A Neural Network Architecture for Real-Time Facial Attributes Recognition on Low-Power Edge Devices

Amin Banitalebi-Dehkordi[1], Ehsan Mohyedin Kermani, Yuri Moiseyenko, Lanjun Wang, Yong Zhang, and Jiebo Luo
Huawei Technologies, Canada
Vancouver, BC Canada
[1]amin.banitalebi@huawei.com

***Abstract*—Deploying efficient algorithms to low-power edge devices is crucial for cloud computing services. We introduce an efficient unified neural network architecture for real-time recognition of facial attributes on embedded devices. This model can recognize 10 facial attributes on low-power Huawei AI chips in real-time (10-25 fps depending on the settings and number of faces).**

Keywords—neural networks, edge computing, facial attributes

I. Introduction

Recent breakthroughs in artificial intelligence has been largely due to the success of Deep Neural Networks. Thanks to the computing power of Graphics Processing Units (GPUs), training of such networks over large-scale datasets is now practical. However, these advancements have not been fully translated to tangible products in the consumer market yet. The main obstacles are:

a) User-end device limitations: There is still a need for custom-design AI for low power embedded devices such that they can run inference models in real-time.

b) Algorithms limitations: Low-power devices often need very low-cost algorithms resulted from compact models.

In terms of devices, there are a few AI-enabled low-power boards in the consumer market currently. The capabilities of such devices are expected to get improved over time. Demonstrations here are based on using Huawei AI chips. In addition, we use the EI Wizard service on the Huawei Cloud to convert the model automatically to the format that can be run on Huawei's latest AI chips.

On the other hand, several state-of-the-art neural network architectures have been proposed to reduce the computational complexity of various tasks. MobileNet and SqueezeNet are among the reputable proposals. These compact models often come with a trade-off in the accuracy.

Facial attributes recognition is one of the most demanding applications for current AI cloud edge devices. This application requires fast enough models to run on efficient devices. Moreover, since facial attributes are correlated, it makes sense for them to share characteristics. We demonstrate the usage of a single unified multi-task neural network for predicting these attributes, using a low-power device.

II. Methodology and Results

In order to estimate all facial attributes with limited computational resources, a single multi-task neural network has been used (see Fig. 1). Facial attributes predicted using this model are: facial age, gender, smile, emotion (expression), eye/lips makeup, facial hair (sideburns, beard, moustache), and glasses type. As observed from Fig. 1, some tasks are interpreted as classification (e.g. expression), while some are regression (e.g. smile). The optimization cost function includes loss terms for each one of the attributes.

A base Convolutional Neural Network (CNN) model has been used to extract visual features. Several candidate models have been evaluated including: *Inception-ResNet-v1*, *MobileNet-v2*, *SqueezeNet*, *ResNet18*, *50*, and *101*. *ResNet18* resulted in a better trade-off between speed and accuracy than the other models.

It is also worth noting that there is no large enough dataset available which contains labels for the attributes mentioned above. Therefore, we used our in-house dataset of facial attributes for training, validation, and testing.

We conclude that the idea of having a small network performing many tasks seems to be a feasible solution for applications such as facial attribute recognition. Future work includes exploring the use of computationally heavier tasks such as action recognition on low-power edge devices.

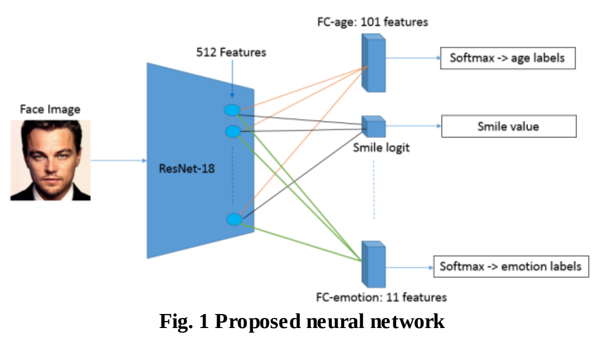

Fig. 1 Proposed neural network

Table 1 Accuracy and MAE (mean average error) for various tasks

Attribute	Accuracy	Attribute	MAE
Gender (M-F)	96 %	Age (0-100)	2.10
Expression (0-10)	87 %	Smile (0-1)	0.06
Glasses (0-2)	99 %	Beard (0-1)	0.04
MakeUp Eye (0-1)	89 %	Moustache (0-1)	0.04
MakeUp Lips (0-1)	87 %	Sideburns (0-1)	0.03