Mar), 643-669.

[25] Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, *22*(10), 1345-1359.

[26] Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and transferring mid-level image representations using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1717-1724).

[27] Van Kasteren, T. L. M., Englebienne, G., & Kröse, B. J. (2010, May). Transferring knowledge of activity recognition across sensor networks. In *International Conference on Pervasive Computing* (pp. 283-300). Springer, Berlin, Heidelberg.

[28] Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2013, April). A public domain dataset for human activity recognition using smartphones. In *ESANN*.

[29] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature, 521*(7553), 436.

[30] Kwapisz, J. R., Weiss, G. M., & Moore, S. A. (2011). Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter, 12*(2), 74-82.

[31] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117.

[32] Ronao, C. A., & Cho, S. B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Systems with Applications*, *59*, 235-244.

[33] Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and trends® in Machine Learning*, *2*(1), 1-127.

[34] Längkvist, M., Karlsson, L., & Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, *42*, 11-24.

[35] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks?. In *Advances in neural information processing systems* (pp. 3320-3328).

[36] Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, *22*(10), 1345-1359.

[37] Bhatia, N., & Rana, M. C. (2015). Deep Learning Techniques and its Various Algorithms and Techniques.

[38] Yao, S., Hu, S., Zhao, Y., Zhang, A., & Abdelzaher, T. (2017, April). Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 351-360). International World Wide Web Conferences Steering Committee.

[39] Lara, O. D., & Labrador, M. A. (2013). A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys and Tutorials*, *15*(3), 1192-1209.

[40] Ronao, C. A., & Cho, S. B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Systems with Applications, 59*, 235-244.

[41] Hu, D., & Yang, Q. (2011). Transfer learning for activity recognition via sensor mapping. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain* (p. 1962).

[42] Bulling, A., Blanke, U., & Schiele, B. (2014). A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, *46*(3), 33.

[43] Chen, Y., & Xue, Y. (2015, October). A deep learning approach to human activity recognition based on single accelerometer. In *Systems, man, and cybernetics (smc), 2015 ieee international conference on* (pp. 1488-1492). IEEE.

[44] Penatti, O. A., & Santos, M. F. (2017). Human activity recognition from mobile inertial sensors using recurrence plots. *arXiv preprint arXiv:1712.01429*.

[45] Zebin, T., Scully, P. J., & Ozanyan, K. B. (2016, October). Human activity recognition with inertial sensors using a deep learning approach. In *SENSORS, 2016 IEEE* (pp. 1-3). IEEE.

[46] Wang, J., Chen, Y., Hu, L., Peng, X., & Philip, S. Y. (2018, March). Stratified transfer learning for cross-domain activity recognition. In *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)* (pp. 1-10). IEEE.

[47] Chen, W. H., Cho, P. C., & Jiang, Y. L. (2017). Activity Recognition Using Transfer Learning. *Sensors and Materials*, *29*(7), 897-904.

[48] Yang, J., Nguyen, M. N., San, P. P., Li, X., & Krishnaswamy, S. (2015, July). Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition. In *Ijcai* (Vol. 15, pp. 3995-4001).

[49] Murad, A., & Pyun, J. Y. (2017). Deep recurrent neural networks for human activity recognition. *Sensors*, *17*(11), 2556.

[50] Ignatov, A. (2018). Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. *Applied Soft Computing*, *62*, 915-922.

[51] Nweke, H. F., Teh, Y. W., Al-Garadi, M. A., & Alo, U. R. (2018). Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications*.

[52] Ordóñez, F. J., & Roggen, D. (2016). Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, *16*(1), 115.

[53] Ronaoo, C. A., & Cho, S. B. (2015). Evaluation of deep convolutional neural network architectures for human activity recognition with smartphone sensors. *한국정보과학회학술발표논문집*, 858-860.

[54] Vrigkas, M., Nikou, C., & Kakadiaris, I. A. (2015). A review of human activity recognition methods. *Frontiers in Robotics and AI, 2*, 28.

[55] Köping, L., Shirahama, K., & Grzegorzek, M. (2018). A general framework for sensor-based human activity recognition. *Computers in biology and medicine*, *95*, 248-260.

[56] Hassan, M. M., Uddin, M. Z., Mohamed, A., & Almogren, A. (2018). A robust human activity recognition system using smartphone sensors and deep learning. *Future Generation Computer Systems*, *81*, 307-313.

[57] Rokni, S. A., Nourollahi, M., & Ghasemzadeh, H. (2018). Personalized Human Activity Recognition Using Convolutional Neural Networks. *arXiv preprint arXiv:1801.08252*.

[58] Walse, K. H., Dharaskar, R. V., & Thakare, V. M. (2016, March). Performance Evaluation of Classifiers on WISDM Dataset for Human Activity Recognition. In *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies* (p. 26). ACM.

[59] Sunny, J. T., George, S. M., Kizhakkethottam, J. J., Sunny, J. T., George, S. M., & Kizhakkethottam, J. J. (2015). Applications and challenges of human activity recognition using sensors in a smart environment. *IJIRST Int. J. Innov. Res. Sci. Technol*, *2*, 50-57.

