De-occluding Faces Using Novel Deep Learning Techniques for Increased Performance of Facial Expression Recognition Networks

Sachin Shah, University of Central Florida, Department of Computer Science, Synthetic Reality Laboratory *(sachin.shah@knights.ucf.edu)*

Abstract

Facial expression recognition (FER) models have many applications for human-computer interaction; however, many current FER models struggle with faces that are partially occluded (hidden). Research suggests trained preprocessors can remove these occlusions, but it is unclear how key facial features can be preserved during the de-occlusion process.

We hypothesize that facial expressions can be preserved by employing our novel expression-consistency loss function which penalizes our network each time the expression of a de-occluded image is mispredicted. This method allows narrow FER networks to focus on ideal conditions, reducing their complexity and training time. Results show this learning technique does preserve expression information and increases FER accuracy over FER networks alone. Further work will explore if preprocessors for other tasks, such as identity mapping, could be similarly created.

Problem Statement

FER models often struggle when portions of facial images are partially occluded because features that contain key expression information are hidden. Literature suggests that preprocessors can be trained to remove such occlusions; however, because current preprocessors optimize for the "realism" of de-occluded images, the underlying expression is not necessarily preserved, and in fact is sometimes altered. This in turn doesn't lead to higher recognition rates of expressions in FER networks.

Figures: Occluded faces. (Source: COFW Dataset)

Methods

The proposed network aims to synthesize occlusion-free images by considering the accuracy of the end FER result, instead of just the realism of the intermediary image, when training the model. This de-occluder model follows an encoder-decoder structure and is trained with supervision from an expression-consistency loss, pixel loss, and a real/synthetic discriminator.

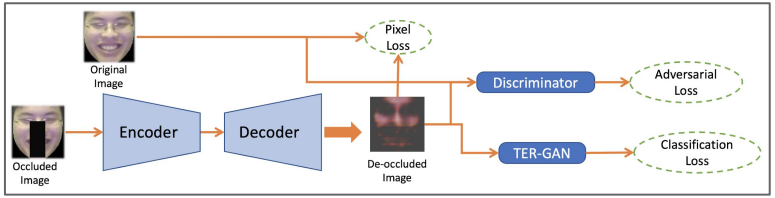

Figure: Learning model architecture diagram. (Source: Researcher)

We compute expression-consistency loss as the cross-entropy loss between the TER-GAN classification on the generated image & the ground truth value. TER-GAN is used as the pre-trained expression classifier with frozen weights.

$$L_{total} = L_{expr} + L_{pixel} + L_{adv}$$

The total loss is the sum of expression-consistency loss, L2 pixel loss between the original unoccluded image and the generated image, and the real/fake adversarial loss from the discriminator.

The model is trained on the Oulu-Casis dataset with a variety of synthesized occlusions. Original images are occluded four times in varying amounts. Experiments compare the TER-GAN FER network accuracy on occluded images with and without preprocessors.

The original and occluded image pair are passed through the network during training to compute each loss value. During inference for the entire FER network, only the encoder-decoder and TER-GAN are needed.

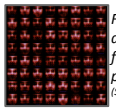

Figure: Sample de-occluded faces after preprocessing. (Source: Researcher)

Experimental Results

Table: (A) TER-GAN without a preprocessor. (B) Preprocessor trained with only adversarial and pixel loss. (C) Preprocessor trained also with expression loss term.

Method	FER Accuracy
Baseline (A)	63%
Trained without L_{expr} (B)	20%
Full Model (C)	71%

Discussion & Conclusion

The un-aided TER-GAN model struggles to classify occluded faces, achieving only 63% accuracy on the test set. The naive preprocessor which uses only pixel loss and adversarial loss in fact *decreases* the accuracy of the model to a abysmal 20%. This suggests that optimizing for "realism" indeed alters the underlying expressions. The model trained on our full loss function achieves 71% accuracy, demonstrating that adding expression-consistency is a promising path for improving preprocessing techniques.

Future Work

- Hyper-parameter tuning, varying weights on loss functions
- Handle more challenging occlusions
- Preprocessors for other tasks (e.g. identity mapping)

Acknowledgements

This research was conducted under the supervision of Dr. Charles Hughes, Co-Director of the Synthetic Reality Laboratory, UCF and Kamran Ali, CS PhD Candidate, UCF.

References:

(1) Kamran Ali & Charles E. Hughes. (2019). All-In-One: Facial Expression Transfer, Editing and Recognition Using A Single Network.
(2) Xiaowei Yuan & In Kyu Park (2019). Face De-Occlusion Using 3D Morphable Model and Generative Adversarial Network. 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea (South), pp. 10061–10070). IEEE.