Mood Detection for Improving Lifestyle of Older Adults in Ambient Assisted Living Contexts

Andrea Caroppo, Alessandro Leone, Pietro Siciliano

National Research Council of Italy - Institute for Microelectronics and Microsystems

Lecce - Italy

andrea.caroppo@cnr.it

alessandro.leone@cnr.it

pietro.siciliano@le.imm.cnr.it

IMM Institute for Microelectronics and Microsystems

Consiglio Nazionale delle Ricerche

SPWID 2020

Introduction and Motivation (1/2)

WHY: Detection of mood through sequences of facial expression for the evaluation of emotional pattern → assess cognitive deterioration/prevent depression in frailty people and/or ageing adults

OBJECTIVES: design and development of new tools and /or technologies for the automatic recognition of emotion or moods performed by older adults in AAL context

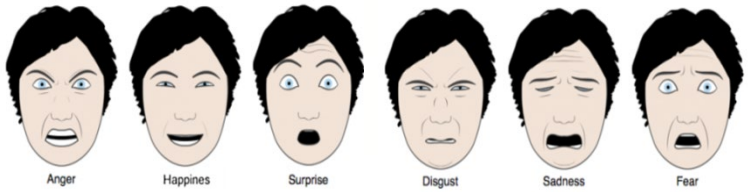
CHALLENGES: existing approaches lack generalizability, effective classifiers ignore the effects of age of the observed subject, very different light conditions and poses

Introduction and Motivation (2/2)

Applications

- Robotics
- Communications
- Security
- Medical and Assistive Technology

6 basic classes for FER (Ekman's classification)



FER main stages: FACE DETECTION → FEATURE EXTRACTION → CLASSIFICATION

Handcrafted features

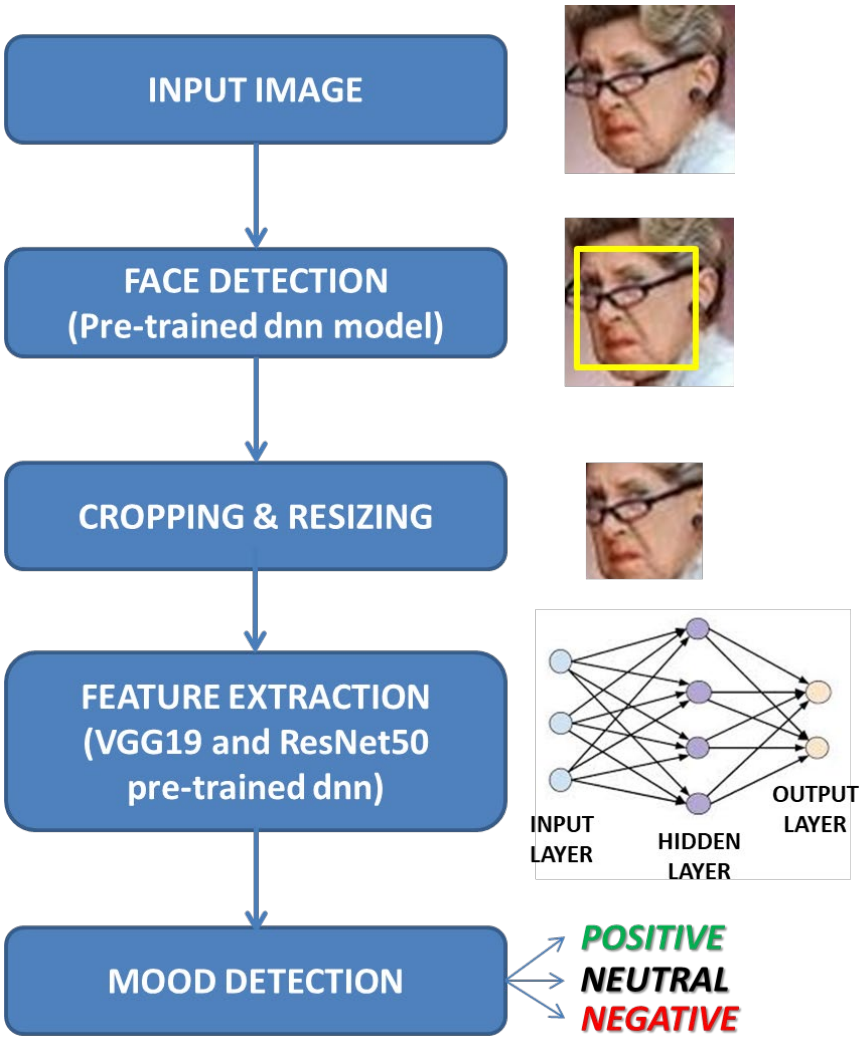
- GEOMETRIC-BASED
- APPEARANCE-BASED
- HYBRID-BASED

