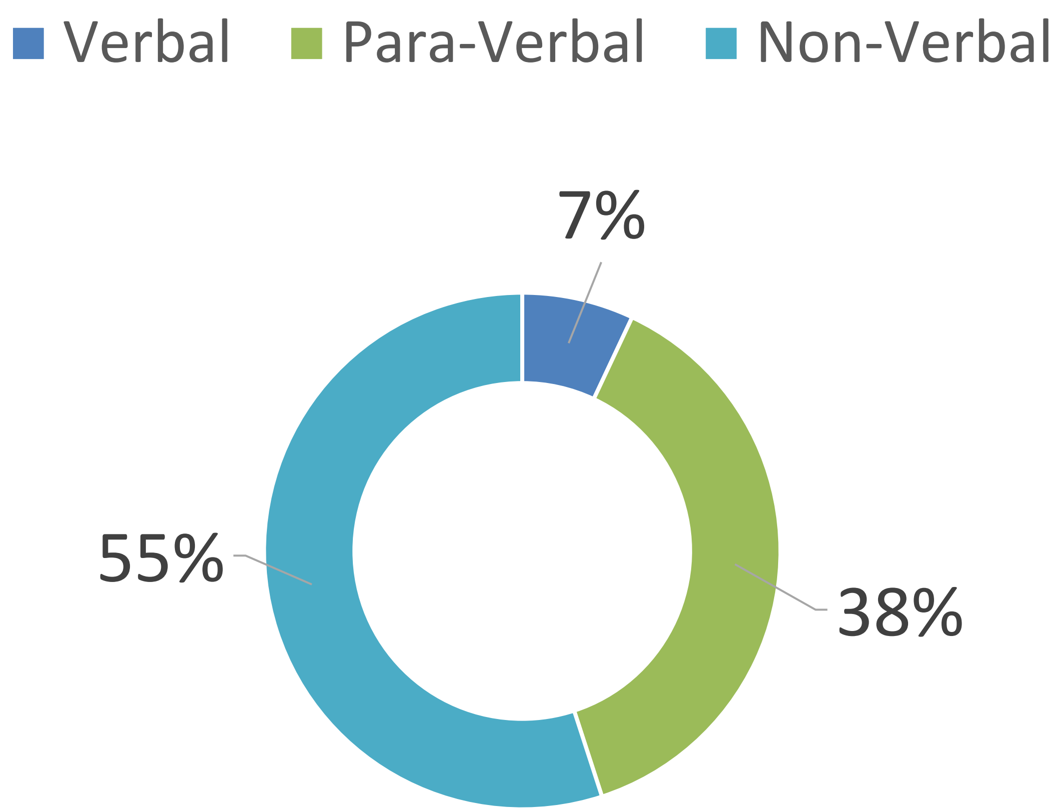
Assessing Accuracy of Ensemble Learning for Facial Expression Recognition

Alessandro Renda[1,2], Marco Barsacchi[1,2], Alessio Bechini[1], Francesco Marcelloni[1]

[1]University of Pisa, Dept. Of Information Engineering

[2]University of Florence, Dept. Of Information Engineering

How We Communicate Our Feelings



Facial Expressions of Emotions

Paul Ekman and Friesen Wallace
Constants across cultures in the face and emotion
Journal of personality and social psychology 17.2 (1971): 124.

- **Universality** of Facial Expressions of Emotion
- Definition of a List of **Basic Emotions**

from http://ocw.mit.edu/courses/brain-and-cognitive-sciences/9-00sc-introduction-to-psychology-fall-2011/emotion-motivation/discussion-emotion

Automatic Facial Expression Recognition

A challenging task:
- Facial Expressions Recognition (FER) from **facial images *in-the-wild***

Fields of **Applications**:
- Human Comuputer Interaction
- Sentiment Analysis
- Behaviomedics
- Deceit Detection
- Emotional Health
- Data Analyitics

