Zurich University
of Applied Sciences

School of
Engineering

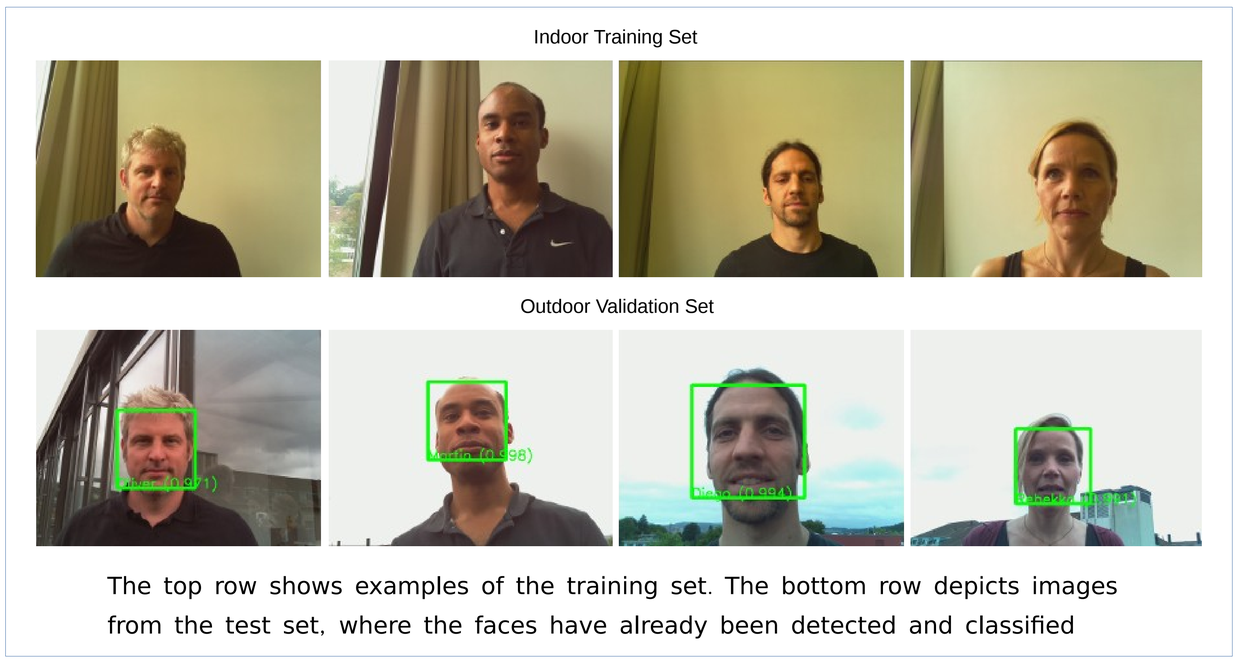
Deep Learning on a Raspberry Pi for Real Time Face Recognition

O. Dürr, D. Browarnik, Y. Pauchard, R. Axthelm, M. Loeser

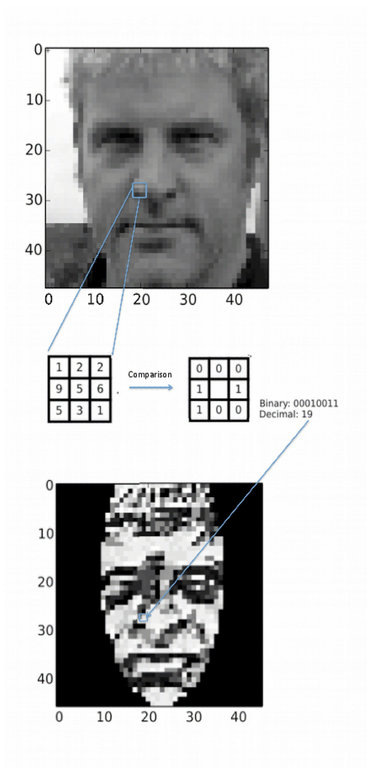
In this paper we describe a fast and accurate pipeline for real–time face recognition. It is based on a convolutional neural network (CNN) and requires only moderate computational resources. After training the CNN on a desktop PC we employ a Raspberry Pi, model B, for the classification procedure. Here, we reach a performance of approximately 2 frames per second and more than 97% recognition accuracy. The proposed approach outperforms all of OpenCV's algorithms with respect to both accuracy and speed and shows the applicability of recent deep learning techniques to hardware with limited computational performance.

Introduction

Automatically recognizing faces and identifying individuals is a problem of large interest and there exists numerous approaches to solve this problem. Recently, convolutional neural networks (CNNs) such as DeepFace have shown impressive classification performance in face recognition tasks. In the following we illustrate how such an approach can efficiently be realized on a Raspberry Pi.



