УДК 004.9, 616.24

Ковенко О.В., студент 4 курсу
Богач І.В., к.т.н., доцент

CREATION OF AN APPLICATION FOR REAL TIME EMOTIONS RECOGNITION ON A LOW RESOURCE MACHINIES

Вінницький національний технічний університет, м.Вінниця
Донецький національний університет імені Василя Стуса, м. Вінниця

Introduction

The improvements related to construction of better algorithms and big data availability gave a possibility to solve a huge number of vision related tasks. One of such problems is understanding and detection of facial expressions. This task is crucial for many applications including market research, making self-driving cars safer, analysis of interviews, etc. The described task is not a new one, thus many solutions exist along with datasets, including FER [1] and AffectNet [2]. However, the task becomes more challenging if we consider detecting a new class of images, the balance between accuracy, latency and size of the model, hyperparameters tuning. The aim of this work is to give exhaustive answers to all the aforementioned questions via description of a successful application serving a neural network trained on a custom dataset for emotion detection.

Actuality

The actuality of our work lies in following:

1. Collection of a medium size dataset of high-quality images of 5 biologically inspired emotions: angry, sad, neutral, surprised, happy and one internet based - ahegao.

2. The creation of an application based on a new dataset, which from our empirical experience achieves the best inference speed with a competitive accuracy.

Prior work

Relative prior work showcases both studies related to improvement of architectures for emotion detection and datasets, and ways of assembling all the components into an application. For instance, Siqueira et al. [3] show that their ESR (Ensemble with Shared Representation) model copes with a problem of unbalanced label distribution and outperforms state-of-the-art deep neural networks on FER dataset, while Hewitt et al. [4] described the creation of emotion recognition application. Similar to Hewitt et al. we used a MobileNet network as a main architecture for all the experiments. The difference between ours and aforementioned work is a post-training quantization used to compress the model and other dataset used for training.

Data creation

In order to create an application for emotion detection, two problems should be solved:

1. Face detection and localization.
2. Emotion recognition on a given patch that contains facial information.

Nevertheless, the first point is an already solved task, it was decided to detect people along with faces, thus the WIDER FACE dataset was additionally annotated with person class. In terms of emotion recognition, the data was scraped from YouTube, Instagram and other datasets such as IMDB (fig. 1).

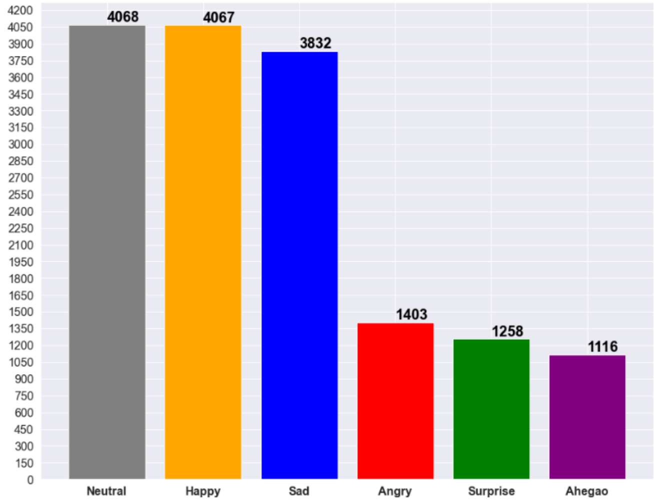

Figure 1 - Distribution of images per class in a collected dataset

Models' training

For detection of people and faces an SSD with a MobileNet V1 was used as a backbone. The model was trained out of the box with default parameters using tensorflow-zoo.

For emotion recognition the same MobileNet V1 architecture was used. It was trained with weights for imbalanced classes with a mu parameter, which affected the weight creation procedure. To account for the overfitting, dropout and L1 regularization was used along with data augmentation. MobileNet V1 has a parameter which accounts for the size of the network named alpha, in our experiments we set it to 1. For training an Adam optimization algorithm was used with a learning rate reduction schedule. The learning rate was reduced with a factor of 0.5, when results on a validation set weren't improving for more than 10 epochs. Right feature extractors N additional dense layers were added, where N was set as a hyperparameter. For choosing best hyperparameters, 10 experiments were conducted, where best ones were chosen using bayesian optimization with respect to validation precision.

Due to results of validation precision per class, the model trained during 549's experiment was chosen as the best one (fig. 2).

Precision per class

	Angry	Ahegao	Happy	Neutral	Sad	Surprise	Average
Experiment							
549	0.79	**0.98**	0.93	**0.72**	**0.77**	**0.84**	**0.84**
571	0.73	0.97	0.93	0.65	0.73	0.80	0.80
633	**0.80**	**0.98**	**0.94**	0.67	0.62	0.75	0.79
489	0.74	0.95	0.93	0.65	0.69	0.81	0.79
597	0.76	0.96	0.93	0.65	0.65	0.75	0.78

Figure 2 - Validation precision per class for top-5 experiments

Models' optimization and deployment

In order to optimize models, the quantizing technique was used. Thus weights of models were quantized from float64 to float16. This optimization reduced the size of the object detection model from **21.7 MB** to **4.51 MB**, whereas for classification one - from **69 MB** to **5.5 MB**. Having optimized models, the overall pipeline was assembled (fig. 3).