Machine learning techniques based on convolutional neural networks (CNNs)

1) Lack of training samples for FER in the wild
2) Datasets for FER at varying of age

Transfer Learning

Overview of the proposed mood detection system



Pre-processing

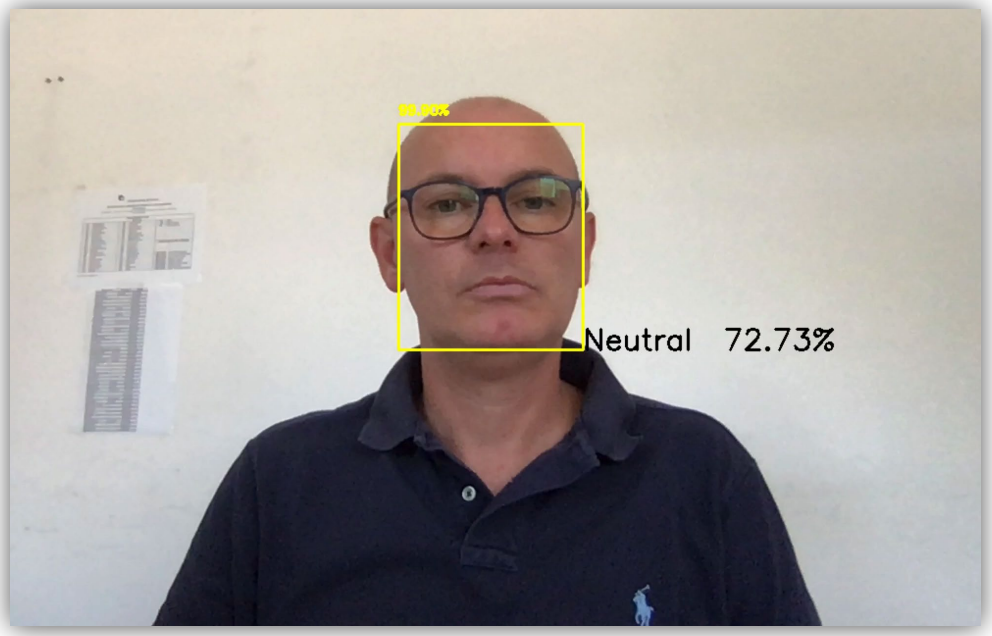
Faces generally occupy very little area -> in AAL context, faces can look very different depending on orientation and pose

Viola-Jones face detector has limitations for multi-view face detection!

Starting from vers 3.3, OpenCV ships out-of-the-box with a more accurate face detector (as compared to OpenCV's Haar cascades).

Ability to detect faces "in the wild" in real-time even if a PC without GPU is used for the processing