The top row shows examples of the training set. The bottom row depicts images from the test set, where the faces have already been detected and classified



Methods

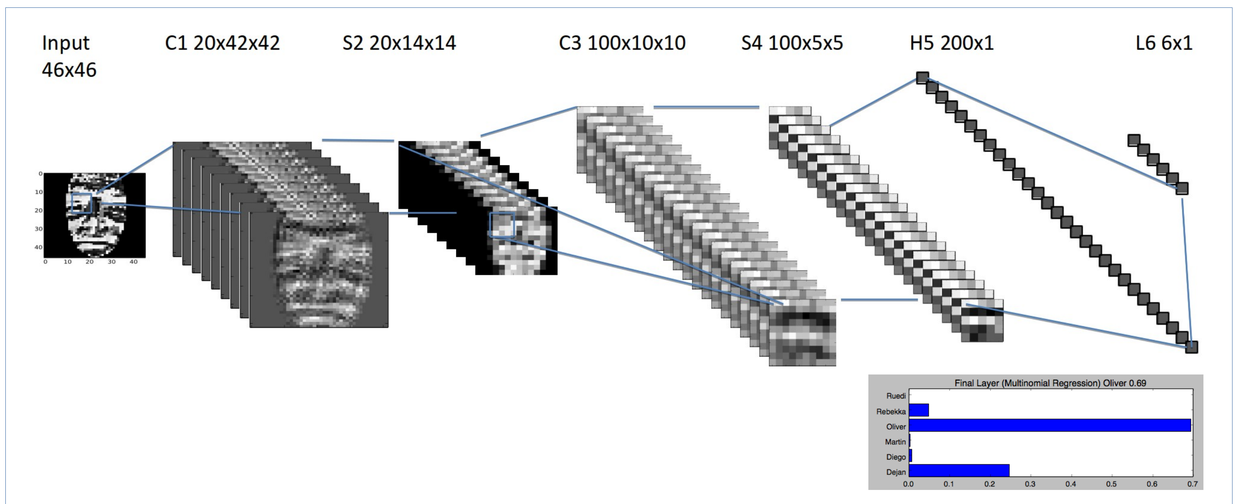
- Training and validation images acquired with a Raspberry Pi camera module (NoIR)
- Six individuals, approx. 40 images per person
- Face detection with Viola–Jones (OpenCV)

Baseline classification:

- Training data augmentation: none
- Alignment: eye detection (Viola–Jones) for rigid and scaling transformation
- Pre–processing: DoG filtering and ellipse cropping
- OpenCV–based classification: Eigen–, Fisherfaces and Local Binary Histogram Pattern (LBP)

Proposed classification:

- Training data augmentation: random rigid and scaling transformations to account for misalignment
- Alignment: none
- Pre–processing: LBP and ellipse cropping
- Classification: Convolutional Neural Network (see above) training data augmentation.



Schematic representation of the CNN featuring convolutional (C), maxpooling (S) and fully connected (H5, L6) layers

Results

Baseline classification:

- Best classification accuracy 96.9% (Fisherfaces with alignment)
- Number of enrolled (successfully aligned) images decreased from N=278 to N_e=192
- Classification time is 958 ms per image
- Without alignment, accuracy drops (89%) and time is reduced (488 ms)

Proposed classification:

- Best classification accuracy 97.5%.
- Number of enrolled images: all 278 test images
- Classification time is 528 ms per image
- Without augmentation, accuracy drops (24%) and time remains constant

Conclusion

Our results show that image alignment necessary for baseline classification is computationally expensive and can be replaced in the proposed method by augmenting the training data. In conclusion, our approach leads to significant improvements with respect to both speed and accuracy compared to OpenCV. Thus, recent deep–learning techniques are also applicable on hardware with limited resources.

Method	Accuracy	Classification Time [msec]	Enrollment Rate N_e/N	Total Time Per Face [msec]
CNN (p_0=0.85)	99.59%	105 +/- 8	250/278	529 +/- 64
CNN (p_0=0.00)	97.48%	105 +/- 8	278/278	529 +/- 64
Fisherfaces (no al.)	88.5%	54 +/- 11	278/278	511 +/- 89
Fisherfaces (al.)	96.87%	535 +/- 89	192/278	1006 +/- 18

Accuracy and performance for various approaches.

Contact

Dr. Martin Loeser, School Of Engineering, 8400 Winterthur
martin.loeser@zhaw.ch, https://home.zhaw.ch/~loma