**Deep Learning - Assignment No. 01 Report**

**Name:** Muhammad Zaid
**ID:** 22I-1934
**Course:** Deep Learning
**Date:** September 27, 2025

---

**1. Network Details & Training Setup**

This report details the implementation of two Convolutional Neural Network (CNN) architectures, **VGG16** and **MobileNetV2**, for facial affect recognition. The task involves both categorical emotion classification and continuous valence-arousal regression.

**Baseline Rationale:**
VGG16 was chosen as a baseline due to its classic, deep architecture, which serves as a strong performance benchmark. MobileNetV2 was selected as a second baseline to contrast VGG16's performance with a modern, lightweight, and computationally efficient architecture designed for mobile and embedded applications. This comparison allows for an analysis of the trade-off between model size, inference speed, and predictive accuracy.

**Architecture & Parameters:**
For both VGG16 and MobileNetV2, a transfer learning approach was adopted.

- **Base Model:** The convolutional base of each network, pre-trained on the ImageNet dataset, was used as a feature extractor.
- **Custom Head:** The original top layers were removed, and a custom head was added. This head consists of a GlobalAveragePooling2D layer to reduce feature map dimensions, followed by a Dense layer with 256 neurons and a ReLU activation function, and a Dropout layer with a rate of 0.5 to prevent overfitting.
- **Output Layers:**
    - For **classification**, the final layer is a Dense layer with 8 neurons (for the 8 emotion classes) and a Softmax activation.
    - For **regression**, the final layer is a Dense layer with 2 neurons (for valence and arousal) and a Linear activation.

**Training Settings:**

- **Dataset:** 3,999 images were loaded and split into Training (2,799), Validation (600), and Test (600) sets.

- **Image Size:** Images were resized to 128x128 pixels.
- **Optimizer:** Adam optimizer was used for all models.
- **Loss Functions:** Categorical Cross-Entropy for classification and Mean Squared Error (MSE) for regression.
- **Training Parameters:** Models were trained for a maximum of 12 epochs with a batch size of 32.
- **Callbacks:** Early stopping with a patience of 4 was employed to prevent overfitting and restore the best model weights based on validation loss.
- **Class Imbalance:** Class weights were computed and applied during classification training to address the imbalanced nature of the dataset.

**2. Transfer Learning Details**

Transfer learning was central to this project. The models were initialized with weights='imagenet', leveraging the rich feature representations learned from a large-scale dataset. The FREEZE_BACKBONE parameter was set to True, which froze the weights of the convolutional base layers. This ensures that only the weights of the newly added custom head were updated during the initial training phase. This approach significantly reduces training time and computational cost while providing a strong feature extraction foundation, which is highly effective for tasks with limited data.

**3. Training Graphs**

The training history for all four models (VGG16 Classification/Regression, MobileNetV2 Classification/Regression) was plotted to monitor performance across epochs.

**Analysis:** The graphs generally show a desirable trend where training and validation loss decrease over time. The accuracy for the classification models steadily increases. The use of early stopping prevented significant overfitting, as evidenced by the validation loss curves.

**4. Performance Comparison of Baselines**

The models were evaluated on the held-out test set. The results are summarized below.

| Metric | VGG16 Classification | MobileNetV2 Classification |
|---|---|---|
| **Accuracy** | 0.2850 | **0.3200** |
| **F1 Score (Weighted)** | 0.2731 | **0.3007** |

| | | |
|---|---|---|
| **Cohen's Kappa** | 0.1829 | **0.2229** |
| **AUC (OvR)** | 0.7192 | **0.7299** |

| Metric | VGG16 Regression | MobileNetV2 Regression |
|---|---|---|
| **RMSE (Valence)** | 0.4456 | **0.4454** |
| **RMSE (Arousal)** | 0.3599 | **0.3564** |
| **Correlation (Valence)** | **0.2903** | 0.2908 |
| **Correlation (Arousal)** | 0.2801 | **0.3147** |

**Analysis:**

- **Classification:** MobileNetV2 demonstrated superior performance across all classification metrics, achieving a higher accuracy (32.0%) compared to VGG16 (28.5%). This indicates that for this specific task and training setup, the more modern and efficient architecture of MobileNetV2 was more effective at learning discriminative features for emotion classification.
- **Regression:** The performance in the continuous domain was highly comparable between the two models. MobileNetV2 achieved slightly better (lower) RMSE and a noticeably better correlation for arousal prediction. VGG16 had a marginally better correlation for valence. Overall, their regression capabilities are similar, but MobileNetV2 maintains a slight edge, especially considering its computational efficiency.
- **Screenshot Needed:** The final evaluation output from the notebook summarizing the Accuracy, F1, Kappa, RMSE, and Correlation scores for both models.