Outline

1. Introduction to the traditional and modern approaches to the FER problem
2. Motivations and Objectives
3. Our Experimental Setup
   - FER2013 *in-the-wild* dataset
   - Ensemble Design Strategies under different scenarios
4. Results and Conclusions

Traditional approaches to the multi-class Image Classification Problem

Pre-Processing

- Face Detection
- Face Registration

↓

Features Extraction

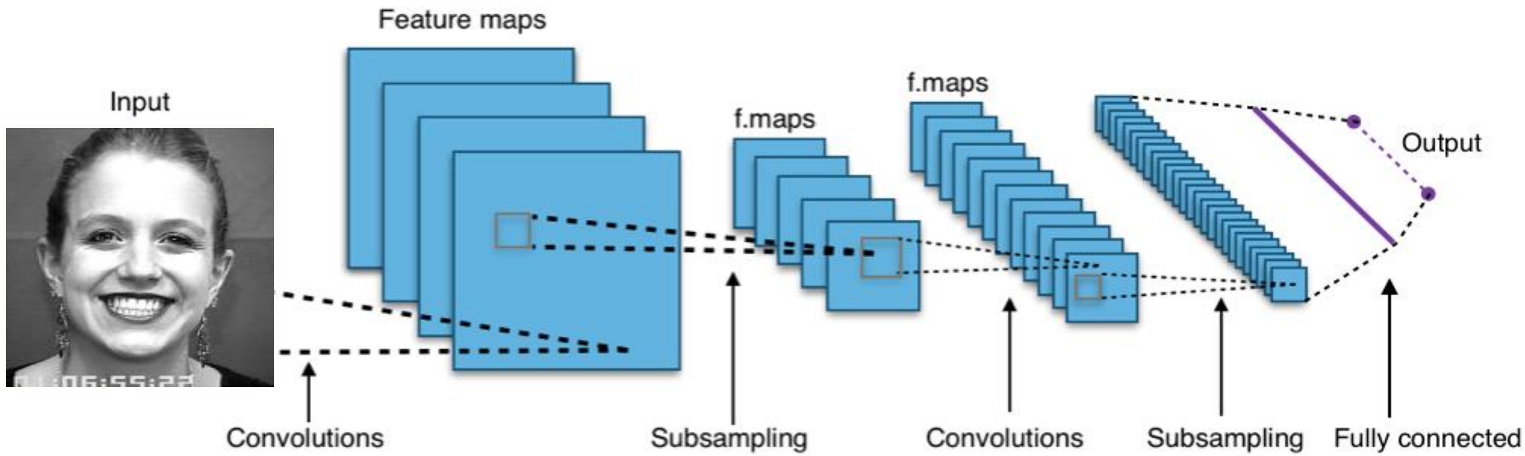
- **Hand Crafted Features**
- Useful, Discriminative Representation

↓

Classification

- N-Way Classification

Deep Learning approach



Convolutional Neural Networks (CNNs)

