1. Background

- Facial Action Units (AUs) represent the smallest visually observable facial muscle movements.
- An automatic system for AU recognition has applications for human-computer interaction (HCI), interactive games, computer-based learning, entertainment, and psychiatry.

2. Motivation & Related Work

2.1 Motivation

- Standard CNN-based methods employ predefined fixed and filter sizes in each convolutional layer.
- Fixed filter sizes are not optimal for all tasks and for different image resolutions.
- Different AUs causing facial appearance changes over various regions may prefer different filter sizes.

2.2 Related Work

- The best filter size is often selected experimentally or by visualization.
- Inception modules concatenate the activations from multiple filter sizes.
- Pyramid filters

The proposed OFS-CNN can learn the filter sizes of all convolutional layers simultaneously.

3. Overview

- We proposed to learn the convolutional filter size from training data.
- In the backpropagation stage, the filter size can be updated based on SGD.

4. Methodology

4.1 Definition of upper and lower bound filters

Given a continuous filter size $k$, the upper-bound filter size $k_{\text{ub}} = k + z - 1$ and the lower-bound filter size $k_{\text{lb}} = k - z + 1$.

4.2 Forward convolution with $k \in \mathbb{R}$

$y_{ij}(k) = w_{ij}(k) + (1 - a) y_{ij}(k-1) = w_{ij}(k) x_{ij} + b_{i}$

Linear interpolation of $y_i(j)$ and $y_i(j)$

4.3 Backward propagation - updating filter size

$\frac{\partial J(k)}{\partial k} = y_{ij}(k + \delta k) - y_{ij}(k - \delta k)$

$\frac{\partial J(k)}{\partial k} = 2w_{ij}(k) x_{ij} = 2w_{ij}(k) x_{ij} - 2k \Delta w_{ij}$

Based on the chain rule $\Delta w_{ij} = \sum \frac{\partial J}{\partial w_{ij}} \frac{\partial W_{ij}}{\partial w_{ij}}$

Use gradient descent to update the filter size $k^{t+1} = k^t - \eta \frac{\partial J(k)}{\partial k}$

5. Experimental Results

- OFS-CNN vs exhaustive search on SEMAINE database in terms of F1 score and 2AFC (Area under ROC) using mean std from 5 runs.

6. Conclusion

- We proposed a novel OFS convolution layer to learn the filter size from the training data.
- Achieved comparable performance as the best filter size by the exhaustive search with much less training time.
- The OFS-CNN can learn adaptive filter sizes in different resolution of input images.