

BABEŞ BOLYAI UNIVERSITY, CLUJ NAPOCA, ROMÂNIA
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Emotion detection in preschoolers' behaviour

– ITSG report –

Team members

Dragos-Alexandru Grigore, dragos.grigore@mail.com
Ana-Corina Todoran, todorananacorina13@gmail.com
Radu-George Vele, georgevele2016@gmail.com
Software Engineering, 258

Abstract

The purpose of this project is to present a new application of emotional recognition, being used for children, by applying a filter to the video they are watching based on their face expression. This is different from other projects by the fact that it is specifically created for children. We are using artificial intelligence in order to find a child's reaction to different visual actions (showed by sample videos).

In this project, we are using an intelligent algorithm - Convolutional Neuronal Network such that we can identify the different children expressions. The training and evaluation has been done using Fer2013 dataset, which is comprised of 28.709 training and 3589 testing images.

Contents

1	Introduction	1
1.1	What? Why? How?	1
1.1.1	What?	1
1.1.2	Why?	1
1.1.3	How?	1
1.2	Paper structure and original contribution(s)	1
2	Scientific Problem	3
2.1	Problem definition	3
3	State of the art/Related work	4
3.1	State of art	4
3.2	Useful tools	5
4	Investigated approach	6
5	Application	9
5.1	Methodology	11
5.2	Data	11
5.3	Results	11
5.4	Discussion	12
6	Conclusion and future work	13

List of Tables

4.1 Model layers	8
----------------------------	---

List of Figures

4.1 Emotion based filters	7
5.1 Login screen	10
5.2 Application dashboard	10

Chapter 1

Introduction

1.1 What? Why? How?

1.1.1 What?

We are addressing the problem of boredom and sadness that comes at kindergarten children when using the computer. This problem usually occurs because children usually get bored very fast when doing any kind of action.

1.1.2 Why?

The emotion of a kid has a big impact on his interaction with other people and it also affects his or her childhood, therefore this issue must be fixed as soon as possible.

1.1.3 How?

Our basic approach in order to solve this issue is to create an application that is worthy enough for children's attention such that we can capture the transition of their emotions and show them this transition in real time. This will be done by applying filters to the video they are choosing to watch, based on their face.

1.2 Paper structure and original contribution(s)

This paper contains bibliographical references is structured in 6 chapters that are describing the experimental study we have done in this semester on the problem of recognizing the face expressions of children. The research presented in this paper advances the theory, design, and implementation of several particular models.

The main contribution of this report is to present an intelligent algorithm for solving the problem of identifying the face expression of several children.

The second contribution of this report consists of building an intuitive, easy-to-use and user friendly software application.

The first chapter is a short introduction to the idea of this paper, while it also discusses the reasoning behind the paper.

The second chapter describes the scientific problem.

The third chapter presents the state of the art and the related work to the paper that we have written.

The fourth chapter aims to present the investigated approach by our team and the studies we have done.

The fifth chapter presents the application, the numerical results, together with the methodology used.

The sixth chapter aims to present a conclusion of this paper, while also presenting the future work to be done.

Chapter 2

Scientific Problem

2.1 Problem definition

The problem is the accurate detection and interpretation of emotion in kindergarten kids. It needs to be solved with an intelligent algorithm because we do not know how to define emotions programmatically.

The advantage of solving it with a neural network classifier is that the algorithm is adaptable, based on our data set, and it can figure on its own what it actually means for a kid to be happy, sad, surprised or any other emotion.

The inputs will be images of kids, and the output will be their emotion. Based on this, we will make a streaming service that will apply different filters on the videos based on the detected emotions. One of the challenges in doing this could be the different structure of the face for a child in comparison to an adult.

Chapter 3

State of the art/Related work

In the following, a few of the related work in the field of facial emotion recognition are going to be presented. Also, a section is dedicated to the useful tools in this domain.

