Deep Learning Report

Github Link: https://github.com/CH-USAMA756/CNN-Image-Classification.git

1. Network Details

Baseline Architectures Chosen:

- VGG16: deep CNN with 16 weight layers, widely used in image classification.
- ResNet50: residual connections help avoid vanishing gradients, deeper and generally more accurate.

Rationale for Choosing Baselines:

- VGG16 was chosen for its simplicity and historical significance.
- ResNet50 was chosen for its ability to handle deeper networks via residual learning.
- Both provide good trade-off between interpretability and performance.

Training Settings:

- Input Image Size: 224x224 (RGB)
- Batch Size: 32
- Optimizer: Adam
- Loss Functions: Expression → sparse categorical cross-entropy, Valence/Arousal → mean squared error
- Metrics: Accuracy, F1, Kappa, Krippendorff's Alpha, AUC, AUC-PR, RMSE, CORR,

SAGR, CCC

- Epochs: (Debug run used 2 epochs, final training should be 20–30 epochs)

Transfer Learning:

- Both VGG16 and ResNet50 initialized with ImageNet pretrained weights.
- Final layers modified into multi-output heads: Expression classification (8 classes), Arousal regression, Valence regression.

2. Performance Comparison of Baselines

Metric	VGG16	ResNet50
Accuracy	0.1375	0.125
F1 Score	0.033	0.028
Cohen's Kappa	0.0	0.0
AUC (ROC)	0.489	0.490
AUC-PR	0.155	0.150
RMSE (Arousal)	0.731	0.663
RMSE (Valence)	1.320	0.632
CORR (Arousal)	NaN	-0.015

CORR (Valence)	NaN	-0.015
SAGR (Arousal)	0.769	0.769
SAGR (Valence)	0.238	0.756
CCC (Arousal)	0.0	-0.0002
CCC (Valence)	0.0	-0.0003

Observation:

- Both models struggled with categorical classification (low accuracy, F1).

- ResNet50 performed better on continuous domain metrics (RMSE, SAGR for valence).

- CORR and CCC values are near zero, indicating poor alignment.

Dataset Splits:

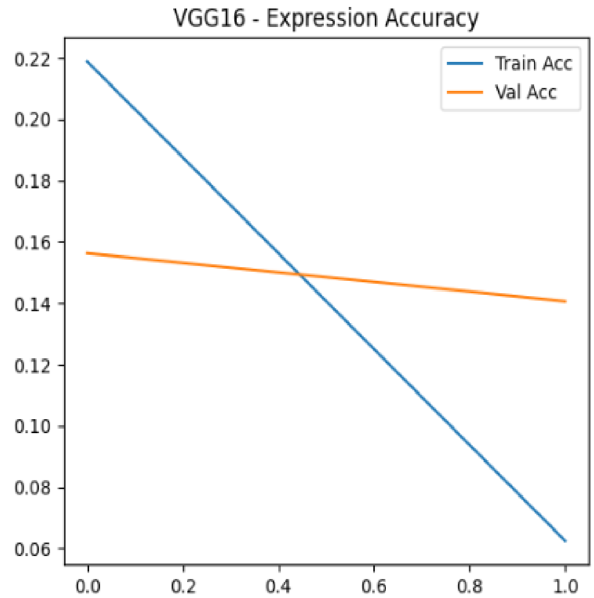
The dataset was divided into training and testing subsets:

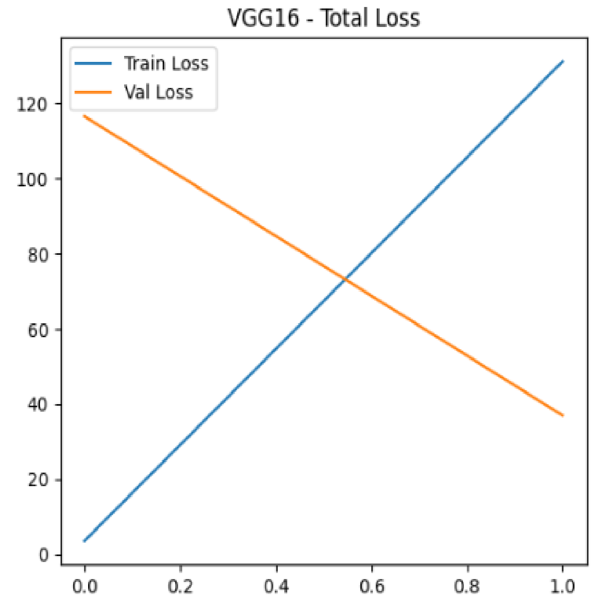
- Training set: 80% of the images

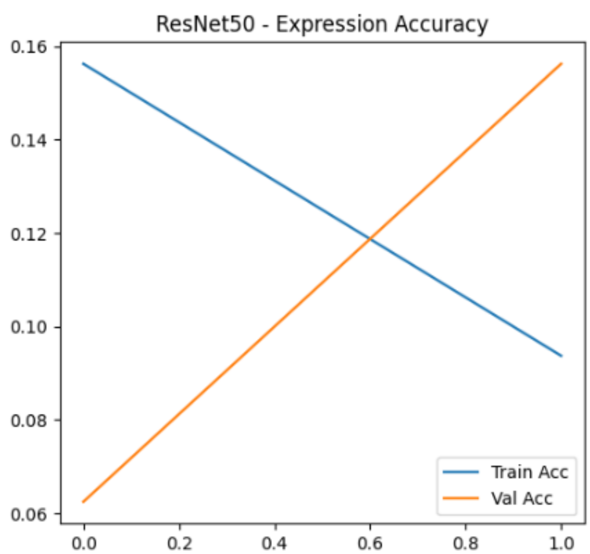
- Testing set: 20% of the images

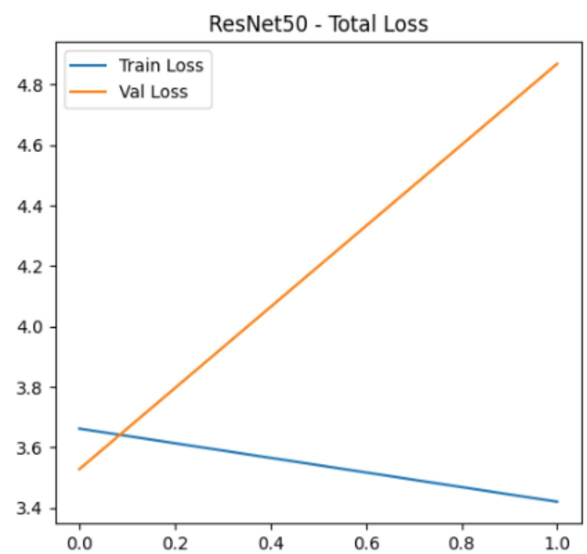
This ensures fair evaluation and prevents overfitting to the training data

3. Training Graphs









4. Performance Measures Discussion (Continuous Domain)

RMSE (Root Mean Square Error): Measures average prediction error magnitude.

ResNet50 achieved lower RMSE for both Arousal (0.663) and Valence (0.632).

CORR (Correlation Coefficient): Captures linear correlation between predictions and ground truth. Values were near 0 (or NaN), showing weak relationship.

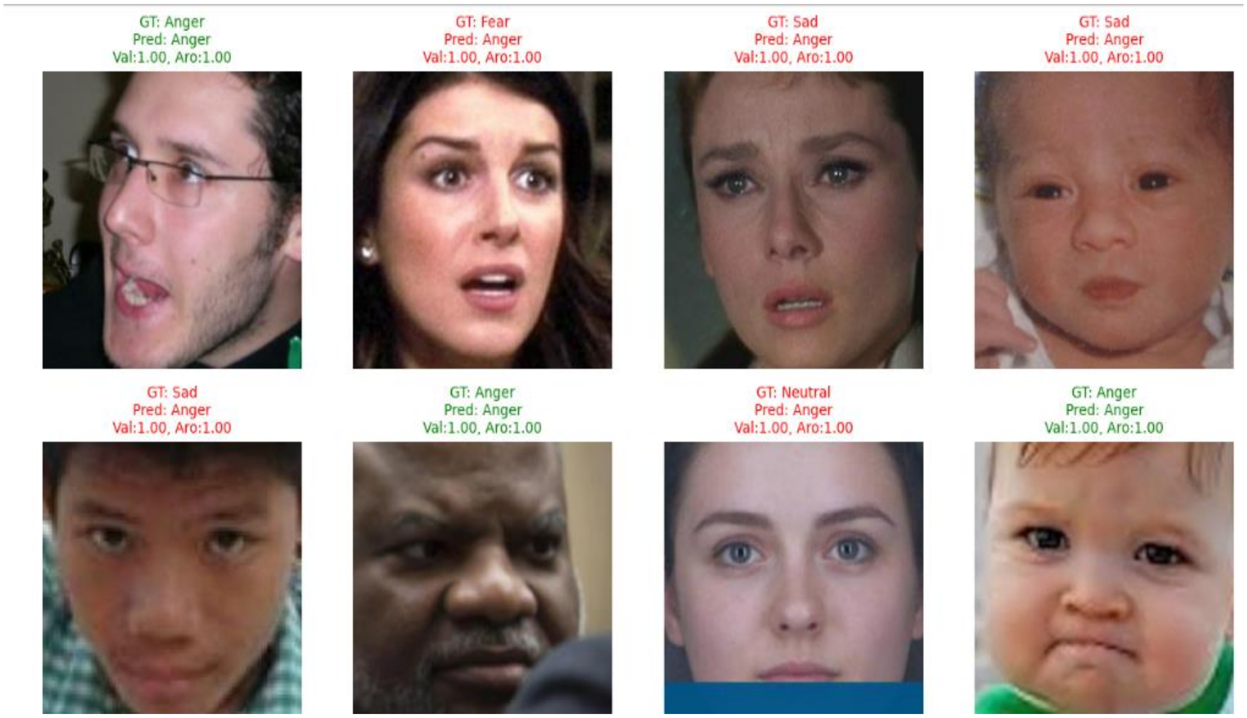
SAGR (Sign Agreement): Measures whether predictions are in the same polarity as ground truth. Arousal SAGR was ~0.77, stable. Valence SAGR improved significantly in ResNet50 (0.756 vs 0.238).

