Moiz Tanvir
22I-1932
Deep Learning
Assignment 1 - Short Report

1. Network Details

This project implements a **multi-task convolutional neural network** to predict three outputs from face images:

- **Expression (exp):** 8-class categorical classification
- **Valence (val):** continuous regression
- **Arousal (aro):** continuous regression

Two pretrained backbones are evaluated:

- **ResNet50** and **VGG16**, both loaded with **ImageNet weights** and used as frozen feature extractors (include_top=False).
- A shared **global average pooling** layer feeds a 512-unit ReLU dense layer with 0.5 dropout.
- Three heads branch out:
  - exp: Dense(8, softmax)
  - val: Dense(1, linear)
  - aro: Dense(1, linear)

The model is compiled with **Adam optimizer**, multi-output losses (categorical_crossentropy for exp, mean-squared error for val and aro), and corresponding metrics (accuracy for exp, MSE for val/aro). Each network trains for **10 epochs** with a **batch size of 32**.

2. Dataset and Splits

The dataset directory contains images/ and per-image annotations/ for expression, valence, and arousal. After cleaning invalid labels (-2), the data are split as follows:

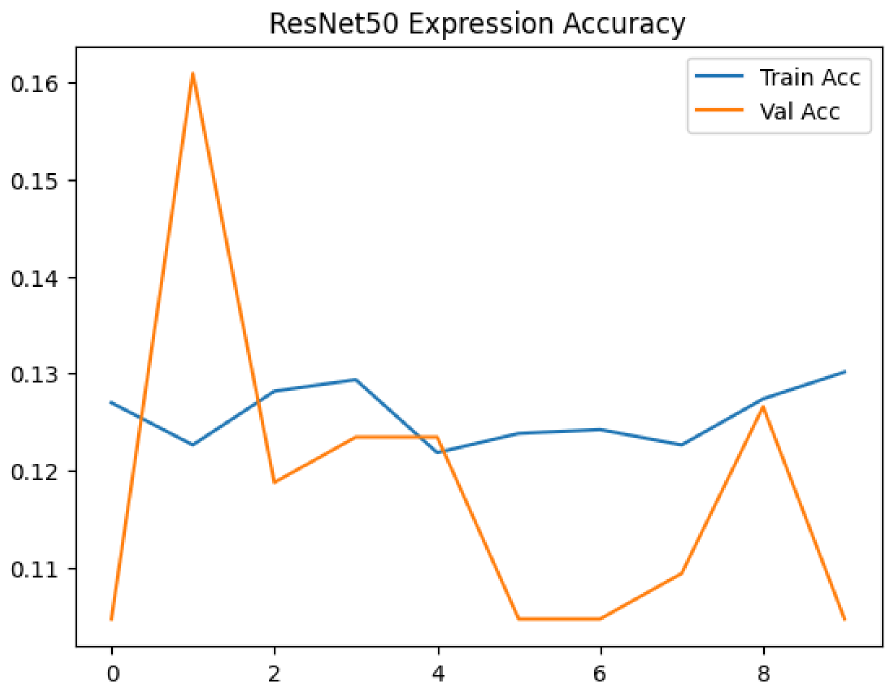
- **Training:** 64 % of total images
- **Validation:** 16 %
- **Test:** 20 %

A custom MultiOutputGenerator performs on-the-fly loading and optional augmentation (rotation, shifts, horizontal flips) for training batches.

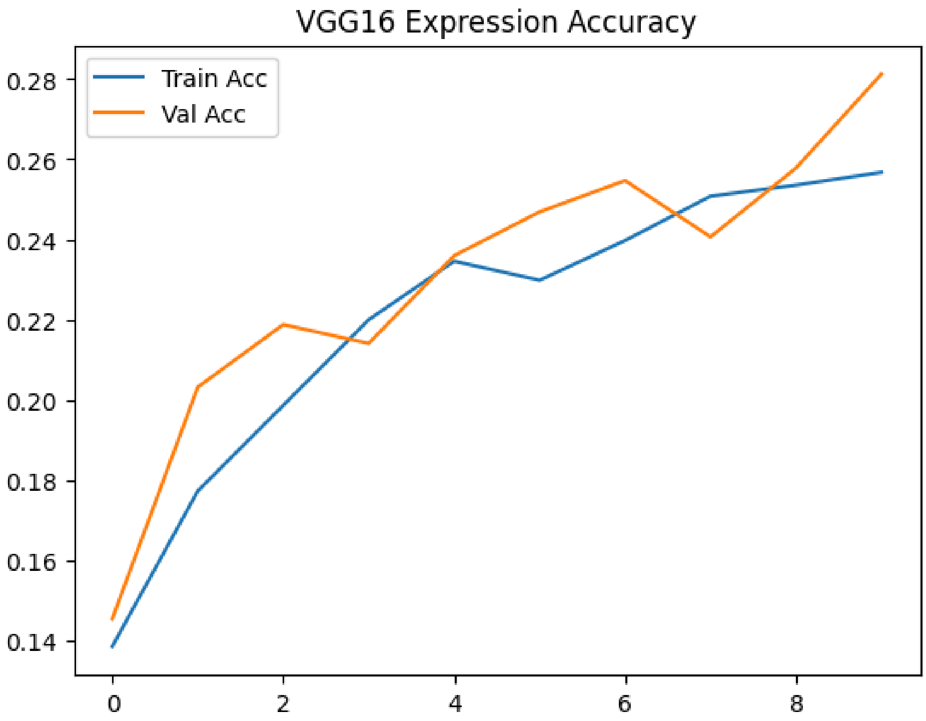
3. Training Graphs

During training, expression accuracy and loss curves were plotted:

- **ResNet50** showed low but slightly improving training/validation accuracy.



- **VGG16** displayed similar behavior but with higher overall validation loss and longer training time.



4. Performance Measures

For **classification**, the following metrics are computed on the held-out test set:

- Accuracy, Macro-F1, Cohen's Kappa, ROC-AUC, and mean Precision-Recall AUC.

For **regression** (valence & arousal):

- Root Mean Squared Error (RMSE), Pearson correlation (corr), Sign Agreement (SAGR), and Concordance Correlation Coefficient (CCC).

5. Results

Metric	ResNet50	VGG16
Classification Accuracy	**0.1200**	0.1200
F1 (macro)	0.0268	**0.1056**
Kappa	0.0000	-0.0062
ROC-AUC	**0.5376**	0.4924
Valence RMSE	**0.4813**	0.4962
Valence Corr	0.0032	**0.0523**
Valence CCC	0.0002	**0.0250**
Arousal RMSE	**0.3842**	0.3948
Arousal Corr	**0.0781**	0.0109
Arousal CCC	**0.0075**	0.0052
Train Time (s)	**1,194**	5,482

6. Discussion & Comparison

- Both models perform **near chance** for the 8-class expression task (random ≈ 12.5 %), indicating that 10 epochs and frozen backbones are insufficient for meaningful learning.

- Regression outputs show **low correlation** and small CCC, though SAGR values (~0.7) suggest moderate agreement on sign.

- **ResNet50** offers **similar or slightly better regression metrics** and finishes almost **4.5× faster** than VGG16, making it the more efficient backbone under identical settings.

- VGG16 shows marginally higher macro-F1 but requires much longer training time.

7. Conclusion

The multi-output CNN framework successfully integrates classification and regression heads but requires **further tuning**—such as unfreezing deeper layers, longer training, and larger datasets—to surpass baseline performance. Among the tested architectures, **ResNet50 is preferred** for its faster training and slightly stronger overall metrics.