NAME THIS EXPRESSION

SPRING BOARD CAPSTONE PROJECT 2

FINAL REPORT

THE OBJECTIVE

Situation

Facial expressions are an important aspect of social interactions in humans. They extend beyond reflective expressions of our emotions and are a sophisticated communicative system. [1]

Challenge

Children on the autism spectrum have difficulties recognising facial expressions, body language and tonality of voice. [2]

Solution

Games on mobile and web applications are a useful and scalable approach to helping children with autism to practice social skills, such as identifying facial expressions, in environments with less pressure. This builds confidence and assists with the application of the skills in daily life.

Sources:
1 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2781887/
2 http://luxai.com/2017/11/29/improve-emotion-recognition-understanding-children-autism/

THE SOLUTION

A model that can classify facial expressions from photographs with a 90.9% accuracy was built.

The emotions have labels ranging from 0 to 7. (where 0=neutral, 1=anger, 2=contempt, 3=disgust, 4=fear, 5=happy, 6=sadness, 7=surprise).

Surprise

Happy

Disgust

THE MODEL WAS DEVELOPED IN 6 STAGES

1. IMAGE CAPTURE / DATA EXPLORATION
2. IMAGE PROCESSING
3. DATA WRANGLING
4. BASELINE
5. SELECT ARCHITECTURE
6. TUNE BEST MODEL

DATA SELECTION & WRANGLING

TRAIN & TEST CLASSIFICATION MODEL

REFINEMENT

THE RAW DATA WAS DEVELOPED BY RESEARCHERS

CK+ dataset

- Posed and spontaneous expressions
- Mostly grey
- Sequence begins with neutral expression in the first frame and progresses to the peak frame.
- 1 expression label per sequence corresponding to peak frame
- 7 discrete labels

Sources:
1 Cohn-Kanade AU-Coded Facial Expression (CK+) database by P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews.

