HUMAN BEHAVIOR CLASSIFICATION WITH CONVOLUTIONAL NEURAL NETWORKS

Su Myat Thwin

Assistant Lecturer, Faculty of Information Science, University of Computer Studies (Pathein), Ayeyarwady, Myanmar

ABSTRACT

Classification of human behavior is a challenging task that has many applications in various fields. In this paper, convolutional neural networks (CNNs) are proposed to classify the human behavior in the real-world scene. CNNs makes efficient and easy to classify human behavior. In the experiment, a dataset is created that collect image frames of human behavior from video in the scenes such as indoor and outdoor. In the system, videos in KTH dataset are used to classify six activities in human behavior including running, boxing, jogging, hand waving, handclapping and walking. The result shows that CNNs can classify human behavior automatically and can perform to get the high accuracy in classifying human behavior.

Keyword: - *Deep Learning, Human Behavior, Convolutional Neural Networks*

1. INTRODUCTION

With the increasing number of populations in today world, human safety is an important role for security purposes. In recent years, with the development of advanced science and technology, human behavior classification as an attractive area in the fields of computer vision technology. Behavior classification in computer vision is necessarily important to make detecting, tracking and analysing in classifying human behavior action based on videos and it is the main technology in pattern recognition and artificial intelligent control system. Classification of human behavior can be effectively and efficiently implemented by specific body movement and by detecting facial features.

With the advanced in technology, deep learning approaches have been used in various areas such as image recognition, speech recognition and behavior recognition in time series. Deep learning is also a research trend in the last year to develop applications in computer vision fields. The major development in object classification was achieved to have the improvements in object representation and machine learning models. CNNs [4] as a deep learning approach that improves traditional neural network and it does not need to work manually for design features, images as input can be directly into the network. In this paper, the system shows that CNNs makes successfully classify in human behavior action with high accuracy.

The rest section in this paper is organized as follows: The motivation of the research is in section 2. The objectives of this system are in section 3. Some related works for the human behavior classification is in section 4. The methodology used in this system is described in section 5, including model architecture and implementation and evaluation result. Finally, the system is concluded in section 6.

2. MOTIVATION

In recent years, people are the victims of accidents lack of security. With an increasing number of accidents, human behaviors are needed to monitor to prevent from happening unexpected events. As an important role in computer vision technology, human behavior classification has many applications in robot vision, human-computer interaction, etc. The applications based on neural networks can efficient make to classify objects, image segmentation and it has many models to classify behavior with a better performance. Deep Neural Networks (DNNs) becomes a more powerful model in applications that are developed by machine learning. DNNs show the better performance in image classification. CNNs are an effective model in deep learning approaches to understand image content. CNNs consist of multi-layered based on neural network where there is a large number of connections between neurons [3]. CNNs is a self-learning model where the network is trained using a large dataset of dynamic images with their weights. CNNs can give a better performance for visual recognition. CNNs also helps to build an artificial intelligence system in advance which will be able to classify human behavior from a dataset of multiple dynamic images for a specific domain area where human behavior will be classified by the person who is in need of the security.

3. OBJECTIVES

- Developing a system to classify human behavior in the scene using CNNs
- Training a system that automatically recognizes human behavior
- Gaining a high accuracy in classifying human behavior
- Analysing in using a pretrained network based on CNNs to classify an image category as a feature extractor

4. RELATED WORKS

CNNs are one of the main categories in image classification, image recognition. CNNs are widely used in most of the application areas in image processing. CNNs are included in a class of deep learning models. CNNs consist of three stages to form a single neural network which is trained to classify output from raw pixel values inputs. The spatial structure of images is given through the connection between layers (local filters), parameter sharing (convolutions) and special local invariance-building neurons (max pooling). Due to the computational costs, CNNs have been applied in developing small scale image recognition. But development on GPU enables that CNNs is applied in large scale networks with millions of parameters to have significant improvement in image classification. In [1], a large, deep convolutional neural network is trained to classify images of 1.2 million with high-resolutions in the ImageNet LSVRC-2010 contest into the 1000 distinct classes. A model with a large learning capacity is needed to learn about thousands of objects from millions of images. In [5], several deep learning networks are trained to classify 10,000 images in the Tiny ImageNet dataset into 200 different classes. The experimentation and evaluation results show that the common ability about the performance of CNNs in tasks with classifying images. CNNs show that it can give reliable results in object recognition and detection, and efficient and effective models that are helpful in real world applications. CNNs are a very good efficient and developing platform for image and video classification. The system implemented with CNNs will be adaptable in any dynamic environment [6]. In this system, CNNs based on the pretrained network ResNet-50 is used to classify human behavior with high accuracy.

