

Facial Expression Recognition using Residual Masking Network

Facial Expression
Recognition using
Residual Masking
Network

Luan Pham,
Huynh Vu, Tuan
Anh Tran

Luan Pham, Huynh Vu, Tuan Anh Tran
Research & Development - Cinnamon AI
Faculty of Computer Science and Engineering - HCMUT

Introduction

Problem definition

- **Input:** an facial image (without alignment information).
- **Output:** probabilities of seven classes.¹

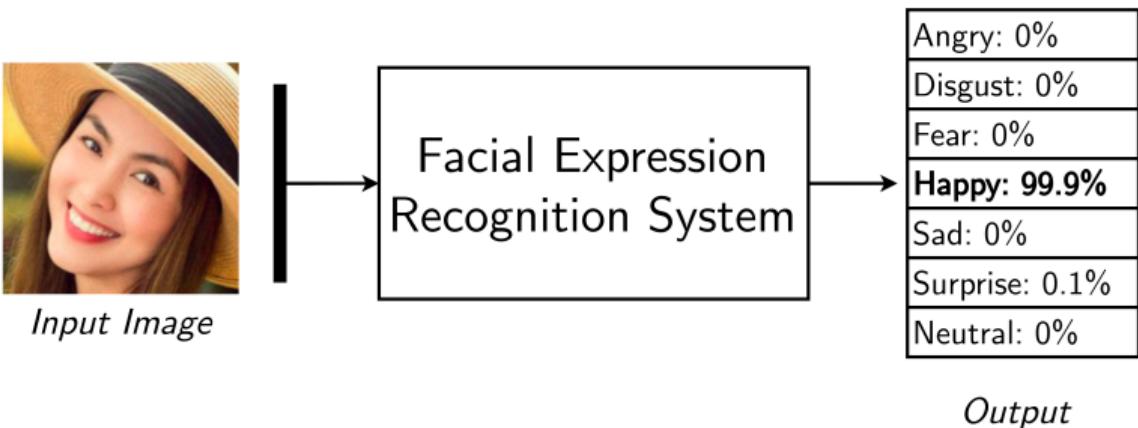


Figure -1.1: Diagram of a Simple Facial Expression Recognition API.

¹Include 6-concrete emotions and neutral state.

Facial Expressions of Emotion



Figure -1.2: Facial action ensembles for common-view facial expressions.¹

¹ Challenges to inferring emotion from human facial movements, 2019 — Barrett, Lisa Feldman, et al.

Facial landmarks

Problems

- Failure in working with small images.
- Poor confidence in complex input.
- Huge cost for labelling.

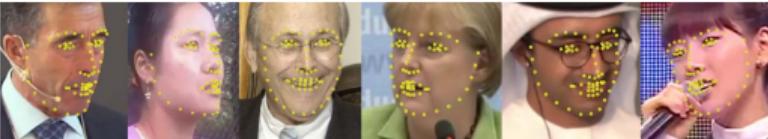


Figure -1.3: Good confidence landmarks

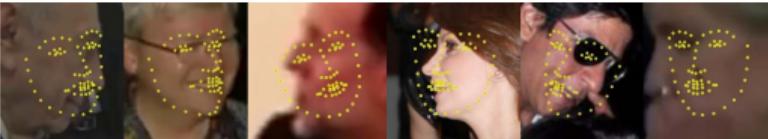


Figure -1.4: Poor confidence landmarks

Q: A end-to-end method to extract facial expression without this intermediate level features?