SEQUENCE LABELS WERE USED TO LABEL IMAGES

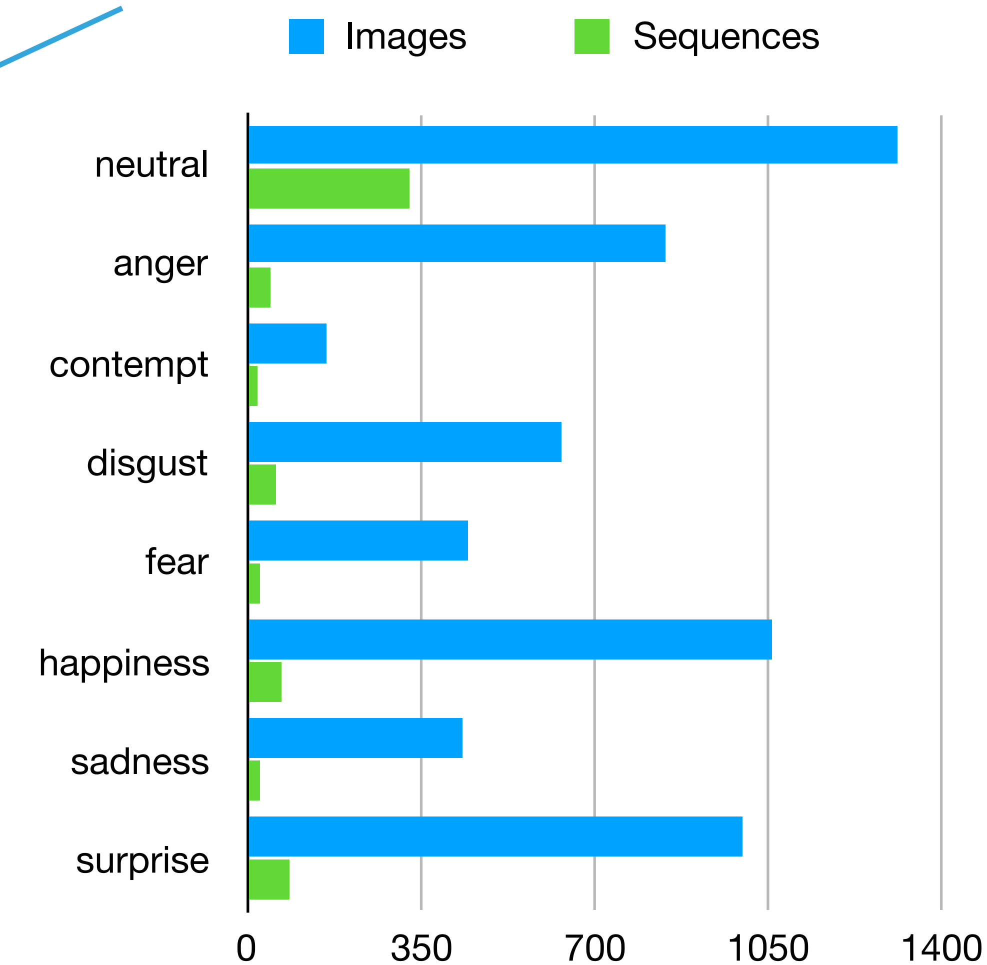
CK+ dataset

Full dataset:

- 123 subjects
- 593 sequences
- 10, 708 images

Labelled dataset:

- 327 sequences
- 5, 876 images



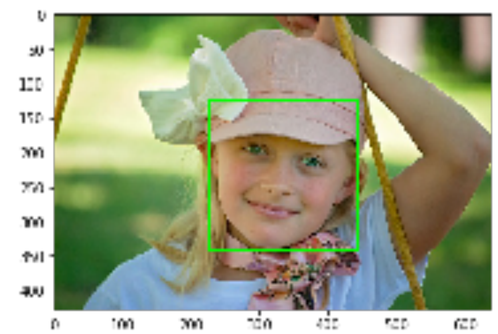
- sequences with no labels were dropped
- images labelled with sequence’s emotion
- set to integer
- first 4 images in sequences labelled neutral (i.e. 0) .

PHOTOGRAPHS PROCESSED TO HOMOGENISE FACIAL IMAGES

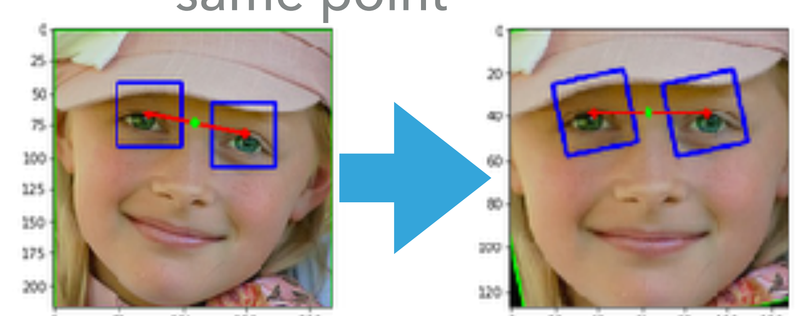
Original image

Processing

1. Identify facial region

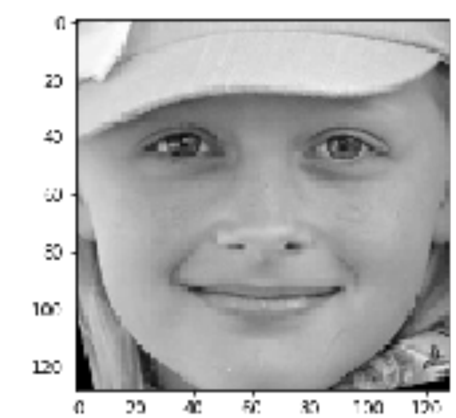


2. Normalise facial image
   i. Adjust the tilt
   ii. Scale the image
   iii. Center the face around the same point



3. Grey scale and save

Desired outcome



THE DATA WAS SPLIT 90/10 BY SEQUENCES

	TRAINING IMAGES	TEST IMAGES
neutral	1176	132
anger	739	103
contempt	151	10
disgust	568	64
fear	413	33
happiness	949	106
sadness	378	57
surprise	923	74
Total	**5297**	**579**
Proportion	**90.1%**	**9.9%**

\

K-NEAREST NEIGHBOUR (KNN) MODEL PROVIDED A BASELINE

RANDOM CLASSIFICATION

Since there are a total of 8 classes, random classification would correspond to an accuracy of 12.5%

