University of Basel

An Empirical Comparison of Deep Learning and 3DMM-Based Approaches for Facial Occlusion Segmentation

Elias Arnold <elias.arnold@stud.unibas.ch>

Department of Mathematics and Computer Science, University of Basel

10.08.2018

Overview

- Motivation
- Occlusion Segmenation with an FCN
- Experiment: Segmentation Quality
  - Qualitative segmenation results
  - Quantitative segmentation results
- Experiment: Fitting under occlusion
- Application to real-life data and conclusion
- Future Work

The Goal


(a) original


(b) mask


(c) resulting fit

Figure: Fitting with a perfect mask

The Problem (Segmentation under Occlusion)

(a) original

(b) resulting fit

Figure: Fitting without a mask

Overview

- Motivation
- Occlusion Segmenation with an FCN
- Experiment: Segmentation Quality
  - Qualitative segmenation results
  - Quantitative segmentation results
- Experiment: Fitting under occlusion
- Application to real-life data and conclusion
- Future Work

Convolutional Neural Network (CNN)

- Artificial Neural Network
- fully connected layers at the end
- makes decisions based on global information

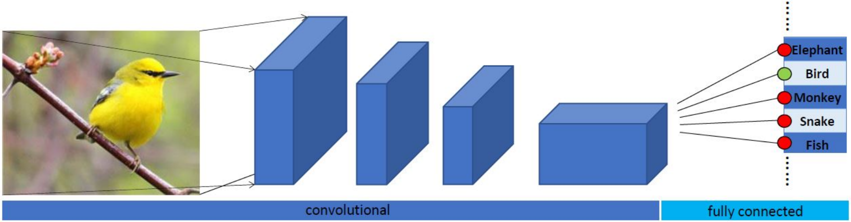


Figure: A classical neural network

Fully Convolutional Network (FCN)

- without fully connected layers at the end
- makes decisions based on spatial information

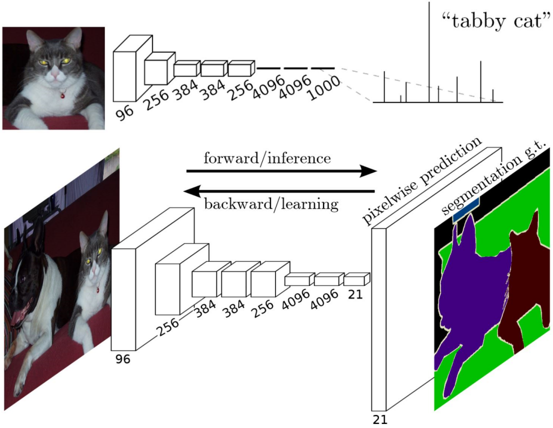


Figure: The transformation from a CNN into a FCN

Training of the FCN

- The FCN was trained by Nirkin et al.
- They used the well known FCN-8s-VGG network
- Training Data: Frames of Videos
- Frame used multiple times (synthetic occlusions)
- Total: 9’818 facial images

Figure: Two images used for training. They are overlaid them with an synthetic occlusion.

Overview

- Motivation
- Occlusion Segmenation with an FCN
- Experiment: Segmentation Quality
  - Qualitative segmenation results
  - Quantitative segmentation results
- Experiment: Fitting under occlusion
- Application to real-life data and conclusion
- Future Work

Qualitative Evaluation on the COFW Dataset

- **C**altech **O**ccluded **F**aces in the **W**ild
- used of Nirkin et al themselves

(a) Six segmentations of the FCN

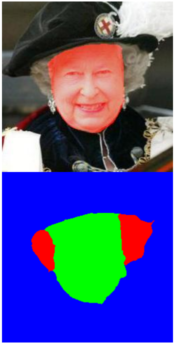
(b) Six segmentations of the approach of Egger et al

Overview

- Motivation
- Occlusion Segmenation with an FCN
- Experiment: Segmentation Quality
  - Qualitative segmenation results
  - Quantitative segmentation results
- Experiment: Fitting under occlusion

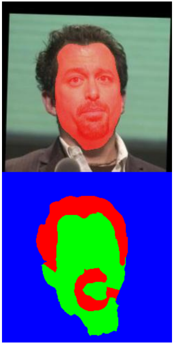
Evaluation on the Parts-LFW Dataset (1)

- **L**abeled **F**aces in the **W**ild
- ground truth labels were provided
- Difficult question: What is an occlusion?