[60] Ranasinghe, S., Al Machot, F., & Mayr, H. C. (2016). A review on applications of activity recognition systems with regard to performance and evaluation. *International Journal of Distributed Sensor Networks*, *12*(8), 1550147716665520.

[61] Shoval, N., & Isaacson, M. (2007). Sequence alignment as a method for human activity analysis in space and time. *Annals of the Association of American geographers*, *97*(2), 282-297.

[62] Aggarwal, J. K., & Xia, L. (2014). Human activity recognition from 3d data: A review. *Pattern Recognition Letters*, *48*, 70-80.

[63] Aggarwal, J. K., & Cai, Q. (1999). Human motion analysis: A review. *Computer vision and image* understanding, 73(3), 428-440.

[64] Aggarwal, J. K., & Ryoo, M. S. (2011). Human activity analysis: A review. ACM Computing Surveys (CSUR), 43(3), 16.

[65] Jiang, W., & Yin, Z. (2015, October). Human activity recognition using wearable sensors by deep convolutional neural networks. In Proceedings of the 23rd ACM international conference on Multimedia (pp. 1307-1310). ACM.

[66] Senin, P. (2008). Dynamic time warping algorithm review. Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA, 855, 1-23.

[67] Berndt, D. J., & Clifford, J. (1994, July). Using dynamic time warping to find patterns in time series. In KDD workshop (Vol. 10, No. 16, pp. 359-370).

[68] Tran, S. D., & Davis, L. S. (2008, October). Event modeling and recognition using markov logic networks. In European Conference on Computer Vision (pp. 610-623). Springer, Berlin, Heidelberg.

[69] Müller, M. (2007). Information retrieval for music and motion(Vol. 2). Heidelberg: Springer.

[70] Keogh, E. J., & Pazzani, M. J. (2001, April). Derivative dynamic time warping. In Proceedings of the 2001 SIAM International Conference on Data Mining (pp. 1-11). Society for Industrial and Applied Mathematics.

[71] Woodworth, R. S., & Thorndike, E. L. (1901). The influence of improvement in one mental function upon the efficiency of other functions.(I). Psychological review, 8(3), 247.

[72] Thrun, S., & Pratt, L. (1998). Learning to learn: Introduction and overview. In Learning to learn (pp. 3-17). Springer, Boston, MA.

[73] Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10), 1345-1359.

[74] Zhang, J., Li, W., & Ogunbona, P. (2017). Transfer learning for cross-dataset recognition: a survey. arXiv preprint arXiv:1705.04396.

[75] Raina, R., Battle, A., Lee, H., Packer, B., & Ng, A. Y. (2007, June). Self-taught learning: transfer learning from unlabeled data. In Proceedings of the 24th international conference on Machine learning (pp. 759-766). ACM.

[76] Shimodaira, H. (2000). Improving predictive inference under covariate shift by weighting the log-likelihood function. Journal of statistical planning and inference, 90(2), 227-244.

[77] Daume III, H., & Marcu, D. (2006). Domain adaptation for statistical classifiers. Journal of Artificial Intelligence Research, 26, 101-126.

[78] Dai, W., Yang, Q., Xue, G. R., & Yu, Y. (2008, July). Self-taught clustering. In Proceedings of the 25th international conference on Machine learning (pp. 200-207). ACM.

[79] Wang, Z., Song, Y., & Zhang, C. (2008, September). Transferred dimensionality reduction. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 550-565). Springer, Berlin, Heidelberg.

[80] Yao, Y., & Doretto, G. (2010, June). Boosting for transfer learning with multiple sources. In Computer vision and pattern recognition (CVPR), 2010 IEEE conference on (pp. 1855-1862). IEEE.

[81] Argyriou, A., Evgeniou, T., & Pontil, M. (2007). Multi-task feature learning. In Advances in neural information processing systems (pp. 41-48).

[82] Sugiyama, M., Lawrence, N. D., & Schwaighofer, A. (2017). Dataset shift in machine learning. The MIT Press.

[83] Jiang, J., & Zhai, C. (2007). Instance weighting for domain adaptation in NLP. In Proceedings of the 45th annual meeting of the association of computational linguistics (pp. 264-271).

[84] Liao, X., Xue, Y., & Carin, L. (2005, August). Logistic regression with an auxiliary data source. In Proceedings of the 22nd international conference on Machine learning (pp. 505-512). ACM.

[85] Huang, J., Gretton, A., Borgwardt, K. M., Sch 鰈 kopf, B., & Smola, A. J. (2007). Correcting sample selection bias by unlabeled data. In Advances in neural information processing systems (pp. 601-608).

[86] Bickel, S., Br 點 kner, M., & Scheffer, T. (2007, June). Discriminative learning for differing training and test distributions. In Proceedings of the 24th international conference on Machine learning (pp. 81-88). ACM.

[87] Sugiyama, M., Nakajima, S., Kashima, H., Buenau, P. V., & Kawanabe, M. (2008). Direct importance estimation with model selection and its application to covariate shift adaptation. In Advances in neural information processing systems (pp. 1433-1440).