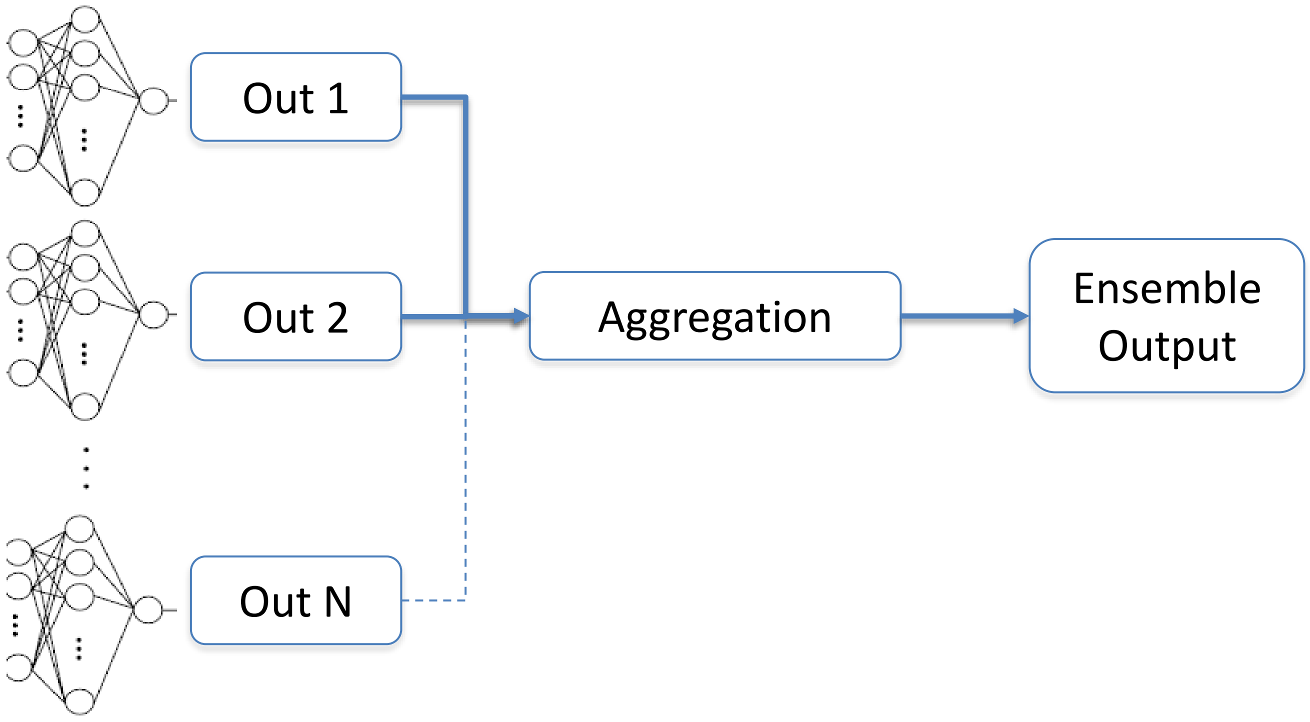
- Automatic learning a **hierarchical features representation**
- **Excellent results** in a wide variety of similar problems
- Represent state-of-the-art also for FER

Drawback:

- Typically rely on ***large* collection of labeled data** for training
- Available FER datasets have limited size