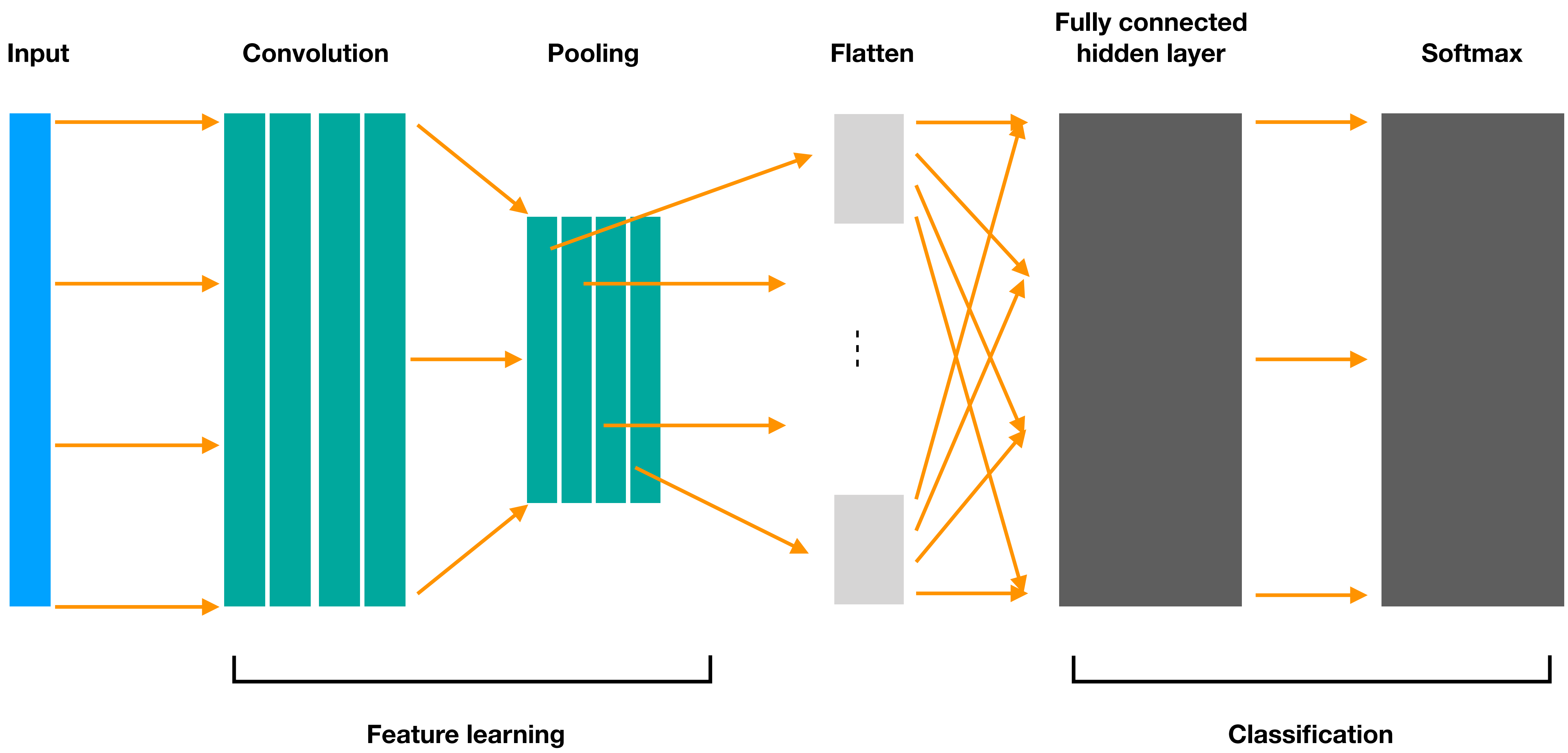
KNN MODEL

- KNN model was constructed using sklearn's KNeighborsClassifier
- This "raw pixel" model was neither scaled (i.e. divided by 255) nor was it normalised (i.e. adjusted histogram)
- The resulting overall raw pixel test accuracy is ~ 23%, an improvement over the 12.5% baseline
- Confusion matrix for KNN test data:

