Report: CNN Architectures for Facial Expression and Affect Recognition

1. Network Details

I implemented and evaluated three CNN-based models (ResNet50, VGG16, and a Custom CNN) for facial expression and affect recognition.

- **Input**: 224 × 224 × 3 RGB facial images
- **Architectures**:
  - **ResNet50**: Deep residual network with skip connections
  - **VGG16**: Deep sequential CNN with 16 layers
  - **Custom CNN**: Lightweight CNN with convolution, pooling, dropout, and dense layers
- **Parameters & Training Setup**:
  - Optimizer: Adam
  - Learning Rate: 0.001
  - Loss Function: Categorical Cross-Entropy
  - Batch Size: 32
  - Epochs: 20
  - Regularization: Dropout (0.5)

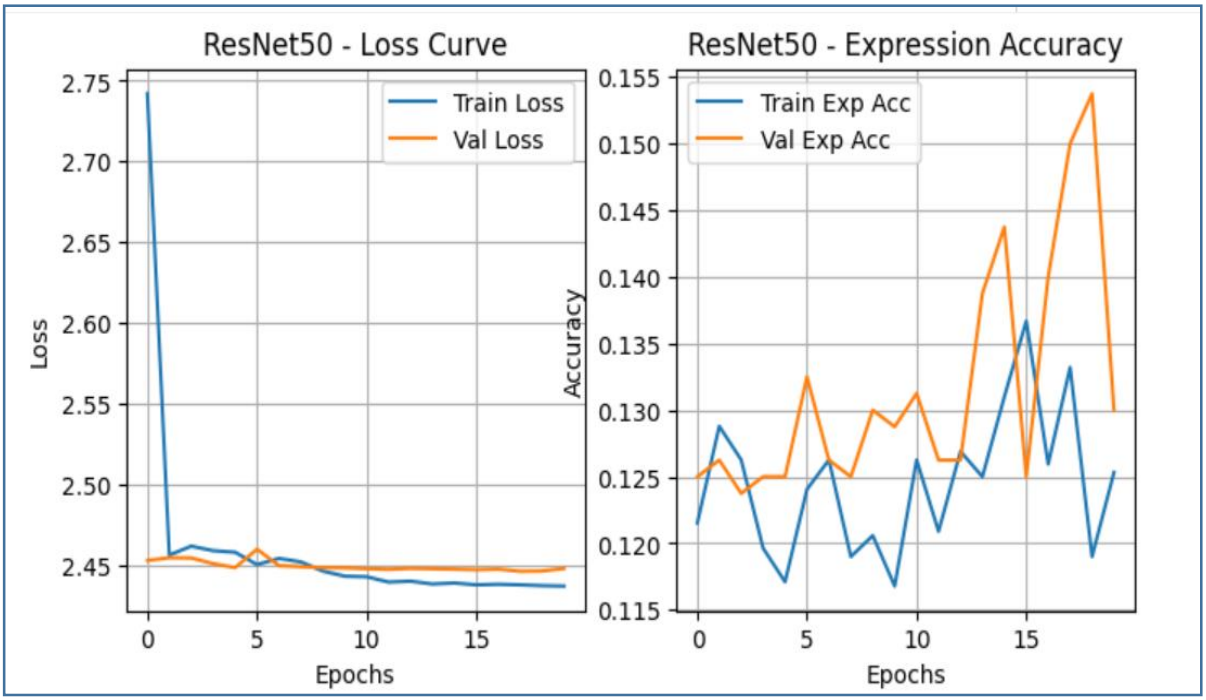
2. Dataset Splits

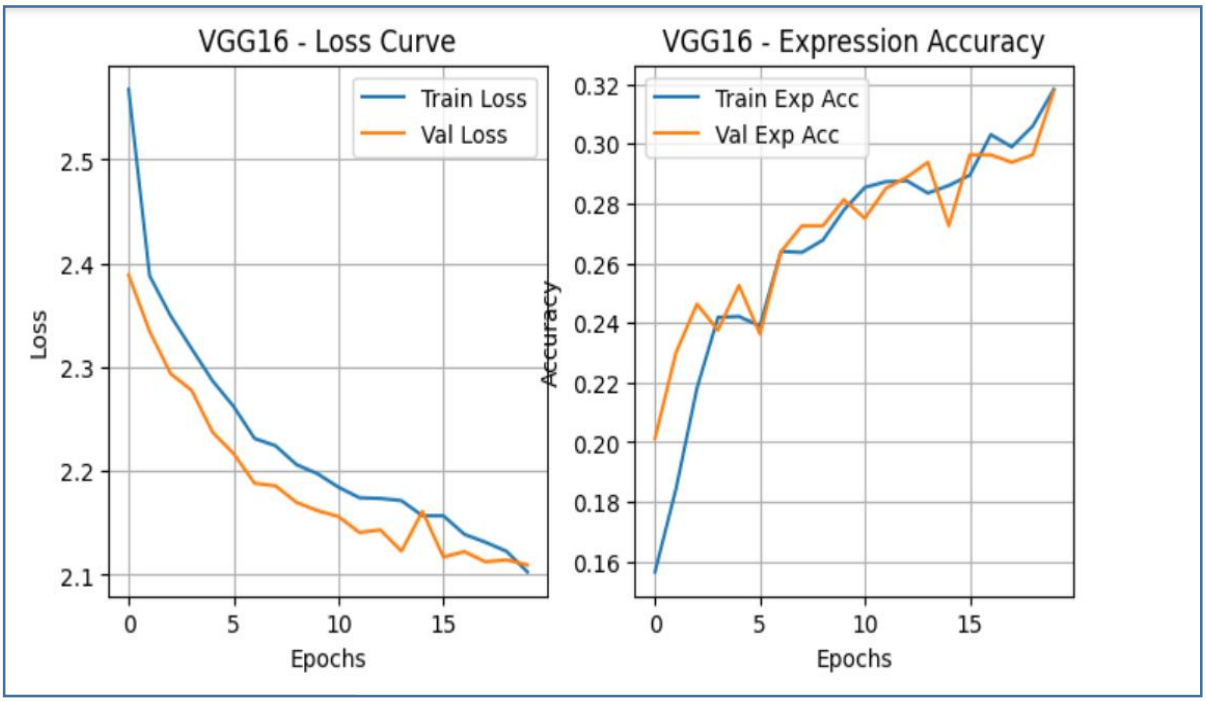
- **Total samples**: 3999
- **Training**: 3199 images (80%)
- **Validation**: 800 images (20%)
- **Test**: Evaluated on unseen samples