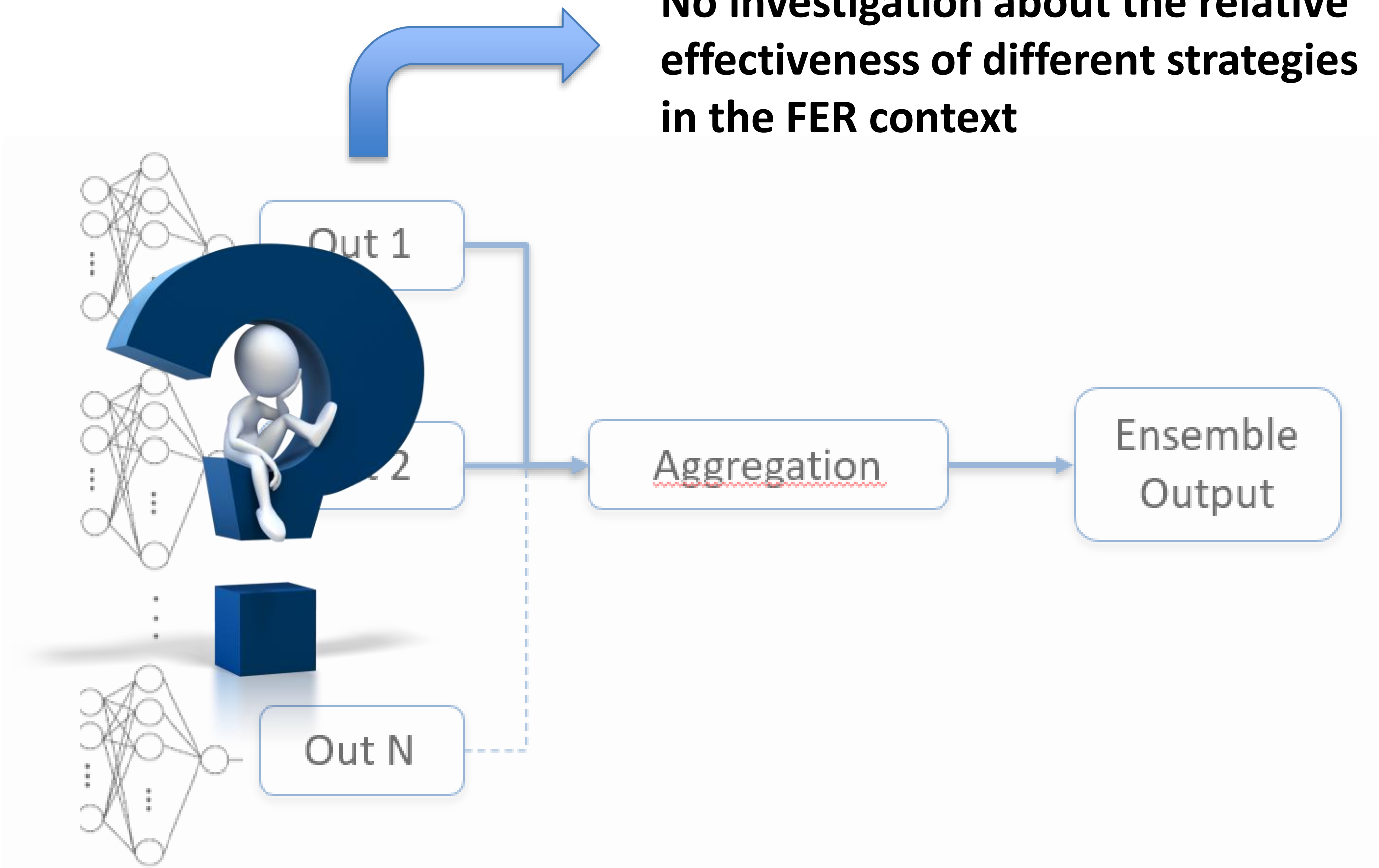
Ensemble Techniques

- Widely exploited in Neural Networks to **boost classification performances**
- Exploit diversity of base classifiers



Ensemble Techniques

No investigation about the relative effectiveness of different strategies in the FER context



Objectives of our work

- Tackling FER problem
  exploiting Ensembles of Deep Convolutional Neural Networks

- Comparative study: Assessing accuracy of two simple techniques to generate diversity across the base classifiers of an ensemble

- Medium-size dataset: Considering two distinct scenarios:
  1. Training from scratch an ad-hoc architecture
  2. Fine-tuning a pre-trained state-of-the-art model

FER-2013: Facial Expressions Dataset

- One of the **largest collection of *in-the-wild* facial images**
- Consisting in **35.876 images** from **7 classes:**