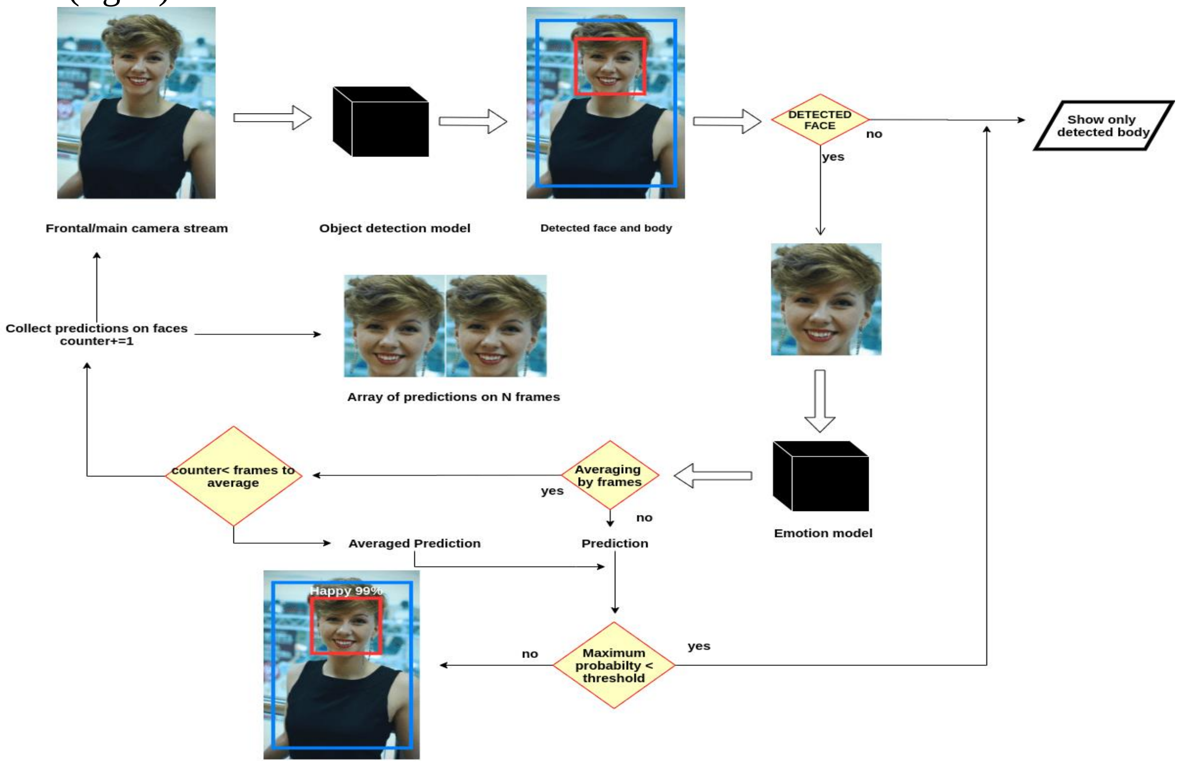

Figure 3 - pipeline of an application for mobile devices

Conclusion

In this work, the process of developing a machine learning application for expression detection was described. The discussion touched on data gathering, models' training and optimization and final application assembling. We presented a new unique medium-size dataset which can be used to solve the problem of emotion classification. Finally, the application was released in the play market [5].

References

1. *Ali Mollahosseini, Behzad Hasani, and Mohammad H. Mahoor, "AffectNet: A New Database for Facial Expression, Valence, and Arousal Computation in the Wild", IEEE Transactions on Affective Computing, 2017.*
2. *I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio. Challenges in representation learning: A report on three machine learning contests. Neural Networks, 64:59--63, 2015. Special Issue on "Deep Learning of Representations"*
3. *Henrique Siqueira, Sven Magg and Stefan Wermter. Efficient Facial Feature Learning with Wide Ensemble-based Convolutional Neural Networks. [Electronic resource] – Electronic data. – Mode of access: https://arxiv.org/pdf/1612.02903.pdf – Title from the screen.*
4. *Charlie Hewitt, Hatice Gunes. CNN-based Facial Affect Analysis on Mobile Devices. [Electronic resource] – Electronic data. – Mode of access: https://arxiv.org/pdf/1807.08775.pdf – Title from the screen.*
5. *Oahega - Emotion detector. [Electronic resource] – Electronic data. – Mode of access: https://play.google.com/store/apps/details?id=org.oahega.com – Title from the screen.*