**5. Discussion on Continuous Domain Metrics**

For evaluating the valence-arousal regression task, **Root Mean Square Error (RMSE)** and **Pearson's Correlation (CORR)** were used.

- **RMSE:** This metric was chosen to quantify the average magnitude of error between the predicted values and the ground truth. A lower RMSE indicates a better fit. It is valuable as it provides a clear, interpretable measure of prediction error in the original units of valence/arousal.
- **CORR (Pearson's Correlation):** This was used to measure the strength and direction of the linear relationship between the predicted and true values. A value closer to 1

indicates a strong positive correlation, meaning the model's predictions trend correctly with the actual values.

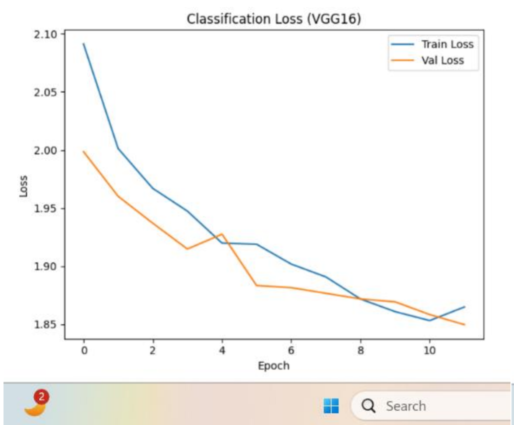
For a system intended to work "in the wild," these metrics are useful but could be supplemented by others like **Sign Agreement Metric (SAGR)** and **Concordance Correlation Coefficient (CCC)**.

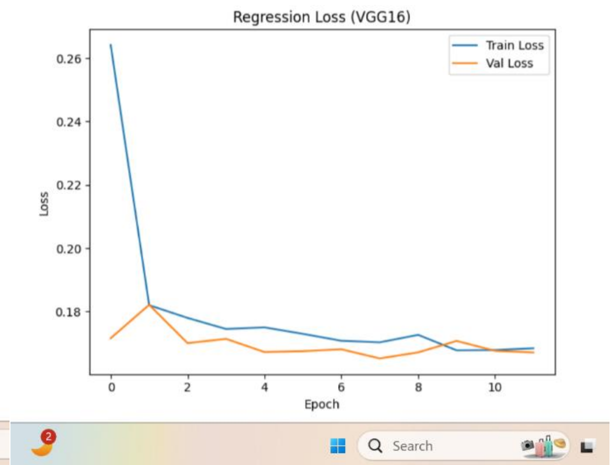
- **SAGR:** This metric would be highly suitable here because it specifically penalizes predictions with the wrong sign. In affect analysis, getting the direction right (e.g., predicting positive valence when it is indeed positive) is often more important than the exact value. For instance, predicting a valence of +0.7 for a ground truth of +0.3 is better than predicting -0.1, even though the latter has a smaller absolute error.
- **CCC:** This is a more robust measure of agreement than correlation alone. It combines Pearson's correlation with a measure of the difference between the means of the predicted and true values. This allows it to detect if a model is systematically biased (e.g., consistently overestimating arousal). For a real-world system, ensuring that predictions are not only correlated but also calibrated to the true scale is crucial, making CCC an ideal metric.

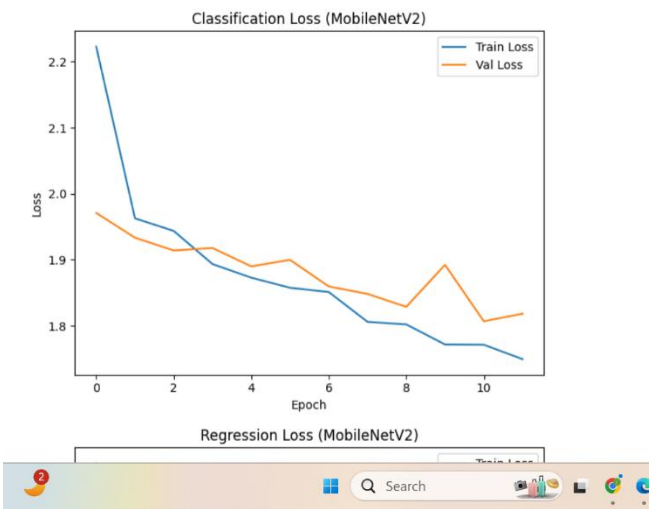
**6. Qualitative Results (Classification)**