Training Set	28699
Validation Set	3588
Test Set	3589

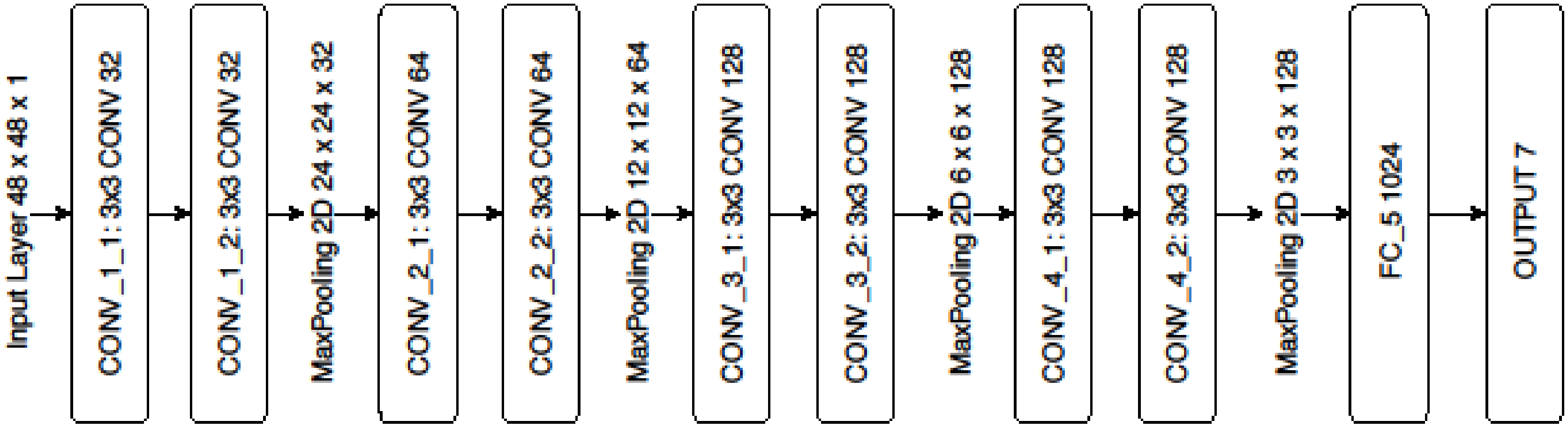
Neutral	6197
Anger	4945
Disgust	547
Fear	5121
Happiness	8988
Sadness	6076
Surprise	4001

- Average Human Accuracy: 65%
- State-of-the-art Accuracy: 75.2% [1]

[1] C. Pramerdorfer and M. Kampel, "Facial expression recognition using convolutional neural networks: State of the art," arXiv preprint arXiv:1612.02903, 2016

First Scenario: CNN10-S

- **Training from scratch an Ad Hoc Architecture**
  - 10 layers, ~ 1.7M parameters
  - Mimics VGG-B architecture, inspired to [1]

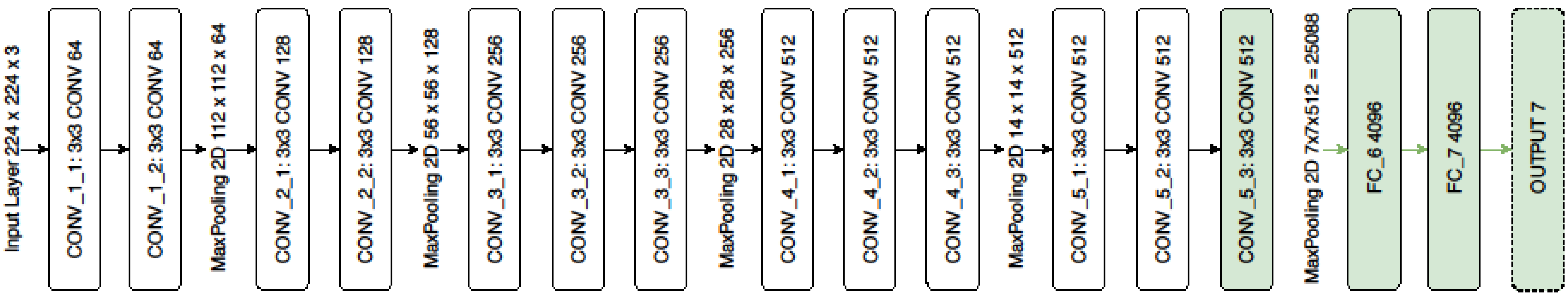


[1] C. Pramerdorfer and M. Kampel, "Facial expression recognition using convolutional neural networks: State of the art," arXiv preprint arXiv:1612.02903, 2016

Second Scenario: VGG16-FT

- **Fine-tuning a Pretrained Model**
  - VGG16 architecture
  - 16 layers, ~130M parameters
  - *Pretrained* on a Face Recognition dataset [1] of 2.6M images

A) Remove original output layer. Add custom output layer (7 units)
B) Train the output layer. Freeze all hidden layers
C) Fine-tune the final (green) layers. Freeze the white layers



O.M. Parkhi et al., Deep face recognition. *BMVC*. Vol. 1. No. 3. 2015.