(a) The desired ground truth labels

(b) Face and skin have same label → False negatives



(c) Facial hair is an occlusion → false positives (hair)

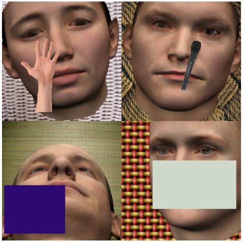
Evaluation on the Parts-LFW Dataset (2)

- statistics with simple pixelwise comparison
- same label for skin parts and face (false negatives)

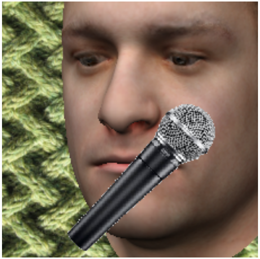
false-positives (hair)	7.04%
false-positives (background)	0.68%
false-negatives (background)	14.13%
right-segmentations	85.87%
right non-sementations	99.32%

Evaluation on Synthetic Data (1)

- Known ground truth labels (c)
- Control of every rotation
- Control of the occluded amount



(a) A synthetic face



(b) A synthetic face with an occlusion

(c) The ground truth mask of (b)

Evaluation on Synthetic Data (2)

- Segmentation is very sensitive to the roll rotation
- Pitch is asymmetric (better recognizable with high angles)
- Yaw rotation has little effect on the segmentation

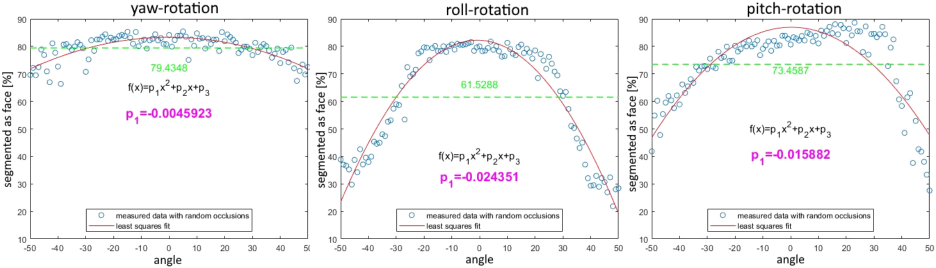


Figure: Plot for every rotation (yaw, roll, pitch). Quadratic fit (red curve/$f(x)$). Average (green). The data for the plot consisted of synthetic facial images with microphones and hands as occlusions.

Overview

- Motivation
- Occlusion Segmenation with an FCN
- Experiment: Segmentation Quality
  - Qualitative segmenation results
  - Quantitative segmentation results
- Experiment: Fitting under occlusion
- Comparison of different approaches
- Future Work

Fitting with the FCN segmentation (1)

	ground truth mask	Egger	FCN	no mask
z-labels				
fits				

Fitting with the FCN segmentation (2)

- Calculation of the error (in σ) in each iteration
- In each parameter-space, we used 50 dimensions
- Average over 10 images (only with hands as occlusions)
- In each iteration, 1000 samples were drawn

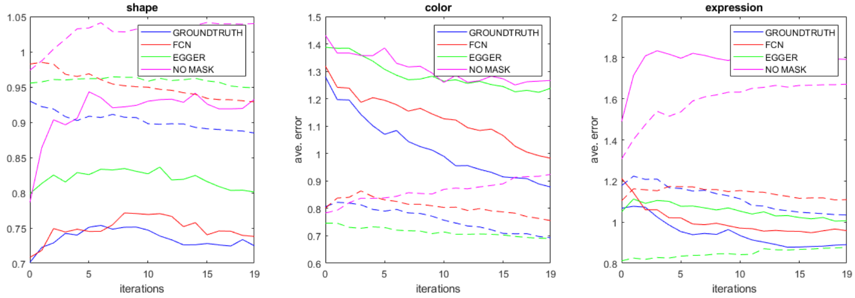


Figure: The average error of the first five (most important) 3DMM parameters is shown with a solid line. The average other 45 parameters is shown with a dashed line.

