Emotion and Hand Recognition with CNN

Sangeeta Kumawat   
Computer Science

Kent State UniversityKent, Ohio USA  
skumawat@kent.edu

Jenifer Werthman  
Computer Science  
Kent State UniversityKent, Ohio USA  
jwerthma@kent.edu

*Abstract*

This paper is a summary of a study of the CNN network applied to the classification of emotions and the classification of determining if there are hands or no hands in a photo. The question we are asking is, how reliable is the CNN algorithms as it relates to these topics and what can we do within the algorithm to improve accuracy.

Keywords—CNN, emotion, hand, touching, object detection, classification, algorithm

# Introduction

In the paper we will discuss the summation of 2 classifications: Emotion detection and hands, no hands, and touching detection within images using the CNN algorithm. The algorithm. The Convolutional Neural Networks (CNN) is an algorithm that can take an image as an input and dissect it into smaller more manageable pieces images and data to compare and differentiate against other images. In doing this it forms patterns that it will compare to determine with any level of accuracy if the original image contains similarities with other images in the model. The question that we are asking is how well does CNN perform and what parameters and layers can we use to help improve this performance.

# Convolutional Neural Networks (CNN)

## Summary

Convolutional Neural Networks (CNN) are an algorithm based on the human brain inspired by the organization of the Visual Cortex developed in the 1980’s. An image is broken down into a matrix of information and pixel values. As you can imagine, this process will produce a lot of data; therefore the larger the images and the amount of images will become computationally intensive. CNN’s have several standard layers, though the methods to process the layers may vary: Input, Convolution Layer, Max Pooling, Fully Connected layer, Output. This algorithm requires a large amount of data to train which means a large computing resources.

## Input

The input layer takes in the train and test images and transforms them into normalized matrixes of information pertaining to each image.

## Convolution Layer

These are typically within what we would call the hidden layers. There could be 1 or more convolution layers used. These layers are the main layers of the network which contain the kernels, parameters, and values used throughout the training. This layer creates an activation map by multiplying the input image by the kernel for every element in the input image matrix. An Activation Map (otherwise known as a feature map) shows the most important areas of an image for a particular prediction.

## Pooling

Three types of pooling is Max Pooling, Min Polling, and Average Pooling. Max Pooling is a pooling operation that calculates the maximum value for each section of the feature map. Min Polling find the minimum value in each section of the activation map. Average Pooling finds the average value for each section of the activation map.

This layer helps to locate share and smooth features for edge and point detection.

## Fully Connected

This is where each neuron transforms the input vector through a weights matrix. All possible neurons are connection layer to layer. The result is that every input will influence the output. This allows the network to learn non-linear combinations for each feature.

## Output

The output of the CNN is a matrix/array the same batch size as the input but other information will/may change.

# Emotions Classification

For the emotional classification, our goal was to build a solution which would feed in the emotion images provided with a level of accuracy over 80% and then play with our options to find out what reduces accuracy and what improves it.

## Initial Analysis

We started this project by reviewing various articles, documentation, and examples of Image Classification using CNN. Next we tried to run some of the examples we found on Google Collab, Spyder, Jupyter, and VS Code. Layers

## Epochs

## Optimizer

## Challenges

We encountered many issues initially where some examples would work on one platform but not on another. Once we did get things working we would get misc random errors at various points.

Once we finally got it running, the next challenge was to figure out how to read images from a file folder rather than an API.

Then we begin seeing run times of 3.5 hours+ for the datasets that we were provided for the emotion dataset.

## Batch Size

## Figures and Results

## Final Results

# Hands-Touch Classification

##### References

For this analysis we analized code provided by our professor Dr JoonYoon Kim as well as several Kaggle references. Ultimately, we combined several ideas to form our finished resources for this project.

1. ??
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.