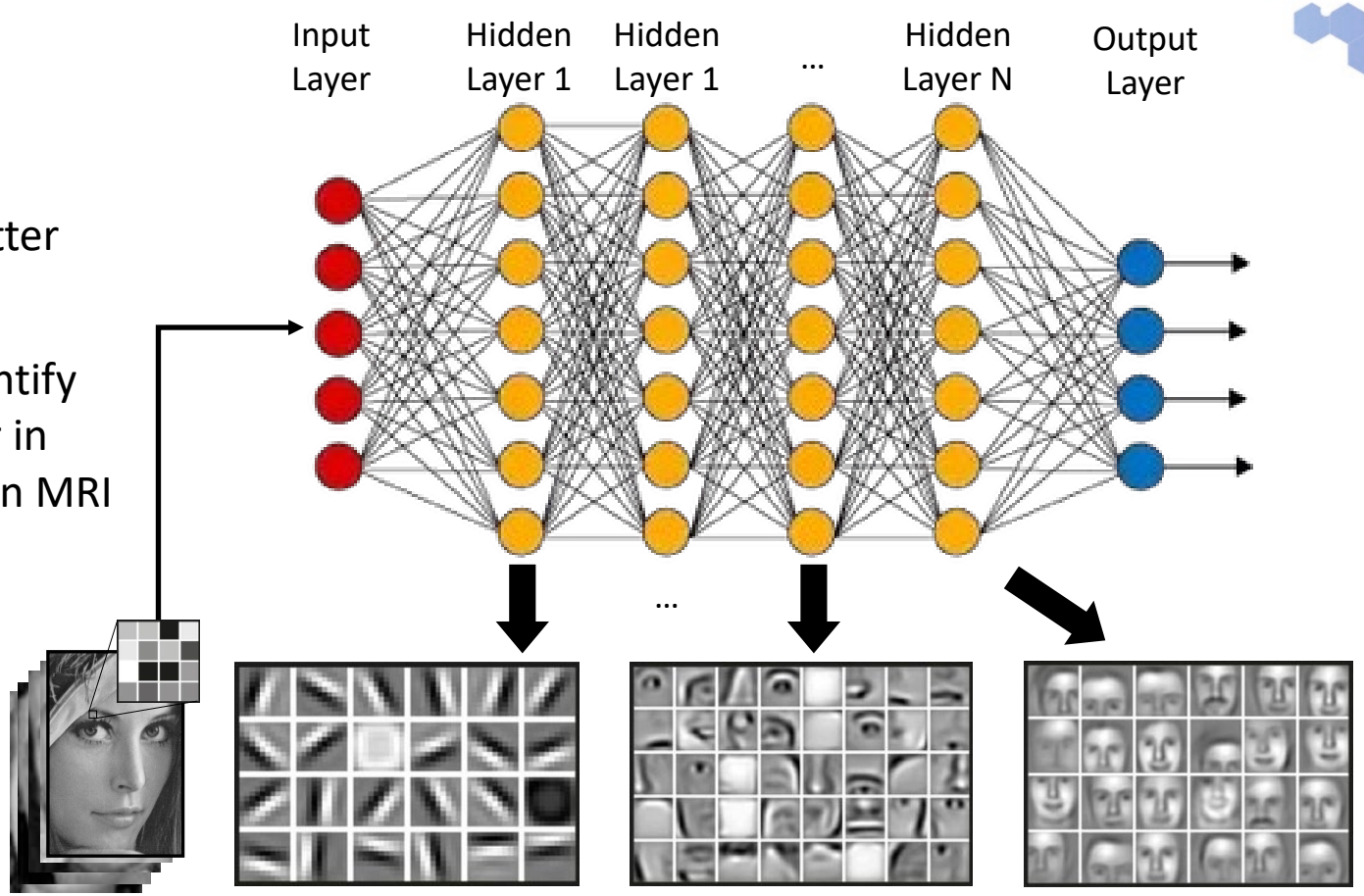
SSD framework (Single Shot MultiBox Detector) with a reduced ResNet-10 model



a) BOUNDING BOX OF THE FACIAL REGION
b) CONFIDENCE INDEX

Feature extraction (1/2)

- Deep learning (DL) is an emerging approach in which features are learned from data using a **general-purpose learning procedure**, involving multiple levels of representation of data.
- Several DL-based computer vision applications are performing even better than humans.
- They are able to identify indicators for cancer in blood and tumours in MRI scans.
- Applications: object detection, speech recognition, face recognition, medical imaging.



Feature extraction (2/2)