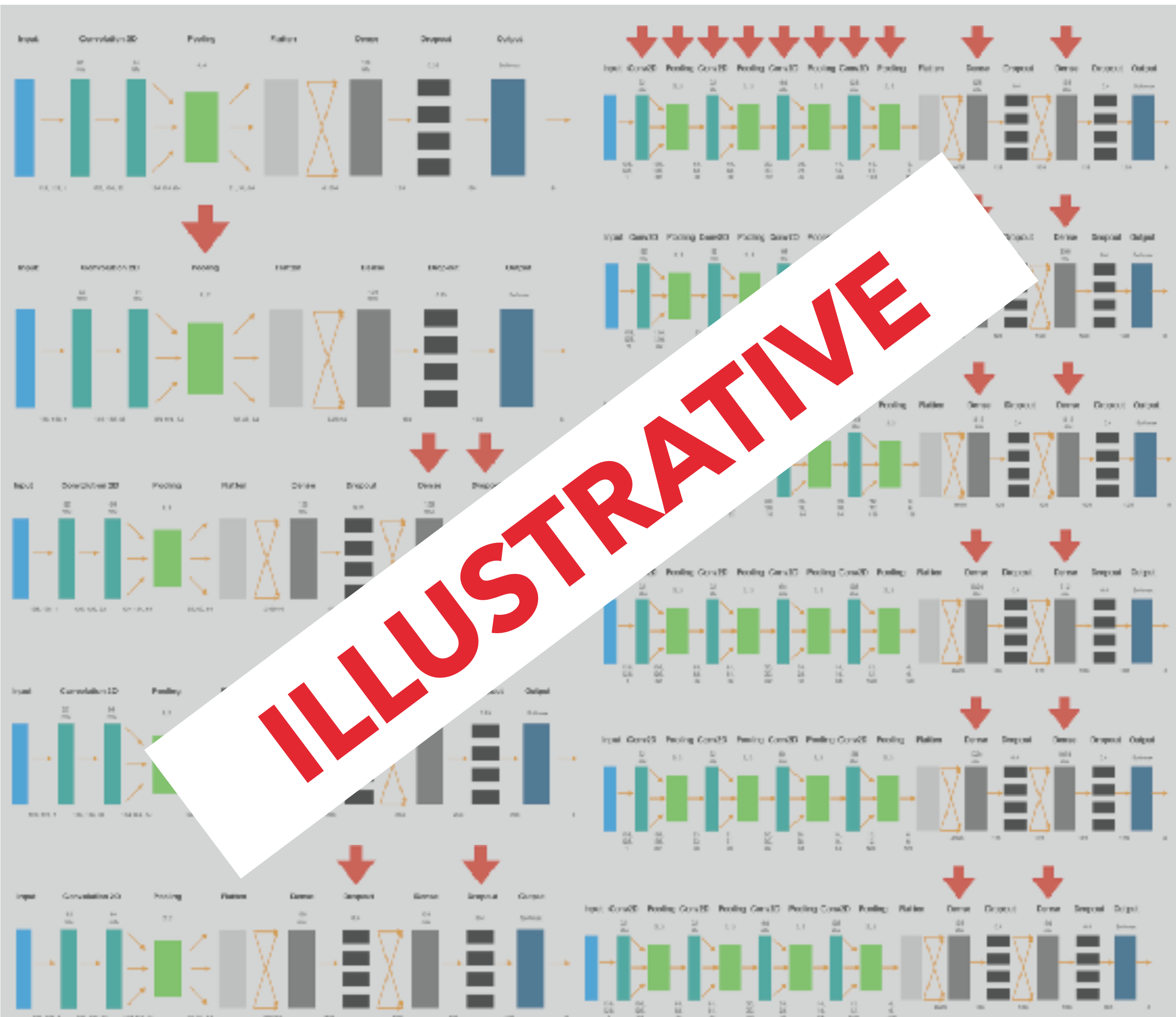
	neutral	anger	contempt	disgust	fear	happiness	sadness	surprise
neutral	111	8	4	0	2	1	4	2
anger	61	5	4	5	7	4	17	0
contempt	5	3	0	0	2	0	0	0
disgust	35	2	3	1	2	21	0	0
fear	23	0	0	0	0	0	0	10
happiness	76	2	0	17	0	3	5	3
sadness	33	10	7	2	0	0	0	5
surprise	45	2	0	10	2	2	0	13