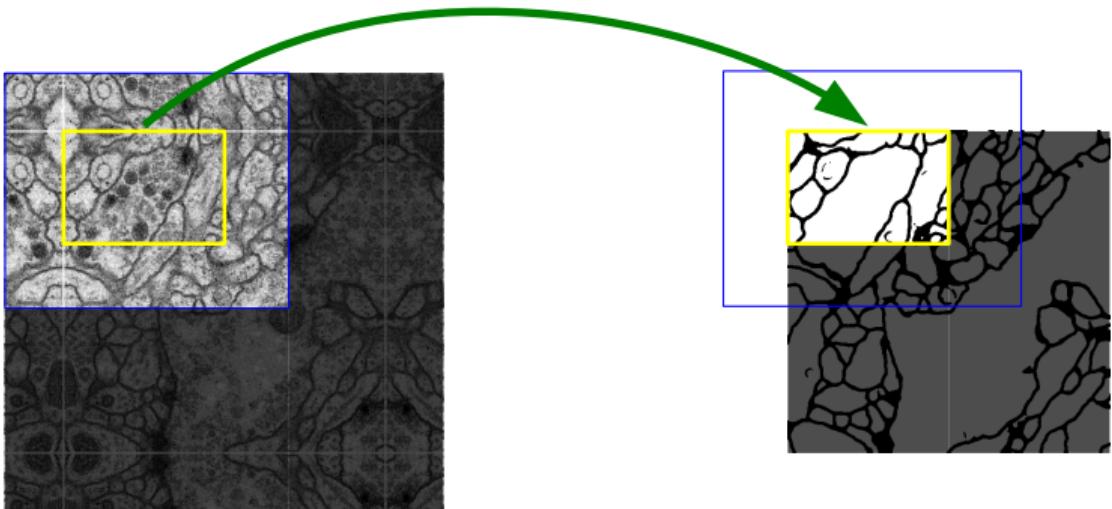


Figure -1.5: Segmentation problem input and output.¹

¹"U-net: Convolutional networks for biomedical image segmentation." — Ronneberger et., MICCAI 2015

Masking Idea

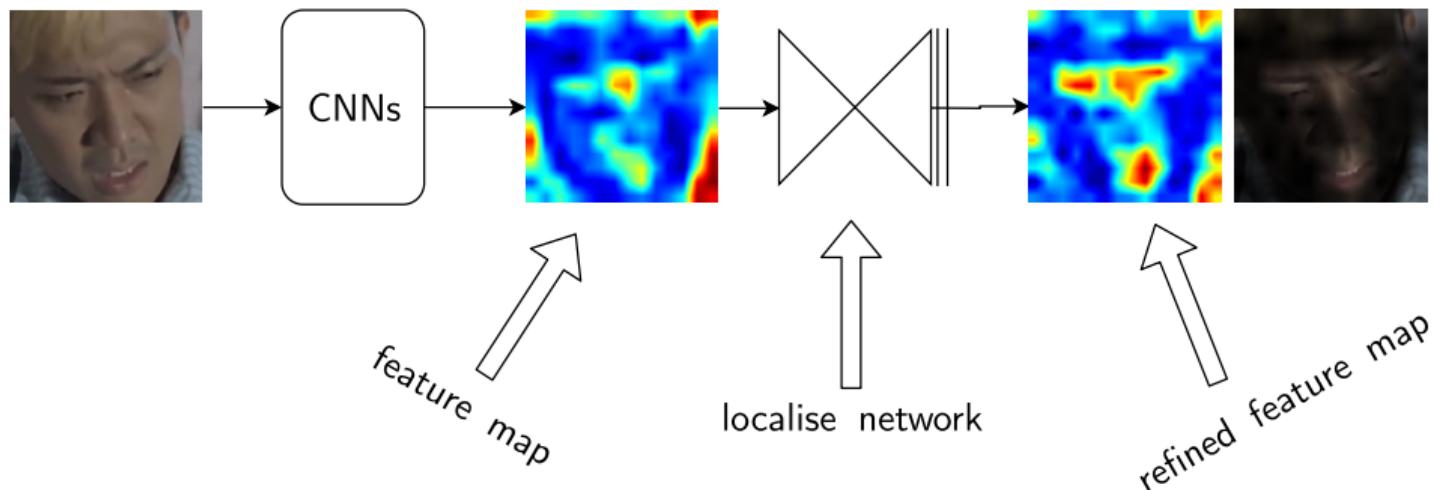


Figure -1.6: Attention to the interested area.

