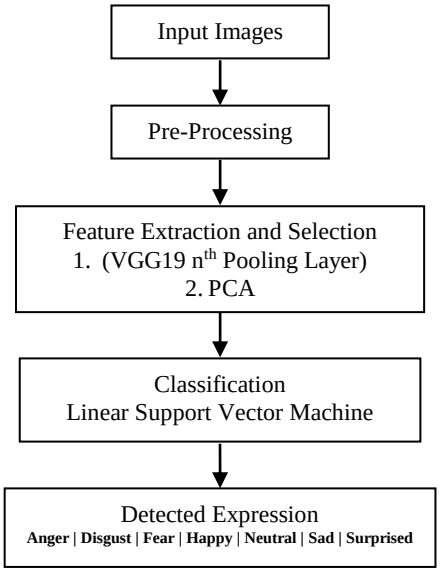
Pre-Trained Convolutional Neural Network Features for Facial Expression Recognition

Aravind Ravi
Department of Systems Design Engineering, University of Waterloo, Ontario CA
aravind.ravi@uwaterloo.ca

***Abstract*— Facial expression recognition has been an active area in computer vision with application areas including animation, social robots, personalized banking, etc. In this study, we explore the problem of image classification for detecting facial expressions based on features extracted from pre-trained convolutional neural networks trained on ImageNet database. Features are extracted, and transferred to a Linear Support Vector Machine for classification. All experiments are performed on two publicly available datasets such as JAFFE and CK+ database. The results show that representations learnt from pre-trained networks for a task such as object recognition can be transferred, and used for facial expression recognition. Furthermore, for a small dataset, using features from earlier layers of the VGG19 network provides better classification accuracy.**

I. System Design



II. Dataset

A. *JAFFE – Japanese Female Facial Expression*

Number of Subjects: 10
Number of Images: 213
Number of Expressions: 7

B. *CK+ – Cohn Kanade Dataset (Subset)*

Number of Subjects: 10
Number of Images: 210
Number of Expressions: 7

A_{JK} – Training accuracy using Jack-Knife Method
A_T – Testing accuracy
N_{PCA} – Number of PCA components

III. Methodology

A. *Feature Extraction*

1. Features from each pooling layer and the fc1 layer of VGG19 are extracted
2. Reduce dimensionality based on chosen N_{PCA} {50,100,150,200}

B. *Feature Selection*

This is based on two parameters:
1. N^{th} VGG19 Layer for feature extraction
2. Find N_{PCA}

Method:
1. Select the parameters of the two highest A_{JK} values.
2. Select the final parameters based on the configuration with the least difference between the A_{JK} and A_T

C. *Classification*

A Support Vector Machine with a linear kernel and C=10 is used as the classifier.

D. *Validation*

The training data for both datasets was split into 80% for training the classifier and 20% was used for testing. These results are validated based on a 10-fold cross-validation, and due to the small size of the datasets, a Jack-Knife validation or leave-one-out validation is also performed.

IV. Results

Table 1 summarizes the training and test accuracies for the chosen number of N_{PCA}. Among the different methods of feature extraction, features from Block4 pool layer of VGG19 provides the highest accuracy for both CK+ and JAFFE dataset. Applying the proposed feature selection methods, an A_{JK} of 92.77% and A_T of 92.86% was achieved for the JAFFE dataset. An A_{JK} of 92.26% and A_T of 92.86% was achieved on the subset of CK+ dataset. It can be observed from the A_{JK} values wherein earlier layers of the CNN provide a better performance than later layers.

TABLE I. TRAINING AND TEST ACCURACIES FOR SELECTED FEATURES OF CK+ AND JAFFE DATASETS

	CK+ Dataset			JAFFE Dataset		
Features	PCA - 100			PCA – 200		
VGG19 Layer (POOL)	Training (80%)		Test (20%)	Training (80%)		Test (20%)
	10-Fold	A_{JK}		10-Fold	A_{JK}	
block1	87.90	88.69	85.71	77.27	79.52	88.10
block2	90.27	90.48	85.71	81.66	83.73	88.10
block3	91.51	93.45	90.48	90.47	89.76	92.86
block4	**94.93**	**92.26**	**92.86**	**92.70**	**92.77**	**92.86**
block5	88.83	90.48	90.48	81.35	82.53	85.71
(Dense) FC1	91.76	89.88	92.86	81.01	76.51	88.10