Two Ensemble Design Strategies: SE - PS

- Fixed size ensemble of **nine networks**
- **Three repetitions** for each strategy and for each scenario

Seed Strategy (SE)

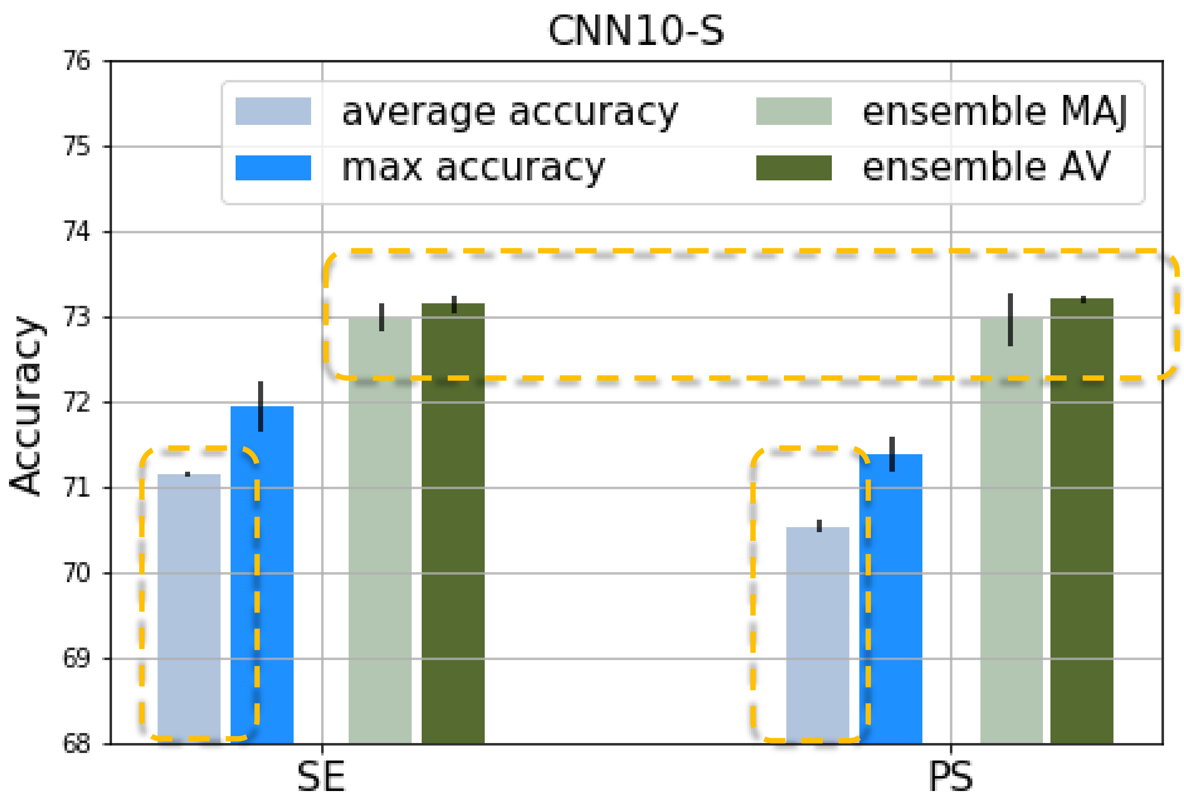
SEED 1	NET 1
SEED 2	NET 2
SEED 3	NET 3
SEED 4	NET 4
SEED 5	NET 5
SEED 6	NET 6
SEED 7	NET 7
SEED 8	NET 8
SEED 9	NET 9

Preprocessing Strtegy (PS) [1]