A CNN MODEL WAS USED TO IMPROVE PERFORMANCE

GENERAL ARCHITECTURE FOR COMPUTER VISION



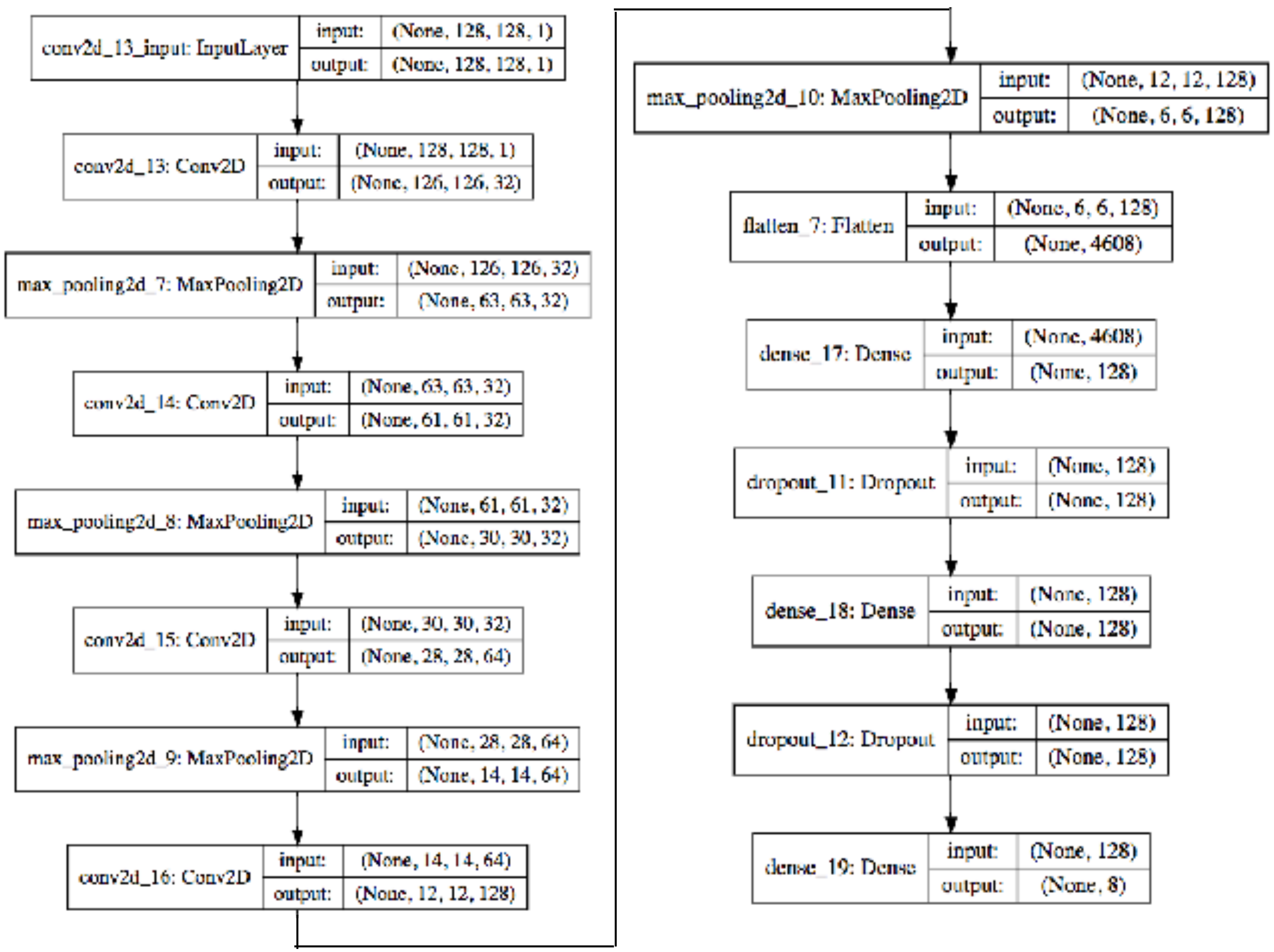
11 CNN ARCHITECTURE VARIANTS WERE USED



- Classifier architectures 1-5 had 2 convolutional 2D layers with varying number of nodes in the dense layers
- Classifier 6 had 4 convolutional 2D layers followed by a dense layer with few nodes. This architecture was inspired by a similar project by Gábor Vecsei (https://github.com/gaborvecsei/Emotion-Recognition)
- Architectures 7 to 11 were based off classifier 6 but with varying number of nodes in the hidden layers
- Training used a maximum of 20 epochs, batch size of 128, a validation split of 0.1 and an early stopping function with a patience of 10 epochs based on validation accuracy