[88] Dai, W., Xue, G. R., Yang, Q., & Yu, Y. (2007, August). Co-clustering based classification for out-of-domain documents. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 210-219). ACM.

[89] Ando, R. K., & Zhang, T. (2005, June). A high-performance semi-supervised learning method for text chunking. In Proceedings of the 43rd annual meeting on association for computational linguistics (pp. 1-9). Association for Computational Linguistics.

[90] Blitzer, J., McDonald, R., & Pereira, F. (2006, July). Domain adaptation with structural correspondence learning. In Proceedings of the 2006 conference on empirical methods in natural language processing (pp. 120-128). Association for Computational Linguistics.

[91] Daum?III, H. (2009). Frustratingly easy domain adaptation. arXiv preprint arXiv:0907.1815.

[92] Argyriou, A., Micchelli, C. A., Pontil, M., & Ying, Y. (2008). A spectral regularization framework for multi-task structure learning, NIPS 20. Journal Publications on Mathematics (Harmonic Analysis).

[93] Lee, S. I., Chatalbashev, V., Vickrey, D., & Koller, D. (2007, June). Learning a meta-level prior for feature relevance from multiple related tasks. In Proceedings of the 24th international conference on Machine learning (pp. 489-496). ACM.

[94] Lawrence, N. D., & Platt, J. C. (2004, July). Learning to learn with the informative vector machine. In Proceedings of the twenty-first international conference on Machine learning (p. 65). ACM.

[95] Bonilla, E. V., Chai, K. M., & Williams, C. (2008). Multi-task Gaussian process prediction. In Advances in neural information processing systems (pp. 153-160).

[96] Schwaighofer, A., Tresp, V., & Yu, K. (2005). Learning Gaussian process kernels via hierarchical Bayes. In Advances in neural information processing systems (pp. 1209-1216).

[97] Evgeniou, T., & Pontil, M. (2004, August). Regularized multi--task learning. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 109-117). ACM.

[98] Gao, J., Fan, W., Jiang, J., & Han, J. (2008, August). Knowledge transfer via multiple model local structure mapping. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 283-291). ACM.

[99] Mihalkova, L., Huynh, T., & Mooney, R. J. (2007, July). Mapping and revising Markov logic networks for transfer learning. In AAAI (Vol. 7, pp. 608-614).

[100] Mihalkova, L., & Mooney, R. J. (2008, July). Transfer learning by mapping with minimal target data. In Proceedings of the AAAI-08 workshop on transfer learning for complex tasks.

[101] Bobick, A. F., & Davis, J. W. (2001). The recognition of human movement using temporal templates. IEEE Transactions on pattern analysis and machine intelligence, 23(3), 257-267.

[102] Gorelick, L., Blank, M., Shechtman, E., Irani, M., & Basri, R. (2007). Actions as space-time shapes. IEEE transactions on pattern analysis and machine intelligence, 29(12), 2247-2253.

[103] Rodriguez, M. (2010). Spatio-temporal maximum average correlation height templates in action recognition and video summarization.

[104] Wang, L., Qiao, Y., & Tang, X. (2015). Action recognition with trajectory-pooled deep-convolutional descriptors. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4305-4314).

[105] Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., & Darrell, T. (2015). Long-term recurrent convolutional networks for visual recognition and description. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2625-2634).

[106] Venugopalan, S., Xu, H., Donahue, J., Rohrbach, M., Mooney, R., & Saenko, K. (2014). Translating videos to natural language using deep recurrent neural networks. arXiv preprint arXiv:1412.4729.

[107] Yacoob, Y., & Black, M. J. (1999). Parameterized modeling and recognition of activities. Computer Vision and Image Understanding, 73(2), 232-247.

[108] Lublinerman, R., Ozay, N., Zarpalas, D., & Camps, O. (2006, August). Activity recognition from silhouettes using linear systems and model (in) validation techniques. In Pattern Recognition, 2006. ICPR 2006. 18th International Conference on (Vol. 1, pp. 347-350). IEEE.

[109] Yamato, J., Ohya, J., & Ishii, K. (1992, June). Recognizing human action in time-sequential images using hidden markov model. In Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on (pp. 379-385). IEEE.

[110] Oliver, N. M., Rosario, B., & Pentland, A. P. (2000). A Bayesian computer vision system for modeling human interactions. IEEE transactions on pattern analysis and machine intelligence, 22(8), 831-843.

[111] Park, S., & Aggarwal, J. K. (2004). A hierarchical Bayesian network for event recognition of human actions and interactions. Multimedia systems, 10(2), 164-179.

[112] Nevatia, R., Hobbs, J., & Bolles, B. (2004, June). An ontology for video event representation. In Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on (pp. 119-119). IEEE.

[113] Vu, V. T., Bremond, F., & Thonnat, M. (2003, August). Automatic video interpretation: A novel algorithm for temporal scenario recognition. In IJCAI (Vol. 3, pp. 1295-1300).

[114] Na, S., Ramachandran, K. M., & Ji, M. Real time Activity Recognition using Smartphone Accelerometer.