Human-Centred Technology Research Centre, Faculty of Science and Technology,
University of Canberra, Canberra, Australia

CNN Depression Severity Level Estimation from Upper Body vs. Face-Only Images

MPRSS Workshop

Dua'a Ahmad, Roland Goecke and James Ireland
Duaa.Ahmad@canberra.edu.au

20/12/2020

Overview

Convolutional Neural Network (CNN)

Alexnet

Resnet34

VGG19

DenseNet169

Depression Severity

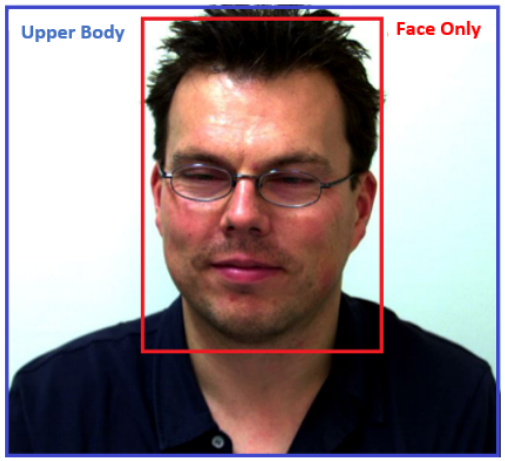
Severely Depressed

27

0

Not Depressed

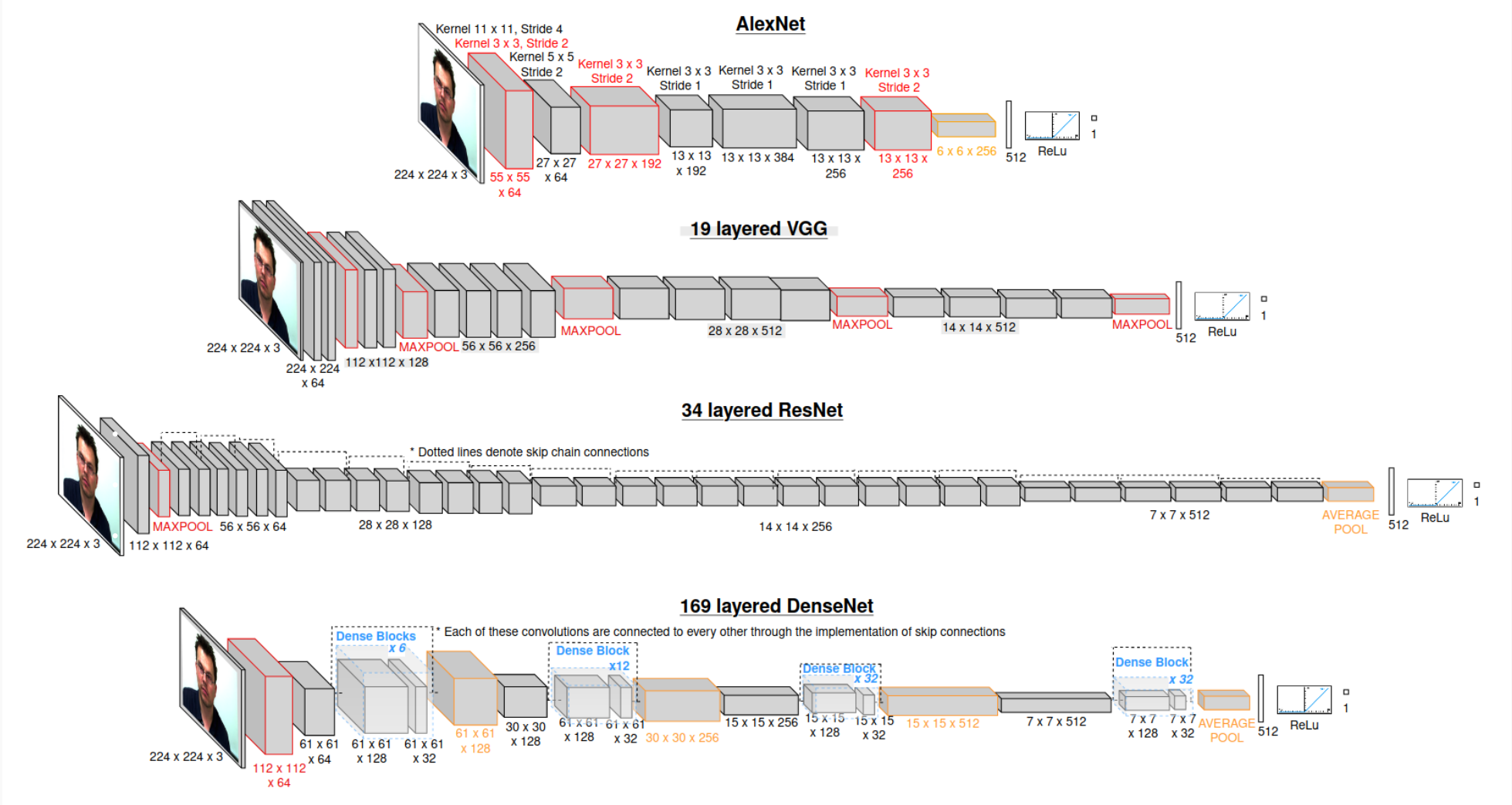
Face vs Upper Body Frames



Method-Experimental Design

Models

CNN architectures used, greyed regions have not been fine-tuned in the pre-trained models



Datasets

The Black Dog Institute Dataset (BlackDog)