CLASSIFIER 6 HAD THE BEST PERFORMANCE

Architecture



OVERALL ACCURACY: 69.6%

IMPROVING INPUT DATA & LABELS BOOSTED PERFORMANCE

Improving labels

- To address mislabeling, a mini dataset of only first 2 frames (neutral) and last 2 frames (peak expression) was created
- The mini dataset comprises of the following:

	Training	Test
neutral	588	66
anger	80	10
contempt	32	4
disgust	106	12
fear	44	6
happiness	124	14
sadness	48	8
surprise	154	12

Improving data

- The dataset is the same as the mini dataset with improved labels
- To improve the input data, the pixel intensities were normalised (i.e. divided by 255)
- X values now ranged from 0 to 1

MODEL PERFORMANCE IMPROVED WITH DATA REFINEMENT

Classifier 6 with full dataset

	0	1	2	3	4	5	6	7
neutral	116	4	0	1	0	1	4	6
anger	37	50	0	7	0	0	7	2
contempt	4	0	1	0	3	2	0	0
disgust	7	0	0	57	0	0	0	0
fear	7	0	0	0	12	0	3	11
happiness	19	0	0	2	4	80	0	1
sadness	32	1	0	0	5	0	19	0
surprise	6	0	0	0	0	0	0	68

OVERALL ACCURACY: 69.6%
OVERALL PRECISION: 74.9%

	Precision	Recall
neutral	50.9%	87.9%
anger	90.9%	48.5%
contempt	100%	10.0%
disgust	85.1%	89.1%
fear	50.0%	36.4%
happiness	96.4%	75.5%
sadness	57.6%	33.3%
surprise	77.3%	91.9%

with mini-dataset

0	1	2	3	4	5	6	7
66	0	0	0	0	0	0	0
0	9	0	0	0	0	1	0
1	0	3	0	0	0	0	0
0	1	1	10	0	0	0	0
0	0	0	0	3	0	1	2
2	0	0	2	1	9	0	0
4	2	0	0	0	0	2	0
0	0	0	0	0	0	0	12

86.4%
85.6%

Precision	Recall
90.4%	100%
75.0%	90.0%
75.0%	75.0%
83.3%	83.3%
75.0%	50.0%
100%	64.3%
50.0%	25.0%
85.7%	100%

with normalised mini-dataset

0	1	2	3	4	5	6	7
66	0	0	0	0	0	0	0
1	9	0	0	0	0	0	0
2	0	2	0	0	0	0	0
0	0	0	12	0	0	0	0
0	0	0	0	4	0	0	2
2	0	0	0	1	11	0	0
2	2	0	0	0	0	4	0
0	0	0	0	0	0	0	12

90.9%
91.6%

Precision	Recall
90.4%	100%
81.8%	90.0%
100%	50%
100%	100%
80%	66.7%
100%	78.6%
100%	50%
85.7%	100%

FUTURE IMPROVEMENTS

To further improve the accuracy of classification, the following actions can be taken:

1. Train the model on more images, particularly classes with fewer samples
2. Add semantic ratings
3. Outputs the probability of a class and label ambiguous results for manual verification
4. Adjust the image's histogram
5. Avoid cropping off parts of the face by identifying more facial landmarks
6. Using a multi-level classification approach where things like gender and race are determined before applying a more sensitive emotional classifier (for example: https://www.wired.com/story/how-coders-are-fighting-bias-in-facial-recognition-software/)

SUMMARY

The final model achieved an accuracy of ~91%. Whilst this level of accuracy might be insufficient for a fully automated labelling system, flagging ambiguous images for manual labelling could provide a very workable compromise of a computer aided solution.

This reports covers aspects of data acquisition from the CK+ dataset, facial image processing and the training and evaluation of KNN and CNN classification models.

The project uses OpenCV for the image processing and Keras' deep learning for the model.

For more information, please refer to:

https://github.com/chriskhoo/machine_vision