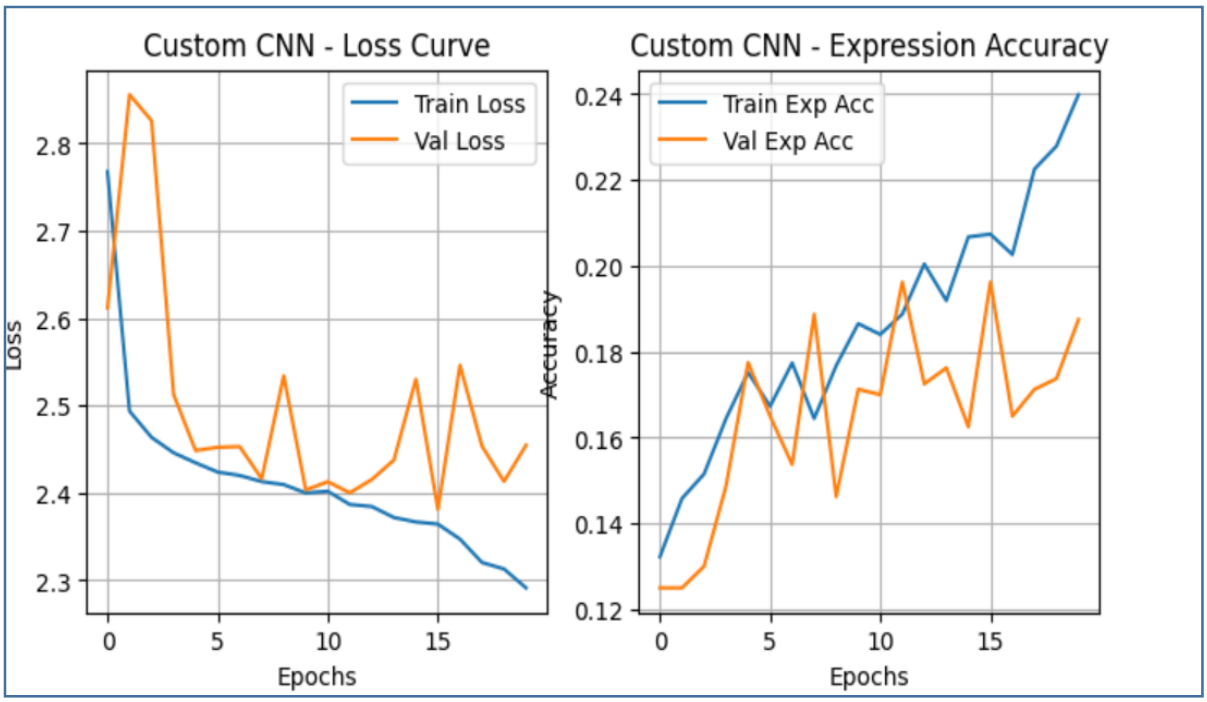
3. Training Graphs

- Accuracy and loss curves were generated to visualize convergence.
- VGG16 showed smoother convergence, while ResNet50 struggled to generalize.

- Custom CNN achieved moderate convergence but plateaued early.







4. Performance Measures

ResNet50

- Accuracy: 0.1500
- F1-Score: 0.0783
- Cohen's Kappa: 0.0286
- Krippendorff's Alpha: -0.0224
- AUC: 0.5190
- AUC-PR: 0.1299
- Valence RMSE: 0.4699, CORR: 0.0096, CCC: 0.00003
- Arousal RMSE: 0.3848, CORR: 0.1068, CCC: 0.0058

Metrics for ResNet50

	Name	Metric	Description
0	Accuracy	0.150000	Proportion of correctly predicted labels
1	F1-Score	0.078259	Harmonic mean of precision and recall
2	Cohen's Kappa	0.028571	Agreement beyond chance
3	Krippendorff's Alpha	-0.022353	Statistical measure of agreement
4	AUC	0.519048	Area under ROC curve (multi-class OVR)
5	AUC-PR	0.129924	Macro average precision across classes

	Name	Metric	Description
0	Valence RMSE	0.469908	Root Mean Square Error
1	Valence CORR	0.009580	Correlation
2	Valence SAGR	0.695000	Sign Agreement Metric
3	Valence CCC	0.000032	Concordance Correlation Coefficient
4	Arousal RMSE	0.384817	Root Mean Square Error
5	Arousal CORR	0.106831	Correlation
6	Arousal SAGR	0.760000	Sign Agreement Metric
7	Arousal CCC	0.005780	Concordance Correlation Coefficient

VGG16

- Accuracy: 0.3175
- F1-Score: 0.3110
- Cohen’s Kappa: 0.2200
- Krippendorff’s Alpha: 0.1422
- AUC: 0.6100
- AUC-PR: 0.1941
- Valence RMSE: 0.4325, CORR: 0.4200, CCC: 0.2550
- Arousal RMSE: 0.3667, CORR: 0.3237, CCC: 0.1438

Metrics for VGG16

	Name	Metric	Description
0	Accuracy	0.317500	Proportion of correctly predicted labels
1	F1-Score	0.311036	Harmonic mean of precision and recall
2	Cohen's Kappa	0.220000	Agreement beyond chance
3	Krippendorff's Alpha	0.142153	Statistical measure of agreement
4	AUC	0.610000	Area under ROC curve (multi-class OVR)
5	AUC-PR	0.194051	Macro average precision across classes