	SEEED 1	**SEEED 2**	**SEED 3**
DEFAULT 	NET 1	NET 2	NET 3
HISTEQ 	NET 4	NET 5	NET 6
INOR 	NET 7	NET 8	NET 9

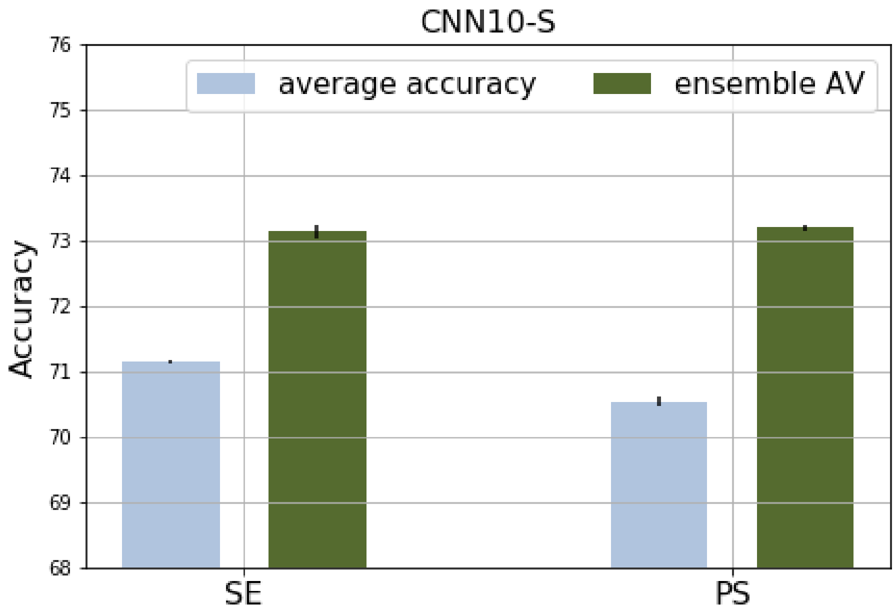
[1] B.-K. Kim et al., Fusing aligned and non-aligned face information for automatic affect recognition in the wild: A deep learning approach, Proc. Of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops, 2016, pp. 1499{1508. doi:10.1109/CVPRW.2016.187.

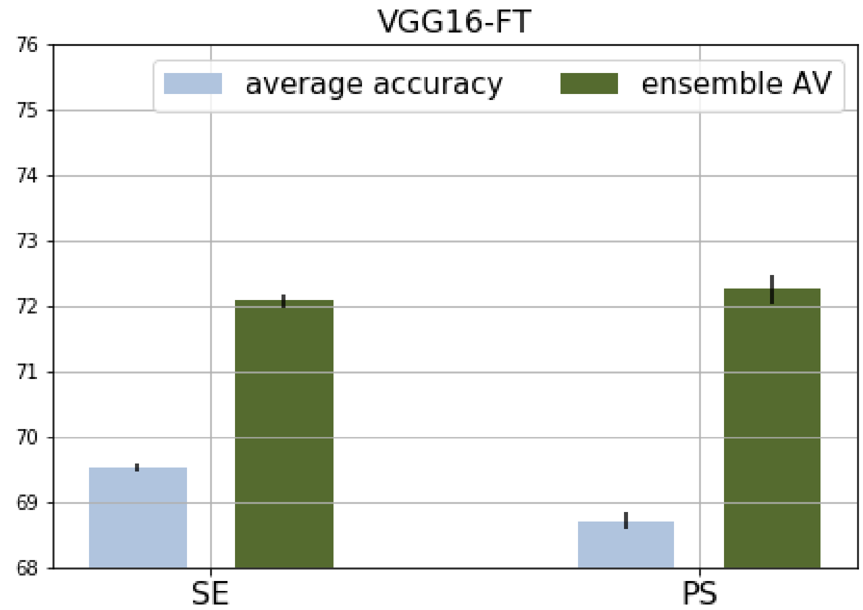
Experimental Results



- Average and Majority voting lead to comparable results
  - The higher variability among networks, the higer ensemble gain