- Clinically validated and self assessed
- 60 Participants
- Based on the Quick Inventory of Depressive Symptomatology (QIDS)-self-assessmet

Audio Visual Emotion Challenge (AVEC) 2013

- Not clinically validated, but self assessed
- 150 Participants (100 were used)
- Based on Beck Depression Inventory (BDI) assessment

AVEC 2013 BDI scales were converted to QIDS for compatability with scores of BlackDog dataset.

Method-Experimental Design

Setup

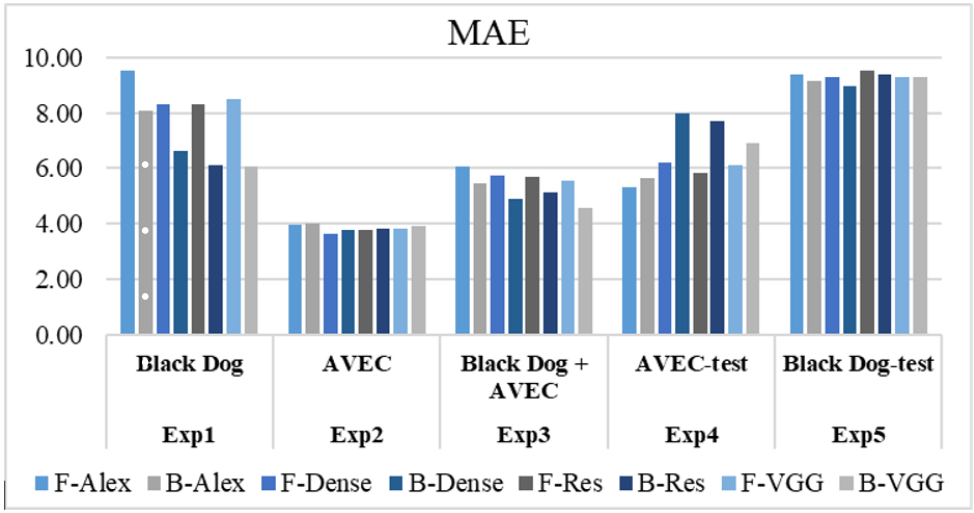
- Data Preprocessing:
  Face frames extraction vs upper body raw video frames
- Model Platform & Training Parameters:
  lr 10e-3, bs 20, epochs 20
- Datasets Split:
  Train/Valid/Test & BlackDog/AVEC2013
- Data Post-processing:
  MSE & RMSE average over video