Masking Idea

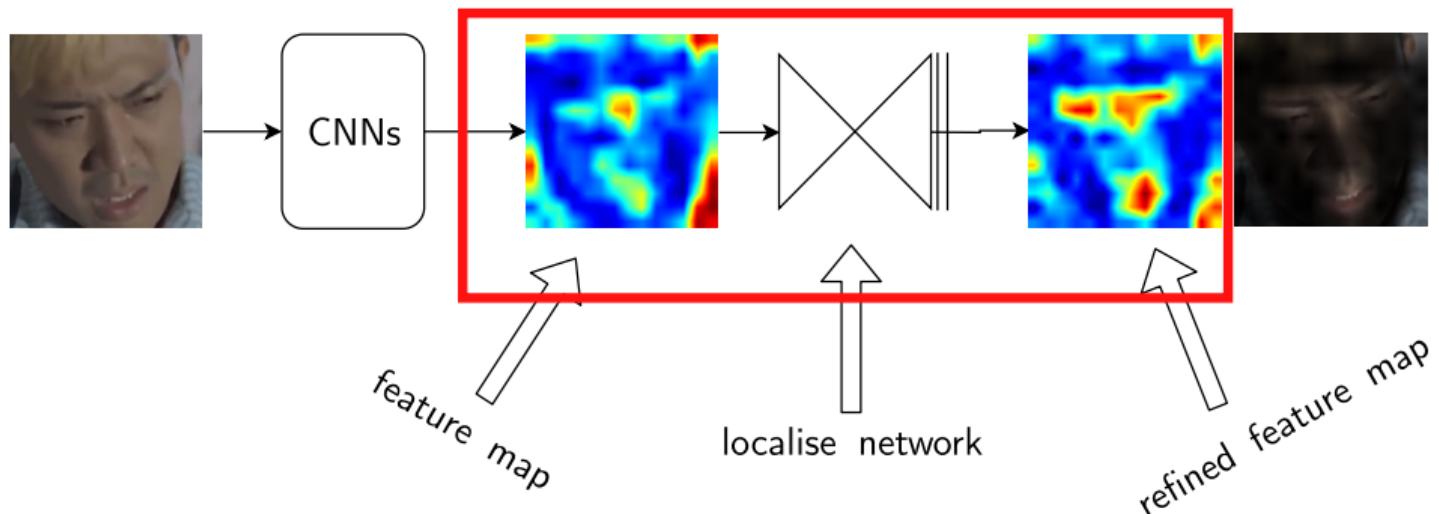


Figure -1.7: Attention to the interested area.