5. METHODOLOGY

5.1 Model Architecture

CNNs are commonly applied in visual imagery and also a class of deep neural networks. CNNs consist of an input layer, multiple hidden layers and an output layer. The intermediate layers make up the parts the CNNs. These are convolutional layers, rectified linear units (ReLU) and pooling layers [3]. These layers are three fully-connected layers and the final layer is the classification layer. The architecture of CNNs as shown in Fig.1,

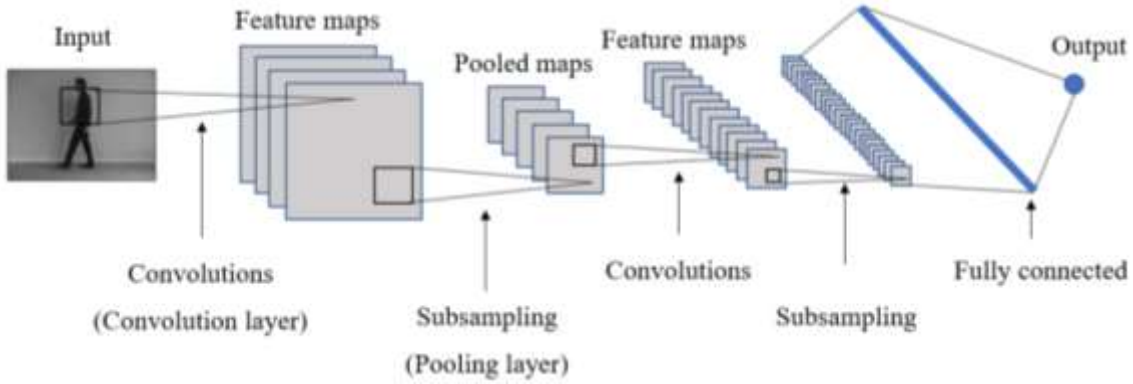

Fig.1: CNNs architecture

The layers of CNNs are organised with three dimensions: width, height, depth. Every layer in CNNs is made up of neurons. CNNs consist of two components including feature extraction and classification. Convolution layer and pooling layer perform as a part of feature extraction and the fully connected layers as a classifier. Convolution is a main block of CNNs.

Firstly, we prepared a dataset of six activities for human action recognition including running, boxing, jogging, hand waving, handclapping and walking. The system analyses images and finds similar features. Images with similar actions form a certain class. Other video frame images are added to test the performance of the system. The first layer in CNNs defines the input dimensions. Image classification takes an input image, process it and classify according the given categories (running, boxing, jogging, hand waving, handclapping and walking). In this system, the size 224 x 224 x 3 is used for input image and 1000 classes from the ImageNet dataset is applied in the classification layers.

5.2 Implementation and Evaluation

In this system, a dataset is created inspired by different image frames in KTH dataset [2] that is used for training to classify human action. This dataset consists of six kinds of human action video which include running, boxing, jogging, hand waving, handclapping and walking. Each category has 25 people action and that is divided into indoor scene and outdoor scene. Some image frames in video from this dataset as shown in Fig.2. Images from the dataset are classified into different categories by using a multiclass linear SVM that trained with CNNs features extracted from the images.

Fig.2: Sample image frames in video from KTH dataset

Firstly, images are loaded into memory and adjust the number of images per category for balancing the number of images in the training set. Next, pretrained network is loaded. There are other popular pretrained networks including

AlexNet, GoogLeNet, VGG-16 and VGG-19. In this system, ResNet-50 model is used to train network. First section of ResNet-50 is visualized as shown in Fig.3.