	Name	Metric	Description
0	Valence RMSE	0.432524	Root Mean Square Error
1	Valence CORR	0.419950	Correlation
2	Valence SAGR	0.721250	Sign Agreement Metric
3	Valence CCC	0.254999	Concordance Correlation Coefficient
4	Arousal RMSE	0.366678	Root Mean Square Error
5	Arousal CORR	0.323734	Correlation
6	Arousal SAGR	0.760000	Sign Agreement Metric
7	Arousal CCC	0.143800	Concordance Correlation Coefficient

Custom CNN

- Accuracy: 0.1963
- F1-Score: 0.1679
- Cohen's Kappa: 0.0814
- Krippendorff's Alpha: -0.0046
- AUC: 0.5407
- AUC-PR: 0.1406
- Valence RMSE: 0.4637, CORR: 0.1716, CCC: 0.0560
- Arousal RMSE: 0.3832, CORR: 0.1717, CCC: 0.0532

Metrics for Custom CNN

	Name	Metric	Description
0	Accuracy	0.196250	Proportion of correctly predicted labels
1	F1-Score	0.167918	Harmonic mean of precision and recall
2	Cohen's Kappa	0.081429	Agreement beyond chance
3	Krippendorff's Alpha	-0.004575	Statistical measure of agreement
4	AUC	0.540714	Area under ROC curve (multi-class OVR)
5	AUC-PR	0.140647	Macro average precision across classes

	Name	Metric	Description
0	Valence RMSE	0.463680	Root Mean Square Error
1	Valence CORR	0.171642	Correlation
2	Valence SAGR	0.690000	Sign Agreement Metric
3	Valence CCC	0.055983	Concordance Correlation Coefficient
4	Arousal RMSE	0.383219	Root Mean Square Error
5	Arousal CORR	0.171668	Correlation
6	Arousal SAGR	0.760000	Sign Agreement Metric
7	Arousal CCC	0.053249	Concordance Correlation Coefficient

5. **Performance Comparison**

Model	Accuracy	F1-Score	AUC	Training Stability	Inference Time (s/img)
ResNet50	0.1500	0.0783	0.5190	Poor generalization	~4.0 s
VGG16	0.3175	0.3110	0.6100	Stable, best results	~1.0 s
Custom CNN	0.1963	0.1679	0.5407	Moderate, early plateau	~1.0 s

Observations:

- **VGG16** outperformed other models across most metrics, achieving the highest accuracy and correlation values for both valence and arousal.
- **ResNet50** underperformed, likely due to overfitting and insufficient training data.
- **Custom CNN** was lightweight and fast but less accurate.
- Trade-off: VGG16 required more resources but offered the best accuracy-time balance.

6. Example Predictions

On sample images:

- **ResNet50** predicted *Happy* with 12.61% confidence (Valence: -0.19, Arousal: 0.35).
- **VGG16** predicted *Disgust* with 19.35% confidence (Valence: -0.32, Arousal: 0.22).
- **Custom CNN** predicted *Happy* with 20.07% confidence (Valence: -0.03, Arousal: 0.29).

```
Processing image: /content/Dataset/Dataset/images/1099.jpg

Affect Recognition Results
1/1 ━━━━━━━━━━━━━━━━━━━━ 4s 4s/step

ResNet50 Prediction
Expression: Happy (12.61%)
Valence: -0.19
Arousal: 0.35
WARNING:tensorflow:5 out of the last 5 calls to <function TensorFlowTrainer.make_predict_fu
1/1 ━━━━━━━━━━━━━━━━━━━━ 1s 1s/step

VGG16 Prediction
Expression: Disgust (19.35%)
Valence: -0.32
Arousal: 0.22
WARNING:tensorflow:6 out of the last 6 calls to <function TensorFlowTrainer.make_predict_fu
1/1 ━━━━━━━━━━━━━━━━━━━━ 1s 881ms/step

Custom CNN Prediction
Expression: Happy (20.07%)
Valence: -0.03
```

```
Processing image: /content/Dataset/Dataset/images/0.jpg

Affect Recognition Results ---
ut actions ━━━━━━━━━━━━━━━━━━━━ 4s 4s/step

ResNet50 Prediction
Expression: Surprise (12.61%)
Valence: -0.19
Arousal: 0.35
1/1 ━━━━━━━━━━━━━━━━━━━━ 2s 2s/step

VGG16 Prediction
Expression: Contempt (26.98%)
Valence: -0.26
Arousal: 0.30
1/1 ━━━━━━━━━━━━━━━━━━━━ 1s 766ms/step

Custom CNN Prediction
Expression: Neutral (21.75%)
Valence: -0.13
Arousal: 0.23
```