Masking Block

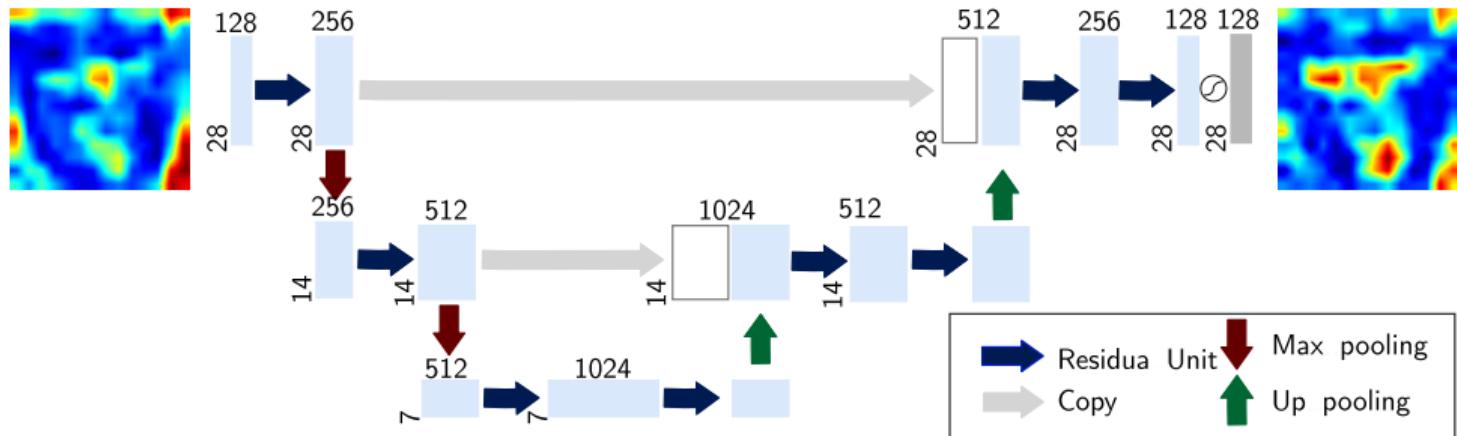


Figure -1.8: Masking Block with $depth = 3$.

The depth of Masking Block is dynamically changed by the spatial size of the input feature map.

Residual Masking Network

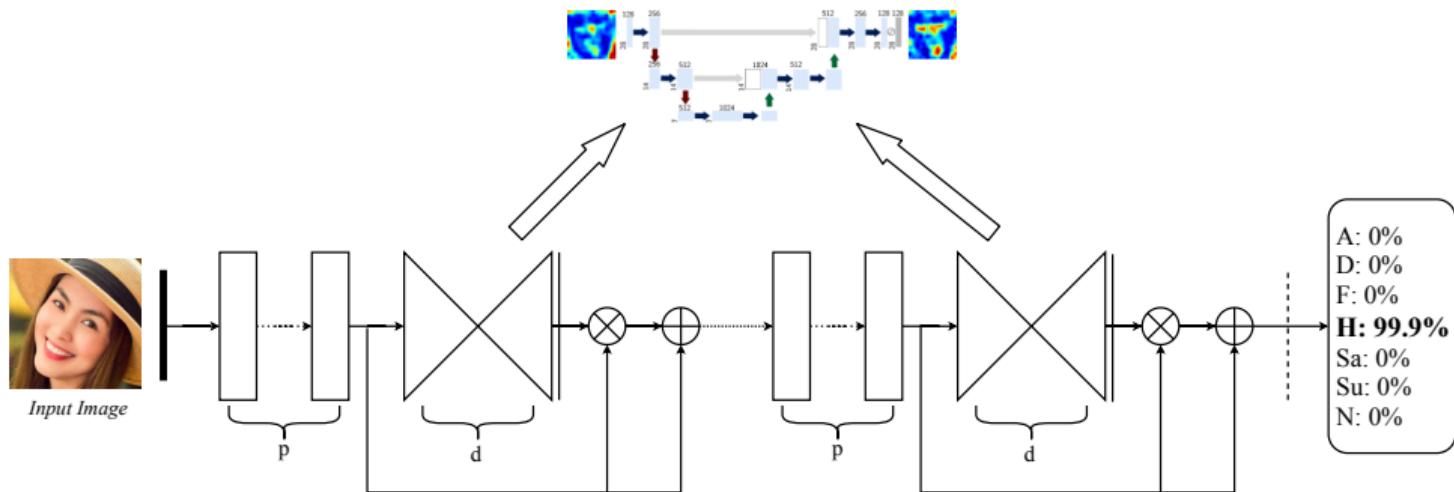


Figure -1.9: Residual Masking Network Overview.

Datasets

FER2013

- 35,887 gray-scale images
- W×H: 48×48
- In-the-wild
- Widely used



VEMO

- 36,470 RGB images
- W×H: Varied
- In-the-wild
- Vietnamese focused



Figure -1.10: Example images

Evaluation results on FER2013

Table -1.1: Results of reproduced training models on FER2013

Name	Acc (%)
VGG19	70.80
ResAttNet56	72.63
Densenet121	73.16
Resnet152	73.22
Cbam_resnet50	73.39
ResMaskingNet	74.14

Table -1.2: Compare with reported results.¹

Name	Acc (%)
Human Accuracy [1]	65 ±5
DNN_L_Reg [2]	66.40
DL-SVM [3]	71.16
CNN-SIFT	73.40
CNNs, BOVW, local SVM	75.42
Ensemble 8 CNNs [6]	75.20
ResMaskingNet + 6	76.82²

¹There is no space left, citations can be found in the dissertation.