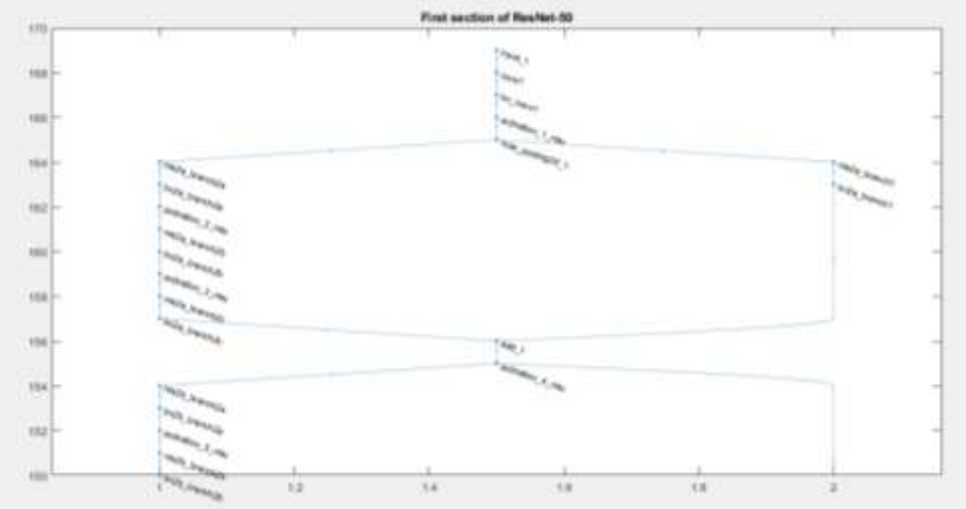

Fig.3: First section of ResNet-50

The sets are split into training data and validation data. Training data is built with 30% of images that are picked from each set and the validation data is 70% of images. The data are randomly split to avoid biasing the results. The CNNs model processes the training sets and test sets. In the pre-processing step for CNNs, RGB images that have 224 x 224 dimensions are used. Image features such as edges and blobs are captured from the layers at the beginning of the network. The network filter weights from the first convolutional layer are visualized as shown in Fig.4. Next, CNNs image features are used to train a multiclass SVM classifier.

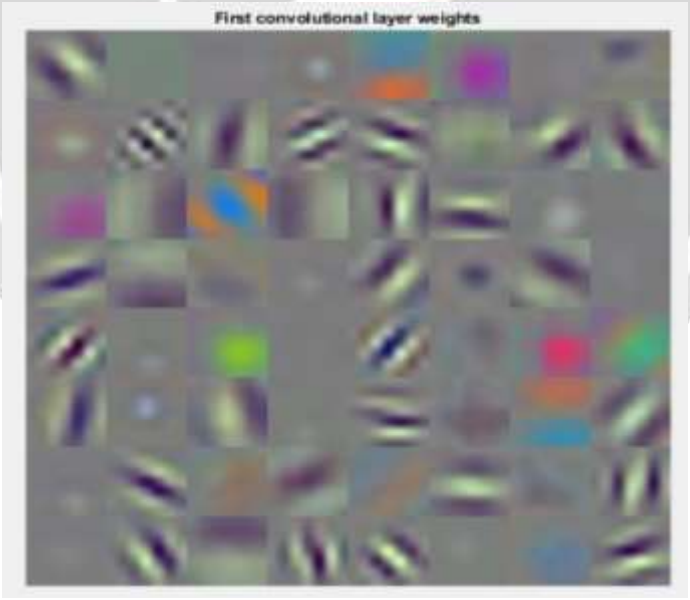

Fig.4: First convolutional layer weights

To measure the accuracy of the trained classifier, the test features can be passed to the classifier. In this system, the accuracy for classifier is 0.9934. Now new images are classified applying with newly trained classifier. The result shows that the CNNs trained on the ResNet-50 can efficiently classify the human behavior according to six human activities such as running, boxing, jogging, hand waving, handclapping and walking.

6. CONCLUSIONS

Human behavior classification is an attractive area of research in the field of Computer Science that has been motivated by the need of automated video surveillance applications. We proposed a deep neural network model CNNs for human behavior classification from video frame. By using a multiclass linear SVM trained with CNNs features extracted from the images, images from video frames are classified into categories. ResNet-50 is one of the pretrained network models that are trained on the ImageNet dataset. ResNet-50 requires Deep Learning Toolbox, Statistics and Machine Learning Toolbox and Deep Learning Toolbox Model for ResNet-50 Network. In this work, neural networks can build high level representation of raw input without any pre-processing. The result shows that CNNs can classify human behavior automatically and can perform to get the high accuracy in classifying human behavior.

Future work will train CNNs with more layers or changing the network architecture in this area. Additionally, the more training data allow the network the better generalize on its classifications.

7. REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. Hinton. "Imagenet classification with deep convolutional neural networks". In NIPS, 2012.

[2] Christian Schuldt, Ivan Laptev and Barbara Caputo; in Proc. ICPR'04, Cambridge, UK.

[3] J. Krause, T. Gebru, J. Deng, L.-J. Li, and L. Fei-Fei, "Learning Features and Parts for Fine-Grained Recognition," 2014 22nd International Conference on Pattern Recognition, 2014.

[4] Le Cun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[5] Leon Yao, John Miller, "Tiny ImageNet Classification with Convolutional Neural Networks", 2012.

[6] MD. Ashik Ahmed, Mushfique Ahmed Isha, Al-Amin Ahmed, "Dynamic Image Analysis for Abnormal Behavior Detection", BRAC University, Dhaka, 2017.