3.1 State of art

1. In a research paper [2], an approach that combines the **Convolutional Neural Network (CNN)** approach with a low-dimensional discriminative embedding step that is learned by using triplet probability constraints in order to address the unconstrained face verification problems. The challenges have been done on the **Intelligence Advanced Research Projects Activity (IARPA) Janus Benchmark A** and labeled faces in the wild databases. The superior performance of this approach are reflected by the robustness of the representation learned by the deep **CNN**.
2. In another research paper [1], the usage of convolutional neural networks on facial recognition was thought to be a good approach, but when they managed to test it on real world data, a lot of issues appeared, mostly caused by light, variations of the head position, very subtle changes in the facial appearance. However, the studies conducted on four benchmark expression databases (Extended Cohn-Kanade, Oulu-CASIA database, MMI database and Static Facial Expressions in the Wild) have proved that island loss **CNN (Island Loss-Convolutional Neural Network (IL-CNN))** outperforms the regular **CNN**, based on the increased inter-class distances and the reduced intra-class variations. However, in this case, none of the databases contained kindergarten children benchmark expressions.

Again, while also true for the second research paper, the conducted tests were not done on kindergarten children.

3.2 Useful tools

- Face-recognition - a Python library that is used in order to recognize and manipulate faces. It can be used from the command line or directly from Python. It is build using dlib's state-of-the-art face recognition and it has a 99.38% on the [Labeled Faces in the Wild \(LFW\)](#) benchmark.
- Keras - an open-source library that provides a Python interface for artificial neural networks. It contains numerous implementations of layers, objectives, activation functions, optimizers and a host of tools to make working with image and text data easier to simplify the coding necessary for writing deep neural network code.
- OpenCV (Open Source Computer Vision Library) - is a library of programming functions mainly aimed at real-time computer vision. The library is cross-platform and free for use.
- Luxand - tool for recognizing and detecting the facial features within our application. It is used to monitor emotions that are associated with visual content (videos or images.)

Chapter 4

Investigated approach

Convolutional Neural Network (abbreviated CNN) is one deep neural network that is usually used for analyzing visual imagery [3]. They are regularized versions of multilayer perceptrons, which are fully connected networks (each neuron from a specific layer is connected to the neurons in the next layer). It takes advantage of a hierarchical pattern in data and assemble more complex patterns using smaller and simpler patterns, that means on a connectedness and complexity scale, they are on the lower extreme.

Our approach was to create an application that lets kids interact with elements with their emotion, and in the process of building this application, we compared 2 models based on the Face Emotion Recognition (Fer2013) dataset, and one provided to us by an external **Application Programming Interface (API)**.

Our application consists of a login based on the face of the user, which is done by using the face_recognition library. Once a user is logged in, he can see 2 screens: one showing his face (by using the PC camera) and one where a video will be played. Below this, 6 video options are presented. Once a video is selected, the video will play and the filter will be applied based on the user's face expression. You can see the expressions and corresponding filters in figure 4.1.



Figure 4.1: Emotion based filters

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 46, 46, 64)	640
conv2d_2 (Conv2D)	(None, 46, 46, 64)	36928
batch_normalization_1 (Batch Normalization)	(None, 46, 46, 64)	256
max_pooling2d_1 (MaxPooling2D)	(None, 23, 23, 64)	0
dropout_1 (Dropout)	(None, 23, 23, 64)	0
conv2d_3 (Conv2D)	(None, 23, 23, 128)	73856
batch_normalization_2 (Batch Normalization)	(None, 23, 23, 128)	512
conv2d_4 (Conv2D)	(None, 23, 23, 128)	147584
batch_normalization_3 (Batch Normalization)	(None, 23, 23, 128)	512
max_pooling2d_2 (MaxPooling2D)	(None, 11, 11, 128)	0
dropout_2 (Dropout)	(None, 11, 11, 128)	0
conv2d_5 (Conv2D)	(None, 11, 11, 256)	295168
batch_normalization_4 (Batch Normalization)	(None, 11, 11, 256)	1024
conv2d_6 (Conv2D)	(None, 11, 11, 256)	590080
batch_normalization_5 (Batch Normalization)	(None, 11, 11, 256)	1024
max_pooling2d_3 (MaxPooling2D)	(None, 5, 5, 256)	0
dropout_3 (Dropout)	(None, 5, 5, 256)	0
conv2d_7 (Conv2D)	(None, 5, 5, 512)	1180160
batch_normalization_6 (Batch Normalization)	(None, 5, 5, 512)	2048
conv2d_8 (Conv2D)	(None, 5, 5, 512)	2359808
batch_normalization_7 (Batch Normalization)	(None, 5, 5, 512)	2048
max_pooling2d_4 (MaxPooling2D)	(None, 2, 2, 512)	0
dropout_4 (Dropout)	(None, 2, 2, 512)	0
flatten_1 (Flatten)	(None, 2048)	0
dense_1 (Dense)	(None, 512)	1049088
dropout_5 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 256)	131328
dropout_6 (Dropout)	(None, 256)	0
dense_3 (Dense)	(None, 128)	32896
dropout_7 (Dropout)	(None, 128)	0
dense_4 (Dense)	(None, 7)	903
<hr/>		
Total params: 5,905,863		
Trainable params: 5,902,151		
Non-trainable params: 3,712		
<hr/>		
shanks@shanks-HP-360:~/Desktop/fer2013\$		