Experimental Results

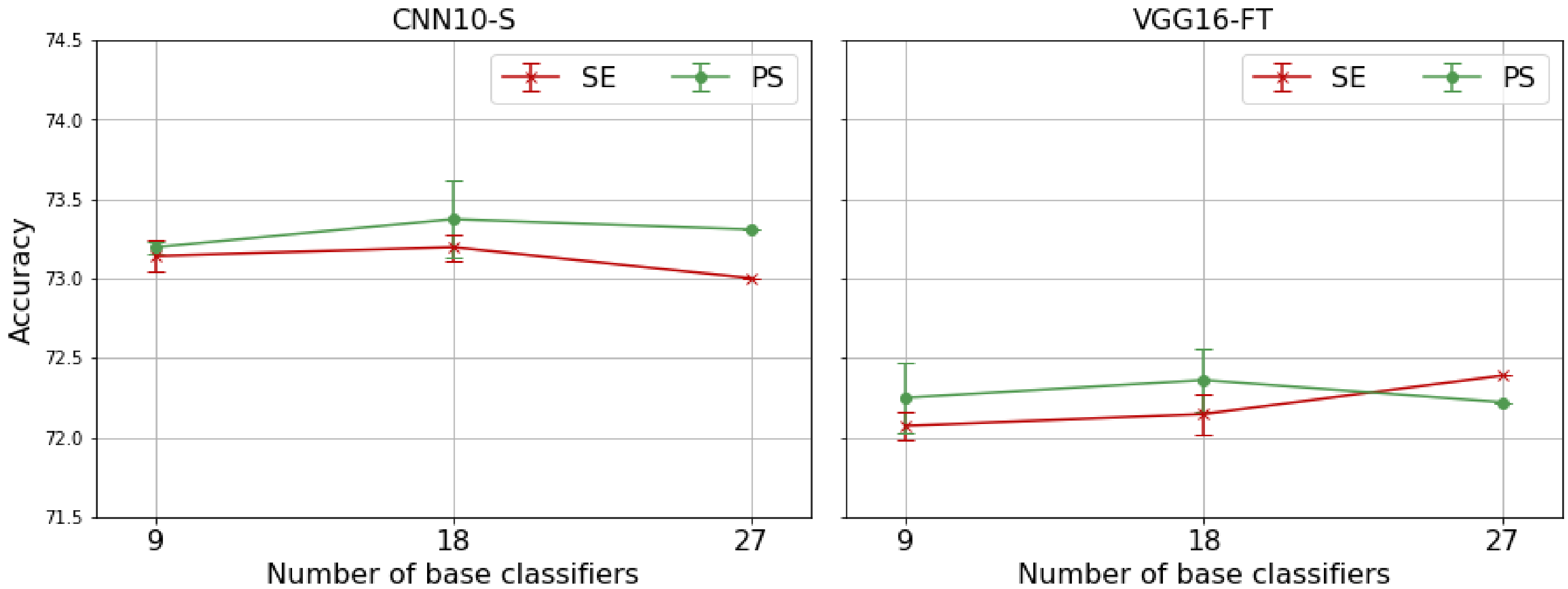




- CNN10 **trained from scratch outperforms the fine tunend** VGG16 model
  - in terms of Base Classifiers
  - In terms of Ensemble Classifiers

Increasing the number of base classifiers

- **3 repetitions** for each strategy, for each scenario → 27 networks



- In general, **no significant benefit** when increasing ensemble size

Conclusions

- Task: *in-the-wild* Facial Expression Recognition
- Assessing the accuracy of different approaches of Ensemble Learning:
  - **Two ensemble design strategies** (SEED vs PREPROCESSING) **achieve comparable results**
  - Two training scenarios (CNN10-S vs VGG16-FT): **Training an ad hoc model from scratch is an appropriate choice in the considered setting**

- Further investigation:
  - Other state-of-the-art models
  - Other pretraining datasets
  - Other factors of variation in the ensemble

Thank you for your attention

Alessandro Renda

PhD program in Smart Computing

Universities of Pisa, Firenze, Siena

Email: alessandro.renda@unifi.it