Oversegmentaion of EGGER

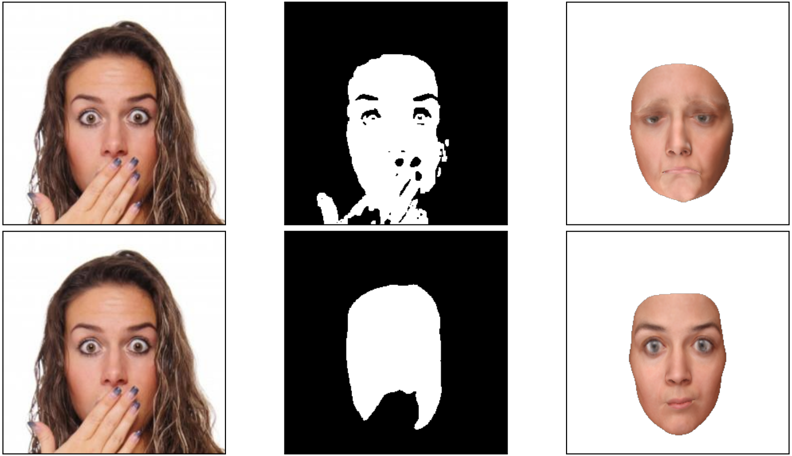
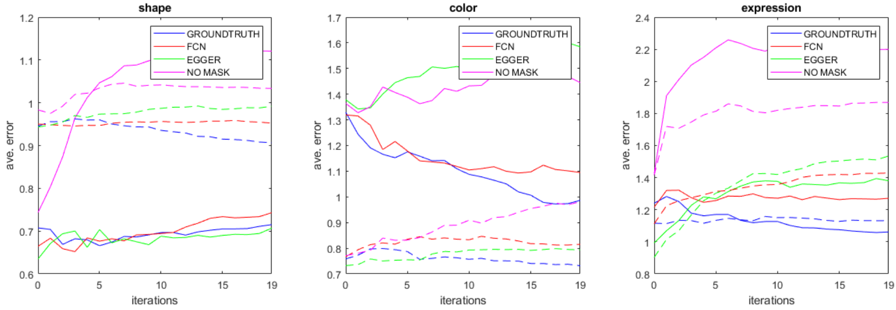


Figure: Fitting with occlusion-aware masks. Mask of Egger et al (top row). Mask of the fully convolution network used in this thesis (bottom row).

Fitting with the FCN segmentation (3)

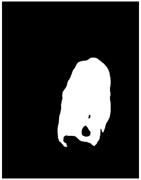
- Biggest weakness of Egger et al's approach: Oversegmentation
- How can we spread between (FCN and EGGER) bigger?
- Synthetic faces with both, the 'face12' and the 'bfm' version



Overview

- Motivation
- Occlusion Segmenation with an FCN
- Experiment: Segmentation Quality
  - Qualitative segmenation results
  - Quantitative segmentation results
- Experiment: Fitting under occlusion
- Application to real-life data and conclusion
- Future Work

Real-Life data

	masks		**fits**	
original	EGGER	FCN	EGGER	FCN
				
				

Conclusion

	advantages	disadvantages
FCN	• Does not oversegment • Segments only facepixels • Detects known occlusions	• Poor detection of thin and new occlusions • Fitting not much faster • Very sensitive to the rotation of the face
EGGER	• Looks at each pixel individually, therefore can detect any occlusion • Stable on real-life data	• oversegments • excludes important details on the mask • computes the mask iteratively

Conclusion

	advantages	disadvantages
FCN	Does not oversegment Segments only facepixels Detects known occlusions	Poor detection of thin and new occlusions Fitting not much faster Very sensitive to the rotation of the face
EGGER	Looks at each pixel individually, therefore can detect any occlusion Stable on real-life data	oversegments excludes important details on the mask computes the mask iteratively

Conclusion

	advantages	disadvantages
FCN	- Does not oversegment - Segments only facepixels - Detects known occlusions	- Poor detection of thin and new occlusions - Fitting not much faster - Very sensitive to the rotation of the face
EGGER	- Looks at each pixel individually, therefore can detect any occlusion - Stable on real-life data	- oversegments - excludes important details on the mask - computes the mask iteratively

Conclusion

	advantages	disadvantages
FCN	- Does not oversegment - Segments only facepixels - Detects known occlusions	- Poor detection of thin and new occlusions - Fitting not much faster - Very sensitive to the rotation of the face
EGGER	- Looks at each pixel individually, therefore can detect any occlusion - Stable on real-life data	- oversegments - excludes important details on the mask - computes the mask iteratively