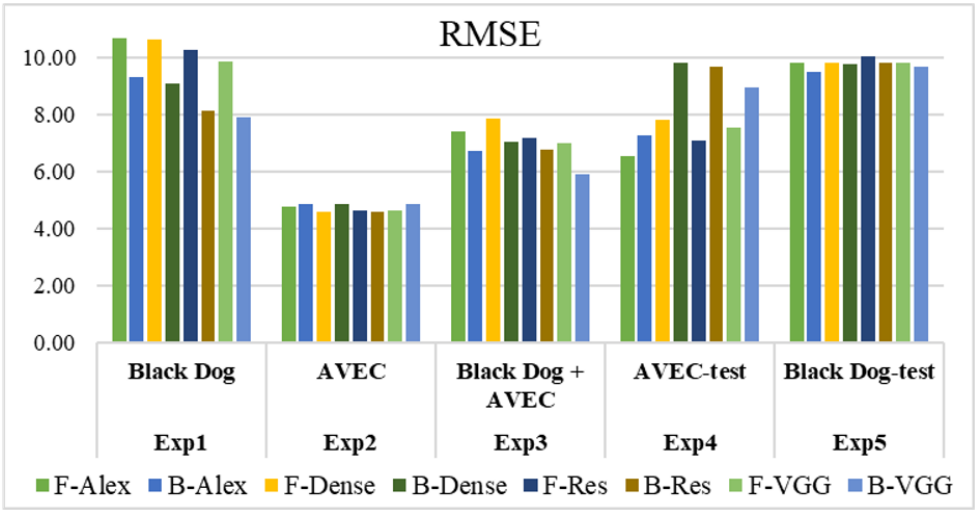
Exp#	Dataset	
Exp1	Black Dog	F
		B
Exp2	AVEC	F
		B
Exp3	Black Dog+AVEC	F
		B
Exp4	AVEC-test	F
		B
Exp5	Black Dog-test	F
		B

Results & Discussion

Models Performance

Results of all model architectures on all data





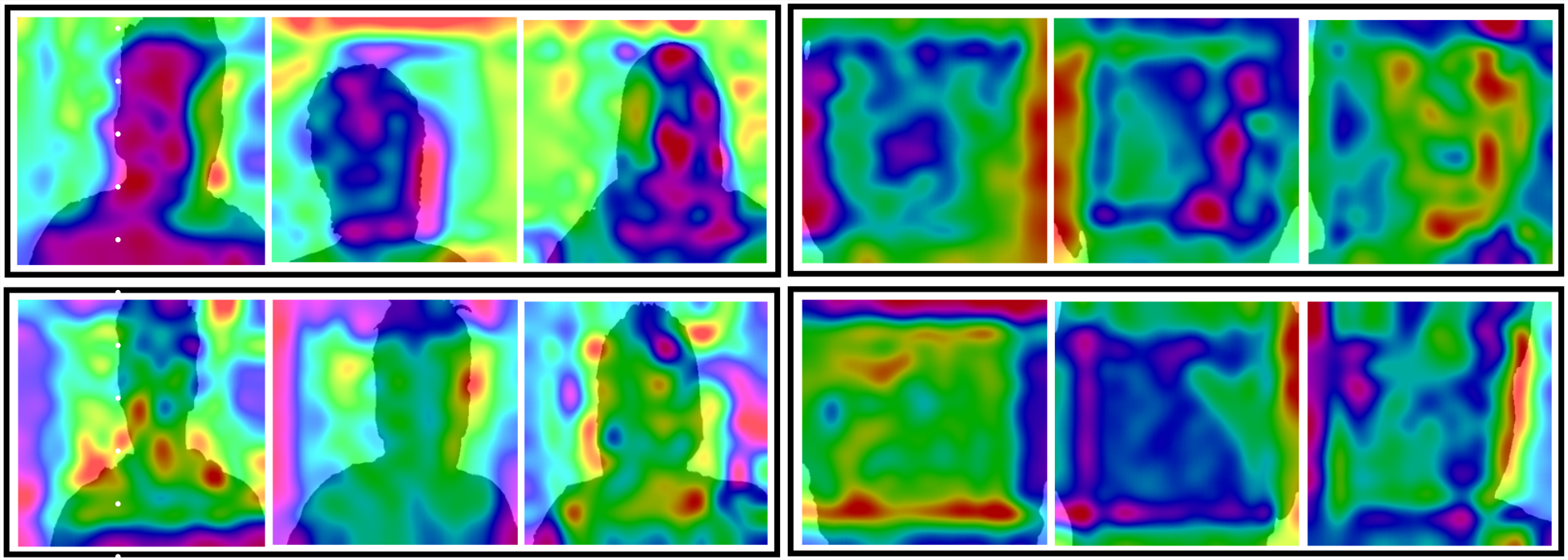
Face (F) and Body (B) Mean Average Error (MAE) & Root Mean Square Error (RMSE)

Alex = AlexNet, Dense = DenseNet169, Res = ResNet34, VGG = VGG19.

Results & Discussion

Activation Maps

Visualisation of activation maps for upper body frames (Left) and face only frames (Right). Predicted severity levels were close to the ground truth (Top row). Predicted severity levels were far from ground (Bottom row)



Conclusion

≈

Thank you for your time!