Table 4.1: Model layers

Chapter 5

Application

In our project we have compared 3 models for detecting human emotions, and one for recognizing peoples faces.

The first model we have tested was generated with a 6 layered Convolutional Neural Network in Keras, trained on 48x48px grayscale images preprocessed from [Facial Expression Recognition \(FER\)](#).

The second model was generated with a 17 layered Convolutional Neural Network also in Keras, on 48x48px images. This was trained both with 30 epochs and 100 epochs, the better one being the one trained with 30.

The third one is a model provided as a service by an external [API](#) called Luxand. It is a paid service, but has a free trial up to 10,000 emotion detections.

The application consists of the following functionalities:

- Login - by using the camera of the device
- Select a video to play
- Apply filter based on actual face expression

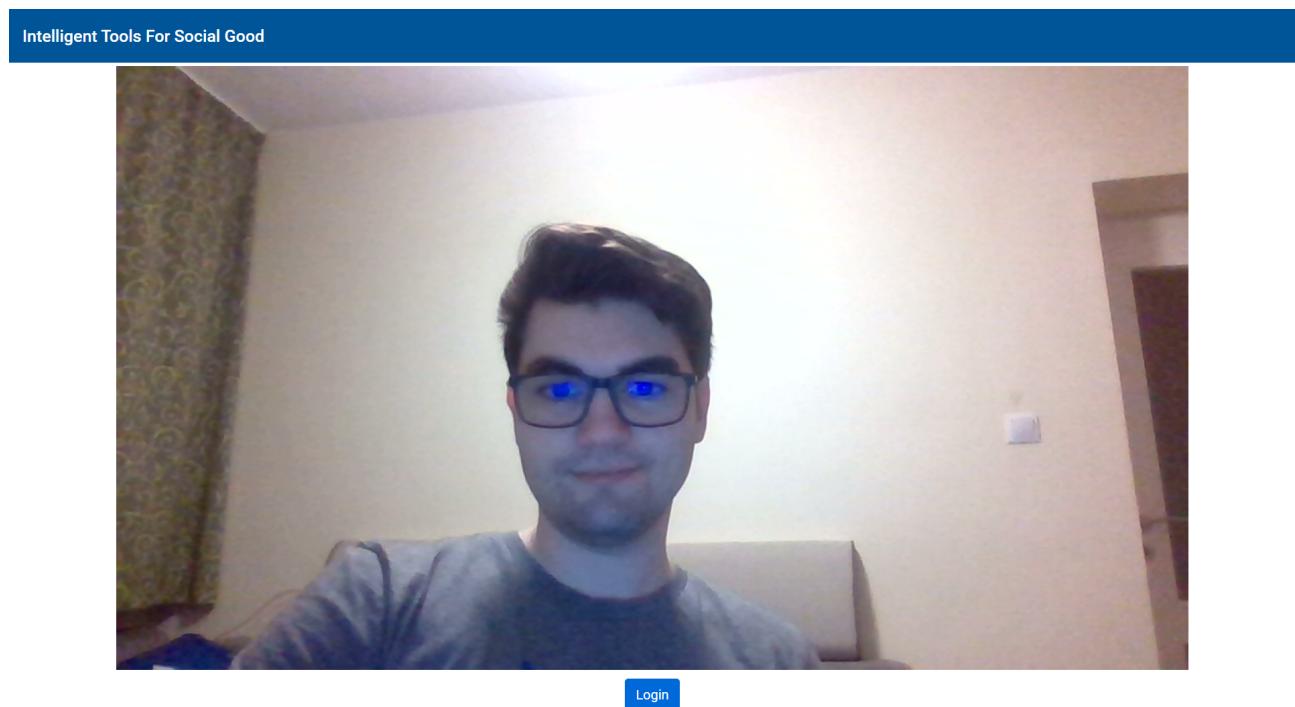


Figure 5.1: Login screen



Figure 5.2: Application dashboard

5.1 Methodology

- What are criteria you are using to evaluate your method?
 - Accuracy and Speed. In our use case, we would like to get emotions as accurately as possible, as fast as possible.
- What specific hypotheses does your experiment test? Describe the experimental methodology that you used.
 - We hypothesise that interaction with software using facial expressions is feasible. We have created an application that uses those exact expressions to apply filters on videos.
- What are the dependent and independent variables?
 - Categorical Cross Entropy loss
 - Adam optimizer
- What is the training/test data that was used, and why is it realistic or interesting? Exactly what performance data did you collect and how are you presenting and analyzing it? Comparisons to competing methods that address the same problem are particularly useful.
 - We used Facial Emotion Recognition (**FER**) 2013. Unfortunately it does not contain children. With the model we trained, on 30 epochs we got an accuracy of 85.26% (training) and 57.73% (validation) and on 100 epochs we got 84.42% (training) and 56.23% (validation).

5.2 Data

The data that is used for training and evaluation is from the Fer2013 dataset. The compressed dataset contains 92MB of compressed pictures. There are 28.709 training and 3589 testing images in the dataset. Each of the images were stored in 48x48 pixels. Unfortunately, the dataset does not contain any children images.