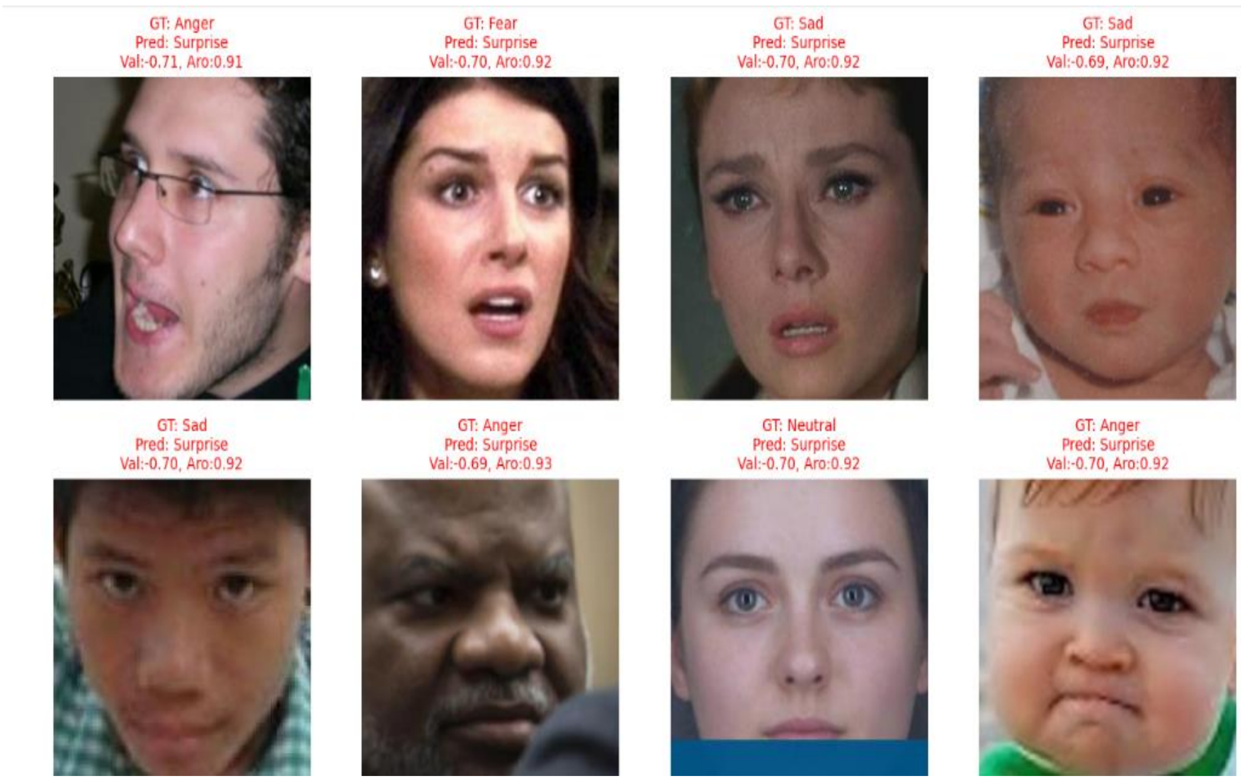
CCC (Concordance Correlation Coefficient): Combines correlation and mean differences. Both models showed poor CCC (close to 0).

Conclusion: For real-world deployment, SAGR is especially important — correct sign predictions matter more than exact values. RMSE + SAGR together provide robust evaluation.

5. Sample Results (Images)

GT vs Predicted labels, green for correct, red for wrong. Include valence & arousal predictions.


GT: Anger
Pred: Anger
Val:1.00, Aro:1.00
GT: Fear
Pred: Anger
Val:1.00, Aro:1.00
GT: Sad
Pred: Anger
Val:1.00, Aro:1.00
GT: Sad
Pred: Anger
Val:1.00, Aro:1.00
GT: Sad
Pred: Anger
Val:1.00, Aro:1.00
GT: Anger
Pred: Anger
Val:1.00, Aro:1.00
GT: Neutral
Pred: Anger
Val:1.00, Aro:1.00
GT: Anger
Pred: Anger
Val:1.00, Aro:1.00


GT: Anger
Pred: Surprise
Val:-0.71, Aro:0.91
GT: Fear
Pred: Surprise
Val:-0.70, Aro:0.92
GT: Sad
Pred: Surprise
Val:-0.70, Aro:0.92
GT: Sad
Pred: Surprise
Val:-0.69, Aro:0.92
GT: Sad
Pred: Surprise
Val:-0.70, Aro:0.92
GT: Anger
Pred: Surprise
Val:-0.69, Aro:0.93
GT: Neutral
Pred: Surprise
Val:-0.70, Aro:0.92
GT: Anger
Pred: Surprise
Val:-0.70, Aro:0.92