²Ensemble the Residual Masking Network with 6 other CNNs.

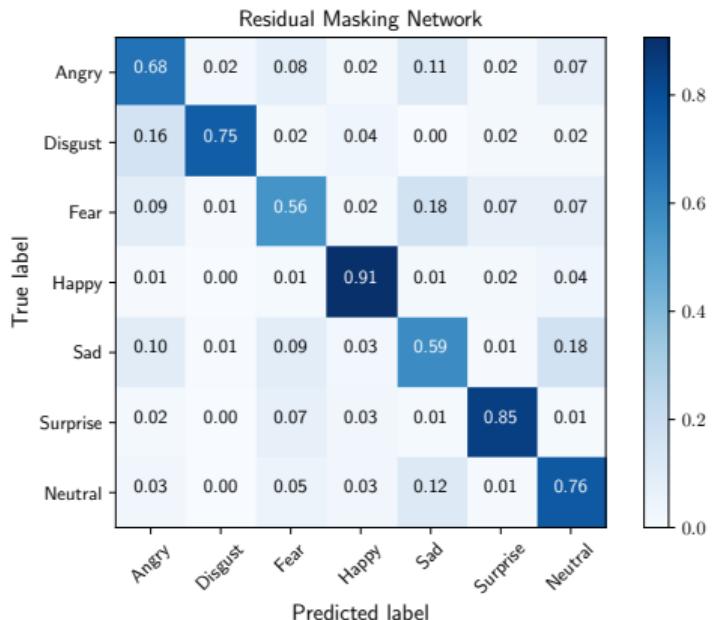


Figure -1.11: CM of RMN on FER2013



Figure -1.12: Failure cases

Evaluation results on VEMO

Table -1.3: Results on VEMO

Name	Acc (%)
ResAttNet56	60.82
Resnet18	63.94
Resnet34	64.84
ResMaskingNet	65.949

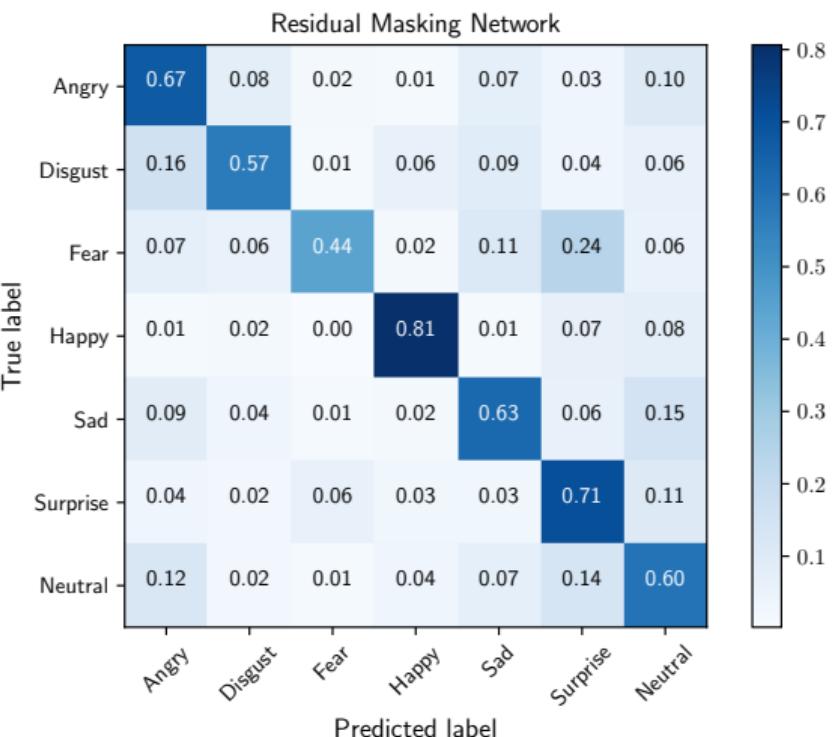


Figure -1.13: CM of RMN on VEMO.

