Lightweight Deep Convolutional Neural Networks for Facial Expression Recognition
Yanan Wang, Jianming Wu, Keiichiro Hoashi
KDDI Research, Inc., Japan

Objectives
- **Facial Expression Recognition (FER)** is an intuitive way to analyze human emotions
- **Deep Convolutional Neural Networks (DNN)** can capture fine-grained facial expression features to achieve high FER accuracy

To maintain the high performance and reduce computational resources, we propose a method to build a lightweight DNN based on a VGG pre-trained network for FER task

Motivation
- **Facial expressions differ among individuals**
- Some facial expressions are very hard to discriminate (ex. “fear” and “disgust”)

The fine-grained facial expression features cannot be obtained by current lightweight DNN without increasing the number of parameters (Such as MobileNets, DenseNet, Xception)

Lightweight architectures
- **Depthwise separable convolutions** [Fig 4]
  (a) Depthwise: apply a $D_x \times D_y$ kernel to each channel
  (b) Pointwise: apply a $1 \times 1$ kernel to the output of depthwise convolution

Number of standard convolution parameters $D_y \times D_y \times M \times D_z + N$
Approximately $D_z^2$ times reduction

Number of depthwise separable convolution parameters $D_x \times D_y \times M + D_y^2 + D_y \times D_x \times M + N$
Depthwise Pointwise

Proposed method
1. Pretrain a high performance VGG model for FER
   ![Fig 2. the structure of the VGG-based pre-trained network.](image)
   “C”, “P”, and “F” denote the convolution, pooling and fully connected layers

2. Conduct two re-training networks through connecting two different part of the pre-trained networks with lightweight architectures

   ![Fig 3. Two types of re-training network structures](image)
   (a) VGG-retrain-v1:
   (b) VGG-retrain-v2:

Discussion (Table 1/2)
- **Accuracy:**
  - VGG-retrain-v1/2: High accuracy comparable to VGG pretrained model (Retrain models retained the fine-grained facial expression features of VGG-retrain model)
  - Lightweight strategy:
    VGG-retrain-v2: Fewer parameters than other lightweight DNNs (Due to lightweight architectures)

Network architecture:
- VGG-retrain-v2 can reduce more parameters than VGG-retrain-v1 while attaining comparable accuracy

Conclusions
Our proposed lightweight FER method reduced parameters approximately 1/20 times that of the VGG-retrain model, and achieved the highest FER accuracy over other lightweight methods (FER2013: 69.8% and AffectNet: 61.1%)