Optimising Facial Keypoint Detection With Deep Learning

Alastair Breeze - Supervised by Stephen McGough, Noura Al Moubayed

Durham University

Facial Keypoints Problem

The facial keypoints problem stems from a branch of computer vision for detecting point of interest locations.

Given an image of a face, pinpoint feature keypoints Fig. 1;4. A competition was standardised at kaggle.com as the 'Kaggle Facial Keypoints Challenge', benchmarking researchers in a leaderboard. The challenge drew interest from industry players such as Google employees and CERN, all competing to reduce root mean squared error RMSE.

Potential applications include: emotion tracking, biometric analysis, generalised keypoint detection and medical diagnosis.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Fig. 4: Prediction examples

What is Deep Learning

In recent years, huge steps in machine learning have been made, thanks to advancements in deep learning.

Deep learning is based around building algorithms that can represent high level abstractions in data through algorithmic layering. It has enabled the representation of higher level patterns in neural networks, at reduced computational cost.

Similar to [5], we developed a convolutional neural network architecture (CNN) [4], made powerful by its translation invariant feature detection.

We explored depth, as it was shown to represent higher level features; and have reduced error in similar architectures [6].

Our Model

We built a CNN Fig. 5 with a Stochastic Gradient Descent algorithm to supervised train our neural network [2]:

$$w_{t+1} = w_t - \mu_t \nabla_w Q(z_t, w_t),$$

Where w_t is a network parameter state at iteration t, and we descend the cost function $Q(z_t, w_t)$ of mini-batch z_t with respect to network weights ∇_w, at learning rate μ_t towards minimum cost [2].

Regularisation is used to avoid over-fitting, where the model fits the training data over the general case. L1 & L2 [2] regularise network weights; dropout [7] uses natural phenomena to add randomisation to the networks; early stopping prevents over-training.

The quality of images vary, ranging from over-exposure to rotations. We tackled image processing as shown in Fig. 2. We also augmented the dataset through rotations, increasing size and diversity of our training set [3].

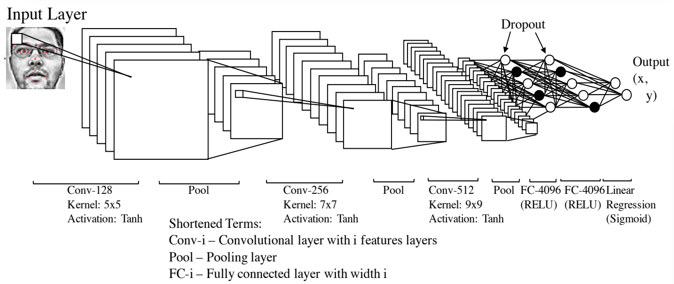
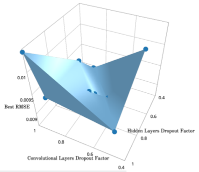


Fig. 5: Convolutional neural network model used in our system

Results

Random grid search [1] helped to quickly optimise parameter configurations with the help of 3D graphs.



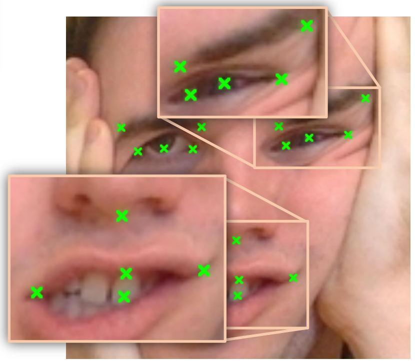
Our model reached 2nd on the worldwide leader-board of researchers (Dec. 15):

#	Δ1w	Team Name * in the money	Score	Entries	Last Submission UTC (Best – Last Submission)
1	—	Jesse Brizzi *	1.83666	29	Sat, 28 Nov 2015 02:41:05
2	↑1	**Alastair Breeze**	**1.85774**	**10**	**Thu, 31 Dec 2015 16:52:05**

Heat map visualisations gave an insight into how the model learned [8]. We investigated an occlusion algorithm Fig. 3, comparing censors and developing a novel extension that prevents additional edges incurred by the censor itself.

Fig. 2: Pre-processing before and after

Fig. 1: Prediction example from live stream tester



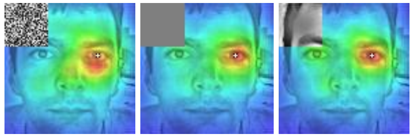


Fig. 3: Random, Grey, Faded-Grey Censors

Conclusion

We validated the CNN model for facial keypoints detection. Fast propagations proved useful when applied to time critical applications, like real time video systems Fig. 1.

Regularisation and pre-processing methods helped in reducing error. Our novel improvement to the visualisation algorithm Fig. 3 showed promise for better understanding models.

Future research could entail scaling down the model to a more portable machine and reducing error for outlier cases.

References

1. *Bergstra & Bengio, 'Random search for hyper-parameter optimization', J. Mach. Learn. Res. 13, 281–305 (2012)*
2. *Bottou, 'Neural Networks: Tricks of the Trade', Vol. 7700 Lecture Notes in Computer Science, Springer Berlin Heidelberg, pp. 421–436 (2012)*
3. *Kimura et al. 'Facial point detection based on a convolutional neural network with optimal mini-batch procedure', IEEE ICIP 2015*
4. *Lecun et al., 'Gradient-based learning applied to document recognition', Proc. IEEE (1998)*
5. *Nouri, 'Using convolutional neural nets to detect facial keypoints tutorial', http://danielnouri.org/notes/2014/12/17/using-convolutional-neural-nets-to-detect-facial-keypoints-tutorial/ (2014)*
6. *Simonyan & Zisserman, 'Very deep convolutional networks for large-scale image recognition', CoRR abs/1409.1556 (2014)*
7. *Srivastava et al. 'Dropout: A simple way to prevent neural networks from overfitting', Journal of Machine Learning Research 15, 1929–1958 (2014)*
8. *Zeiler & Fergus, 'Visualizing and understanding convolutional networks', CoRR abs/1311.2901 (2013)*