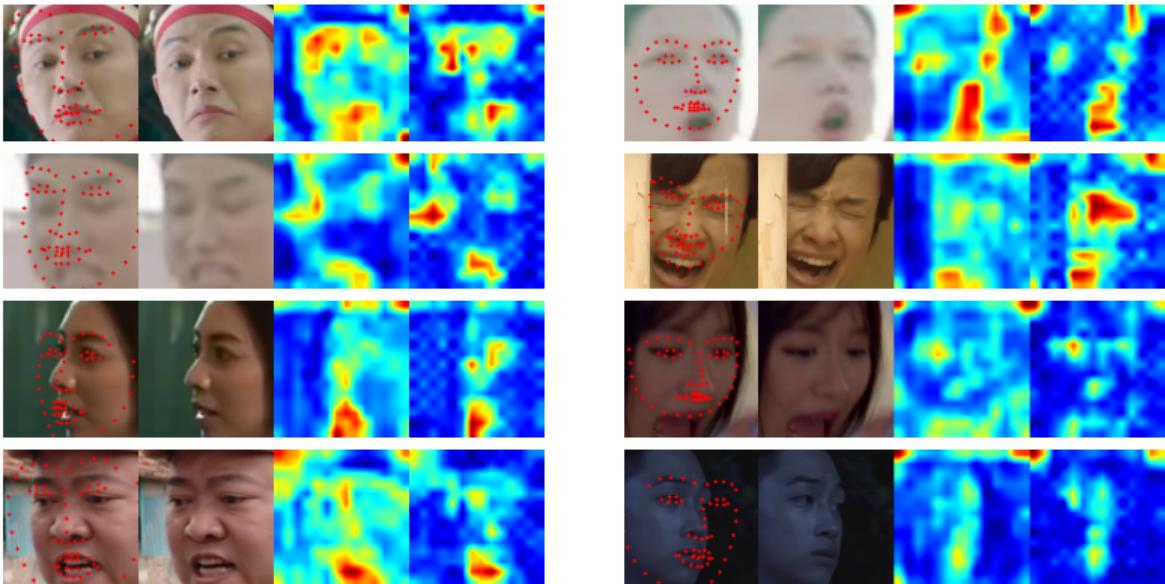


Figure -1.14: Landmark → Image → Before 3rd masking layer → After 3rd masking layer