5.3 Results

- On 30 epochs we got an accuracy of 85.26% (training) and 57.73% (validation)
- On 100 epochs we got 84.42% (training) and 56.23% (validation)
- The 6 layered pre-trained model we have used only outputted surprise and fear, regardless of the actual emotion of the user.

- The Luxand API has yielded the best and consistent results, and that is the one we have chosen to integrate into our application (we also have endpoints for our trained one, but this one is primarily used)

5.4 Discussion

- Is your hypothesis supported?

-Yes, we can create an interactive application based on emotions, and it can be quite consistent.

- What conclusions do the results support about the strengths and weaknesses of your method compared to other methods?

-Our method is certainly not the best, but I believe there is a lot of potential for improvement, by upgrading the dataset and the prepossessing made

Chapter 6

Conclusion and future work

The problem of recognizing the face expressions of children is still a problem nowadays. Even though a lot of progress has been made in this field, by seeing the adoption of face recognition by bigger commercial applications, there still remain some aspects that are yet to be improved. The final accuracy tends to be around 60% which is more than half. This gives us a sign of progress in the field and a considerable step ahead, because children faces have a lack of structure in comparison with adults.

Next step for the application would be a selection of the video based on where the child is looking, as well as trying a new face recognition algorithm.

Acronyms

API Application Programming Interface. [6](#), [9](#), [12](#)

CNN Convolutional Neural Network. [4](#), [6](#)

FER Facial Expression Recognition. [9](#), [11](#)

IARPA Intelligence Advanced Research Projects Activity. [4](#)

IL-CNN Island Loss-Convolutional Neural Network. [4](#)

LFW Labeled Faces in the Wild. [5](#)

Bibliography

- [1] Ahmed Shehab Khan Zhiyuan Li James O'Reilly Yan Tong Jie Cai, Zibo Meng. Island loss for learning discriminative features in facial expression recognition.
- [2] Carlos D. Castillo Rama Chellappa Swami Sankaranarayanan, Axadeh Alavi. Triplet probabilistic embedding for face verification and clustering. 2017.
- [3] N.N.; Lyakhov P.A.; Valuev G.V.; Chervyakov N.I. Valueva, M.V.; Nagornov. Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. 177.