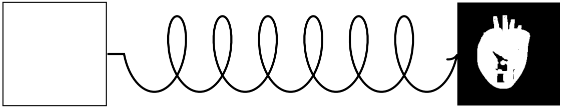
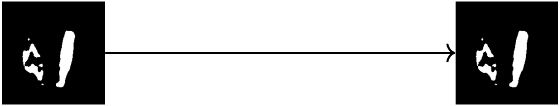
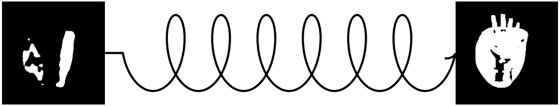
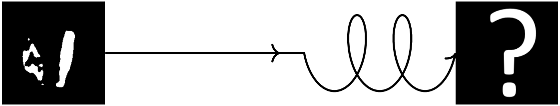
Conclusion

	advantages	disadvantages
FCN	Does not oversegment Segments only facepixels Detects known occlusions	Poor detection of thin and new occlusions Fitting not much faster Very sensitive to the rotation of the face
EGGER	Looks at each pixel individually, therefore can detect any occlusion Stable on real-life data	oversegments excludes important details on the mask computes the mask iteratively

Overview

- Motivation
- Occlusion Segmenation with an FCN
- Experiment: Segmentation Quality
  - Qualitative segmenation results
  - Quantitative segmentation results
- Experiment: Fitting under occlusion
- Application to real-life data and conclusion
- Future Work

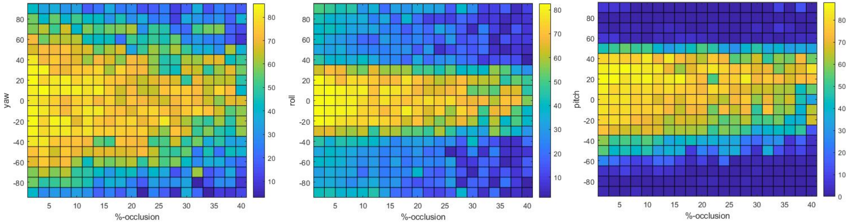
Future Work: Combination of both masks

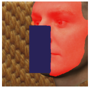
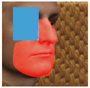
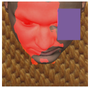
	initial mask → fitting iterations → final mask
EGGER	
FCN	
Our Combination	
Future Work	

Questions?

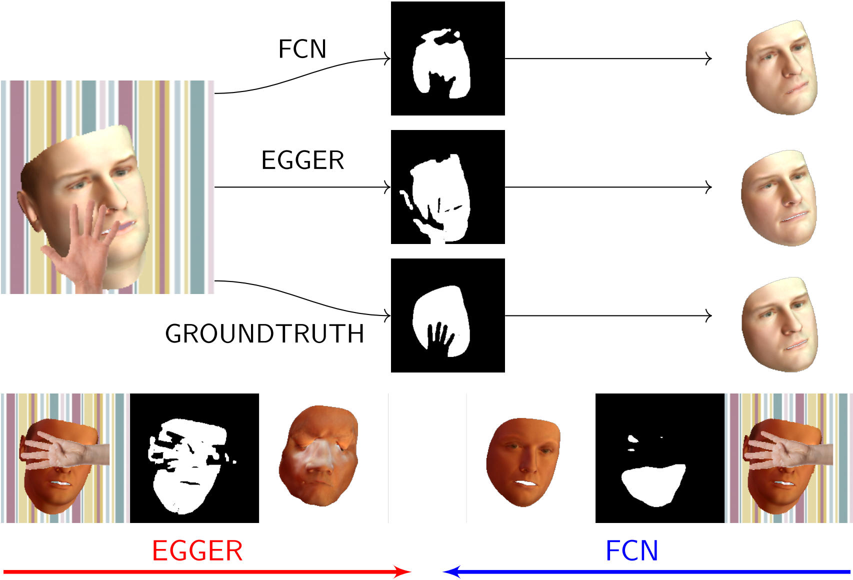
elias.arnold@stud.unibas.ch

Dependence of the Degree of Occlusion



	yaw	roll	pitch
-40°			
40°			

Wrong Color Choice with the Mask of Egger et al



Related Work (Nirkin et al: On Face Segmentation, Face Swapping, and Face Perception)

- Widespread artificial neural networks can be used for face segmentation
- intra-subject swapped faces remain as recognizable as their sources (a)
- less recognizable inter-subject results (b)

(a) Example of Intra-Subject swapping

(b) Example of Inter-Subject swapping