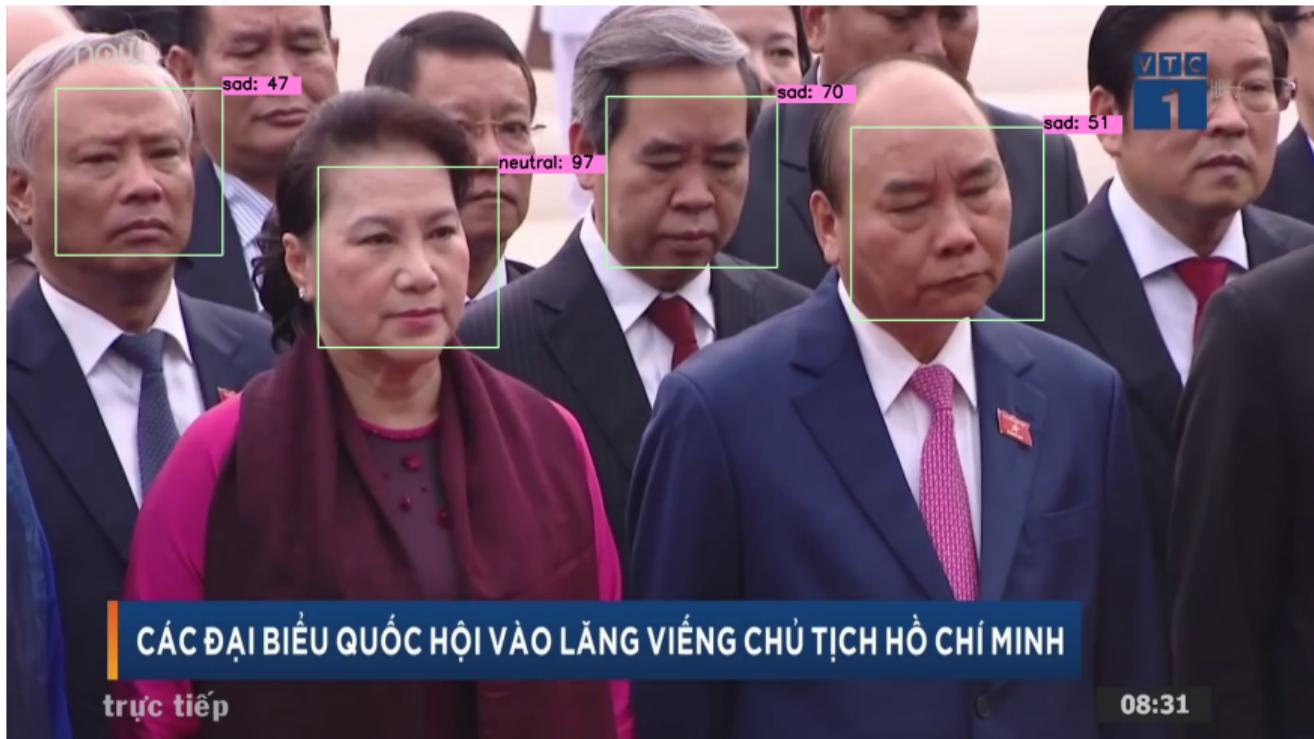


Figure -1.15: National Assembly deputies visit Ho Chi Minh Mausoleum

Demo



Facial Expression
Recognition using
Residual Masking
Network

Luan Pham,
Huynh Vu, Tuan
Anh Tran



Figure -1.16: Dung and Ha Lan are on Honda with Happy face - Image from Mat Biec movie

Demo

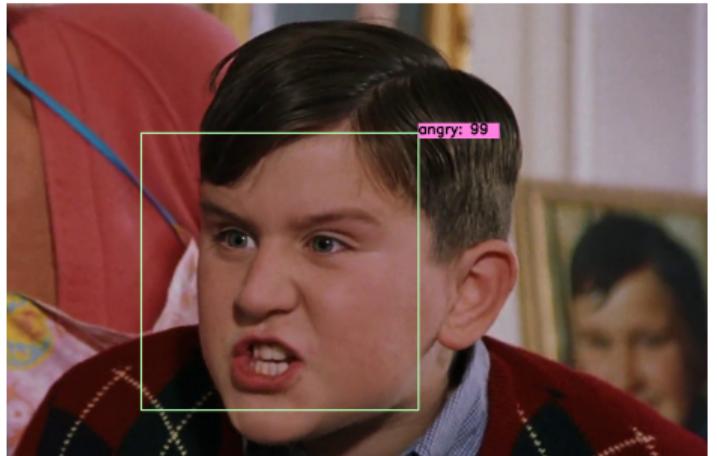


Figure -1.17: Dudley being angry -
Harry Potter movie



Figure -1.18: A Vietnamese actress
being sad

Summary

- Masking Idea can be used to help CNNs focus to regions of interest without facial landmarks.
- The proposed method - Residual Masking Network with Masking Idea, can produce competitive results in FER problem, ensemble with other CNNs can produce state-of-the-art result.
- Create VEMO dataset, the first novel one that supports recognizing facial expression of Vietnamese.



Facial Expression
Recognition using
Residual Masking
Network

Luan Pham,
Huynh Vu, Tuan
Anh Tran

Thanks for listening



Facial Expression
Recognition using
Residual Masking
Network

Luan Pham,
Huynh Vu, Tuan
Anh Tran

Questions?