CNN for FER encloses three learning stages in just one framework

1) FEATURE LEARNING

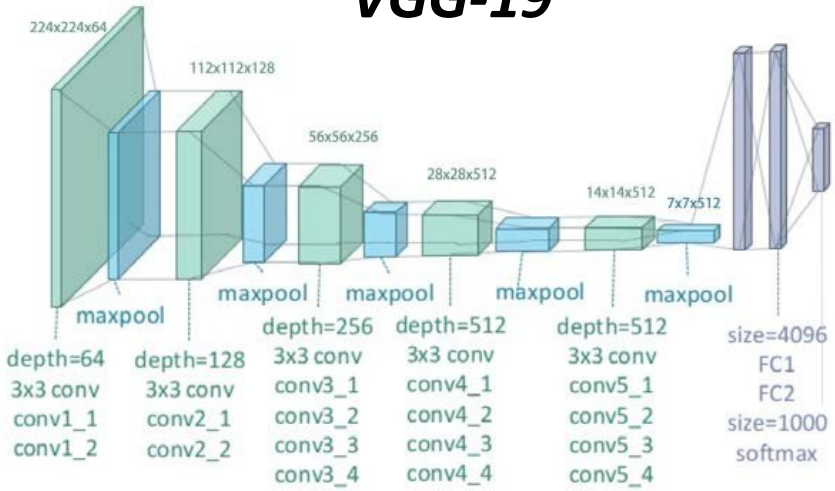
2) FEATURE SELECTION

3) CLASSIFIER CONSTRUCTION

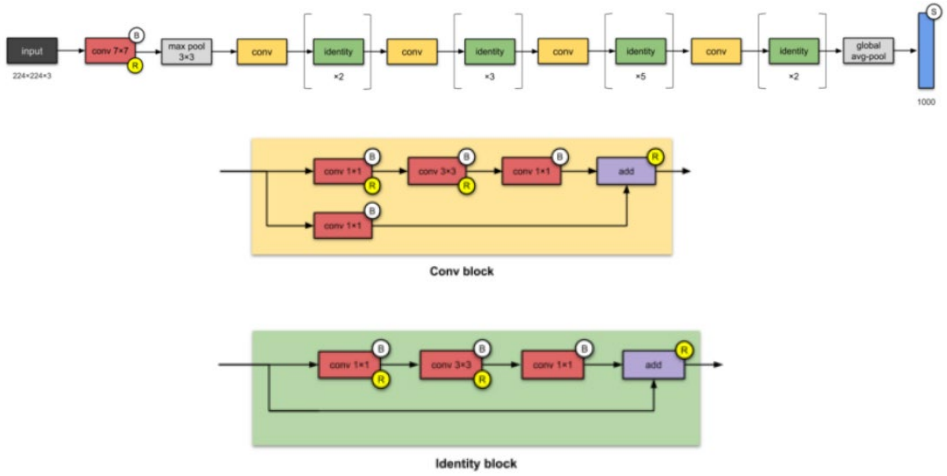
Creating a CNN from scratch is not an easy task!!!

TRANSFER LEARNING

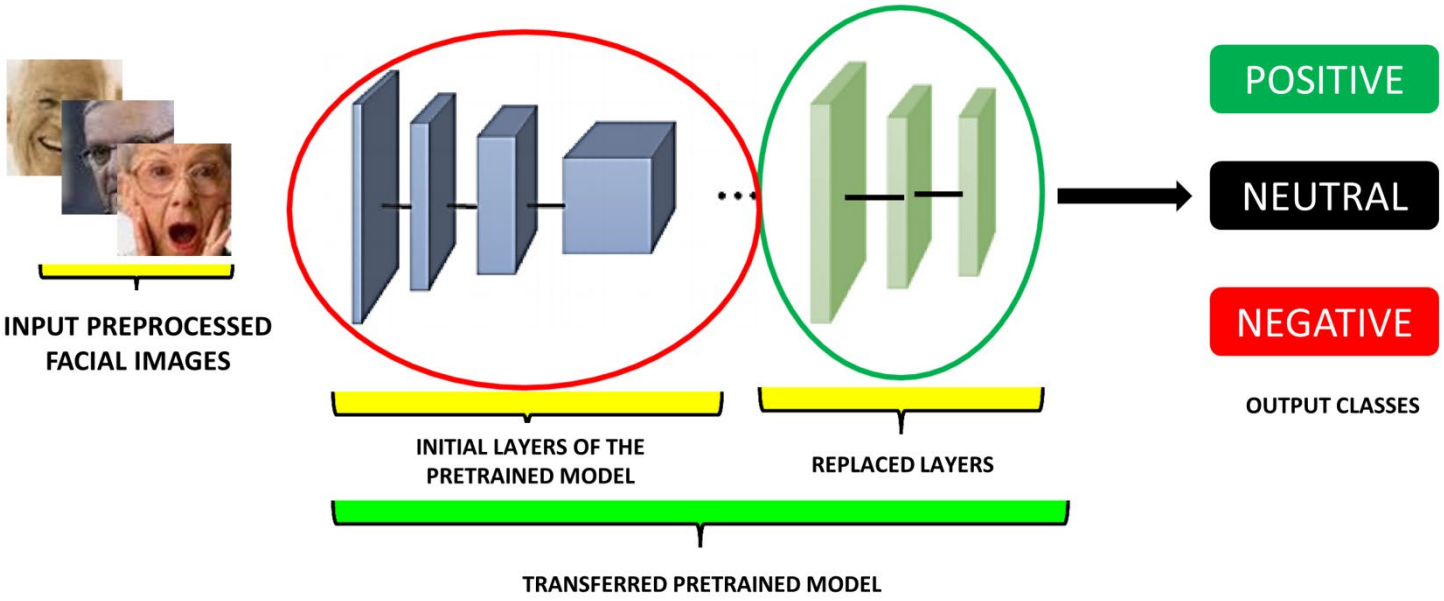
VGG-19



ResNet-50



Classification



Only the last three layers of VGG19 and ResNet50 were replaced to accommodate the new image categories

The traditional facial expression classification has been modified by grouping the expressions into 3 main groups: positive (happiness), negative (fear, disgust, anger and sadness) and neutral

Results (1/3)

The proposed pipeline was tested on two benchmark datasets among the few present in the literature that include facial expressions acquired in uncontrolled conditions and containing subjects of different age groups