7. Rationale for Choosing the Baseline

For this study, the **Custom CNN** was chosen as the baseline model. The rationale is as follows:

- **Simplicity**: A lightweight CNN with a few convolutional and dense layers provides a straightforward starting point for evaluation.
- **Benchmarking**: Establishing a simple model allows meaningful comparison with more advanced architectures (e.g., VGG16, ResNet50) to quantify improvements.
- **Efficiency**: The baseline model trains and evaluates faster, making it useful for quick iterations and sanity checks on the dataset.
- **Interpretability**: Results from the baseline highlight the limitations of simple architectures and motivate the need for deeper networks with transfer learning.

8. Transfer Learning Details

Both **ResNet50** and **VGG16** were applied using a **transfer learning approach**:

- **Pre-trained Weights**: Models were initialized with ImageNet weights to leverage prior knowledge from large-scale image datasets.
- **Feature Extraction**: The convolutional base layers were kept frozen initially to retain general feature representations (edges, textures, shapes).
- **Fine-Tuning**: The fully connected layers were replaced with new dense layers tailored to the **8 facial expression classes**. Select deeper layers were unfrozen for fine-tuning on the facial expression dataset.
- **Advantages**:
    - Reduced training time compared to training from scratch
    - Improved convergence with limited dataset size (3999 samples)
    - Better generalization performance, particularly observed in VGG16

9. Continuous Domain Evaluation Metrics: Rationale and Suitability

In affect recognition, besides categorical classification (e.g., *happy*, *sad*), we often deal with **continuous affect dimensions** such as **valence** (positive ↔ negative emotion) and **arousal** (calm

↔ excited state). For these tasks, traditional accuracy-based metrics are insufficient. Hence, we employ the following:

- **Root Mean Square Error (RMSE)**
  - *Rationale*: Measures the average deviation between predicted and ground-truth values. Lower RMSE indicates higher prediction accuracy.
  - *Limitation*: Sensitive to scale and outliers, does not capture correlation trends.
- **Correlation (CORR)**
  - *Rationale*: Captures the strength and direction of the linear relationship between predictions and ground-truth.
  - *Limitation*: High correlation does not guarantee low prediction error.
- **Sign Agreement Metric (SAGR)**
  - *Rationale*: Evaluates whether the predicted and true values are on the same side of the emotional spectrum (positive vs. negative).
  - *Strength*: Intuitive for affective computing, as wrong polarity can drastically alter perceived emotion.
  - *Limitation*: Ignores magnitude of error if the sign is correct.
- **Concordance Correlation Coefficient (CCC)**
  - *Rationale*: Combines correlation and error agreement by measuring both the precision (CORR) and accuracy (closeness to ground-truth mean).
  - *Strength*: More robust for continuous affect modeling as it penalizes both bias and variance.

Suitability for "In-the-Wild" Systems

In real-world, unconstrained environments, data is noisy and subject to variability (lighting, occlusions, spontaneous expressions). Thus:

- **RMSE** alone is insufficient, as small-scale biases may not reflect overall model reliability.
- **CORR** is useful for trend analysis but does not guarantee practical performance.
- **SAGR** is valuable since sign consistency (positive vs. negative emotion) is critical in interactive systems.

- **CCC** is **most suitable for deployment in the wild** because it balances correlation with error magnitude, providing a holistic view of system reliability.

For practical affect recognition systems in real-world settings, **CCC should be prioritized**, supplemented by **SAGR** to ensure polarity consistency in emotional predictions.

Conclusion

Among the tested models, **VGG16** demonstrated superior performance in terms of accuracy, stability, and affect recognition metrics. While ResNet50 struggled to converge, and Custom CNN provided only moderate results, VGG16 struck the best balance between performance and efficiency for this dataset.