References for Table 4.2

1. Goodfellow, Ian J., et al. "Challenges in representation learning: A report on three machine learning contests." ICML, 2013.
2. Minaee, Shervin, and Amirali Abdolrashidi. "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network." arXiv:1902.01019 (2019).
3. Tang, Yichuan. "Deep learning using linear support vector machines." arXiv:1306.0239 (2013).
4. Kim, Bo-Kyeong, et al. "Fusing aligned and non-aligned face information for automatic affect recognition in the wild: a deep learning approach." CVPR Workshops. 2016.
5. Nguyen, Hai-Duong, et al. "Facial expression recognition using a multi-level convolutional neural network." ICPRAI, 2018.
6. Pramerdorfer, Christopher, and Martin Kampel. "Facial expression recognition using convolutional neural networks: state of the art." arXiv:1612.02903 (2016).

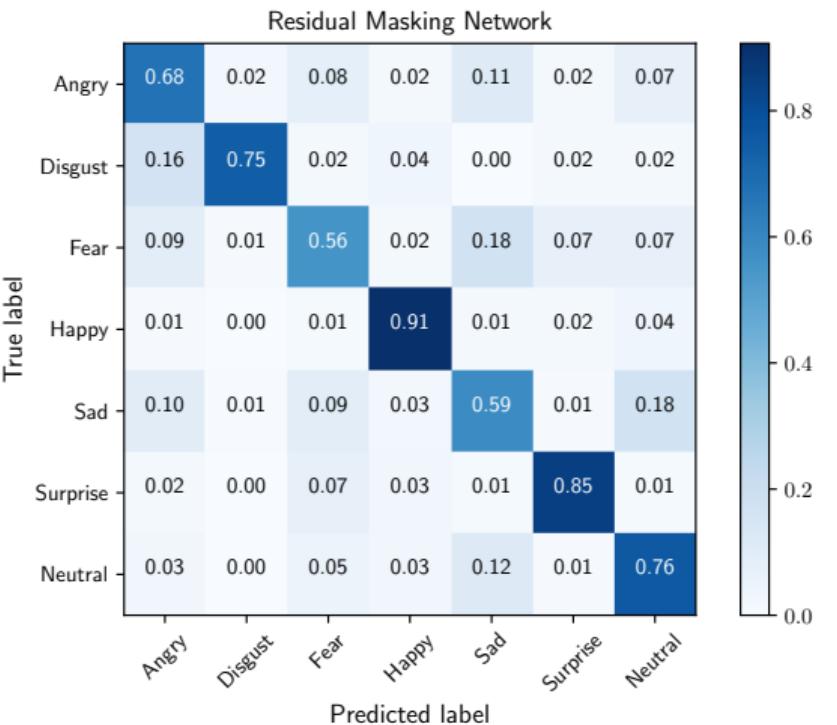
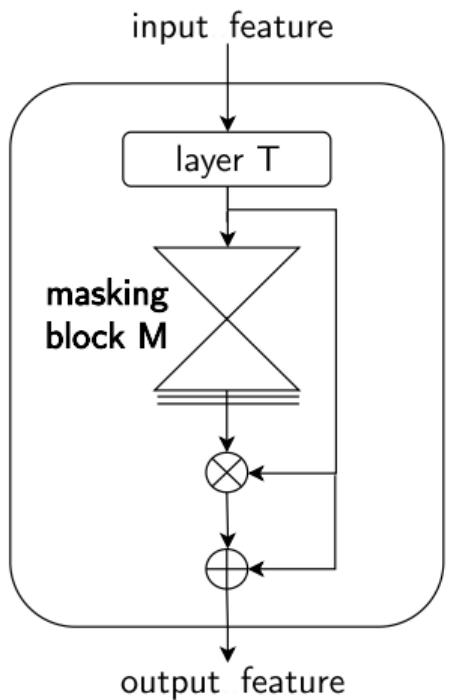


Figure -1.19: Confusion Matrix of Residual Masking Network on FER2013 dataset



Given input feature x ,
with i is spatial index, c is channel
index.

- $t = T(x)$
- $m = M(t)$
- $b_{i,c} = (1 + m_{i,c}) * t_{i,c}$

Figure -1.20: Residual Masking Block.