CIFE

	Age (years)				
	<35	35-55	56-68	>68	Total
#images	5587	4828	2263	2079	14757

FER-2013

	Age (years)				
	<35	35-55	56-68	>68	Total
#images	13560	7432	6128	5178	32298

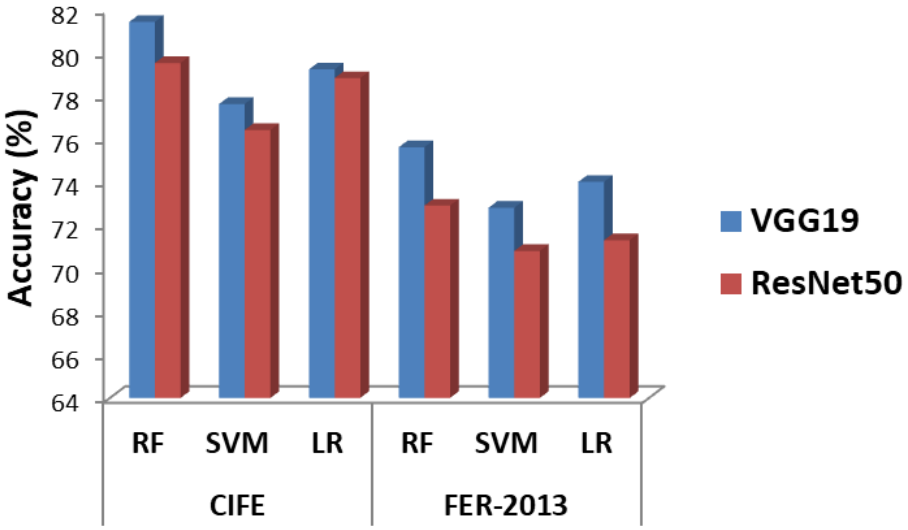
EVALUATION METRIC

$$Acc = \frac{\sum_1^n Acc_{\exp r}}{n} \qquad Acc_{\exp r} = \frac{Hit_{\exp r}}{Total_{\exp r}}$$

n = number of expressions to be considered (in our case *n*=3).

Results (2/3)

For the final classifier layer, Random Forest (**RF**), Support Vector Machine (**SVM**) and Logistic Regression (**LR**) were compared.



The recognition performances of three categories of expressions vary significantly as the dataset changes (FER-2013 is more challenging since the images are grayscale and have a resolution of 48x48 pixels)

Using for pre-training VGG 19 greater accuracy is obtained on both datasets

RF classifier tends to provide an improvement in the results:

- for ***CIFE DATASET*** 3.8% with respect SVM and 2.2% with respect LR
- for ***FER-2013 DATASET*** 2.8% with respect SVM and 1.6% with respect LR

Results (3/3)

FER ACCURACY ON CIFE DATASET

Age Group	VGG19+RF (%)	ResNet50+RF (%)
<35	86.4	85.3
35-55	84.7	81.6
56-68	79.6	77.6
>68	74.9	73.5
Average value	**81.4**	**79.5**

FER ACCURACY ON FER-2013DATASET

Age Group	VGG19+RF (%)	ResNet50+RF (%)
<35	82.5	78.8
35-55	77.1	74.7
56-68	73.2	70.3
>68	69.6	67.8
Average value	**75.6**	**72.9**

Confusion matrices obtained with VGG19 + RF model (only the facial images of older adults with more than 68 years were considered)

	POS	NEU	NEG
POS	88.2	6.3	5.5
NEU	6.1	72.1	21.8
NEG	3.7	31.9	64.4

	POS	NEU	NEG
POS	81.5	11.2	7.3
NEU	5.3	68.4	26.3
NEG	4.5	37.6	58.9

POS : positive expression (happiness)
NEU: neutral expression
NEG: negative expressions (fear, disgust, anger and sadness).

- In the case of implementation of e-coaching platforms: negative expressions are confused considerably with neutral expression
- Negative expressions are symptomatic of the onset or aggravation of diseases !

Conclusions

- The contribution of this study
  - Explore and evaluate two deep transfer learning approach for mood detection in older adults
  - Only three main classes of facial expressions considered: positive, negative and neutral (sufficient for the development of an integrated system capable of implementing e-coaching platforms based on the mood detected.
  - Achieved promising preliminary results, with the pre-trained VGG19 architecture in combination with an RF classifier yielded the best performance for each considered dataset and for each age group in which the dataset has been divided

- Future and ongoing activities:
  - Perform the pre-training of deep architectures on datasets different from ImageNet and more specific for the topic considered (e.g. AffectNet)
  - Extend the number of compared deep learning approaches (Inception-v4 , Inception-ResNet-V2, …)

THANK YOU

for your attention