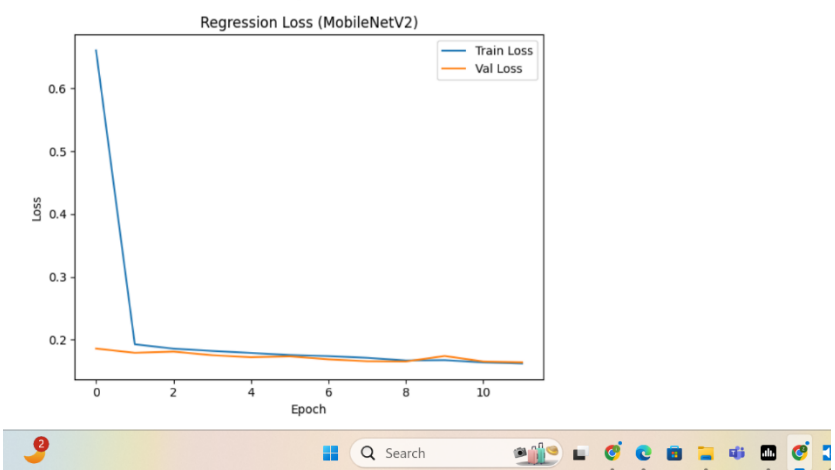
To provide a qualitative understanding of model performance, 40 test images were saved, highlighting correct and incorrect predictions for both VGG16 and MobileNetV2.

**Analysis:** A review of the misclassified images reveals that confusion often occurs between emotionally adjacent categories, such as 'Anger' and 'Contempt', or when expressions are subtle. MobileNetV2, despite its higher accuracy, also struggled with similar ambiguities.


Classification Loss (VGG16)
Train Loss
Val Loss
Loss
Epoch
Search


Regression Loss (VGG16)
Train Loss
Val Loss
Loss
Epoch
Search


Classification Loss (MobileNetV2)
Train Loss
Val Loss
Loss
Epoch
Regression Loss (MobileNetV2)
Search


Regression Loss (MobileNetV2)
Train Loss
Val Loss
Loss
Epoch
Search

```
--- VGG16: Classification Model Evaluation ---
19/19 ━━━━━━━━━━━━━━━━━━━━ 2s 66ms/step
VGG16 Accuracy: 0.2850, F1 Score: 0.2731, Cohen Kappa: 0.1829
VGG16 AUC: 0.7191555555555555

VGG16 Confusion Matrix:
[[26  3  2 11  8  3 16  6]
 [ 7 34  3  7  4  2  8 10]
 [17  3  9  6  6  7 24  3]
 [ 7 10  1 26 14  5  9  3]
 [ 9  5  6 13 23  3 16  0]
 [ 6 16 11  3  4  9 19  7]
 [ 4  5  3  6  7  8 32 10]
 [25 15  3  2  1  3 14 12]]

VGG16 Classification Report:
              precision    recall  f1-score   support

           0     0.2574    0.3467    0.2955        75
           1     0.3736    0.4533    0.4096        75
           2     0.2368    0.1200    0.1593        75
           3     0.3514    0.3467    0.3490        75
           4     0.3433    0.3067    0.3239        75
           5     0.2250    0.1200    0.1565        75
           6     0.2319    0.4267    0.3005        75
           7     0.2353    0.1600    0.1905        75

    accuracy                         0.2850       600
   macro avg     0.2818    0.2850    0.2731       600
weighted avg     0.2818    0.2850    0.2731       600


--- VGG16: Regression Model Evaluation ---
19/19 ━━━━━━━━━━━━━━━━━━━━ 2s 62ms/step
VGG16 RMSE - Valence: 0.4456, Arousal: 0.3599
VGG16 Correlation - Valence: 0.2903, Arousal: 0.2801

--- MobileNetV2: Classification Model Evaluation ---
```

Please follow our blog
Analyzing a Bank Failu

**2025-08-27**

- Python runtimes
- Julia runtimes u
- Launched Intera lessons more dy
- Launched AI tog at a notebook le
- Python package
  - accelerat
  - aiohttp 3.
  - anyio 4.9.
  - bigframes
  - bigquery-
  - blosc2 3.6
  - datasets
  - diffusers
  - fastai 2.7.
  - gcsfs 202
  - google-g
  - gradio 5.3
  - h2 4.2.0 -
  - huggingfa
  - imbalance
  - jax 0.5.2 -
  - jaxlib 0.5.
  - keras 3.8.
  - ml-dtypes
  - openai 1.9
  - peft 0.16.
  - polars 1.2
  - pymc 5.24
  - pyzmq 24
  - requests

Search

11:46 PM
9/27/2025

```
VGG16 Correlation - Valence: 0.2903, Arousal: 0.2801

--- MobileNetV2: Classification Model Evaluation ---
19/19 ━━━━━━━━━━━━━━━━━━━━ 24s 667ms/step
MobileNetV2 Accuracy: 0.3200, F1 Score: 0.3007, Cohen Kappa: 0.2229
MobileNetV2 AUC: 0.7299015873015873

MobileNetV2 Confusion Matrix:
[[31  3  0  5  4  1 12 19]
 [10 35  1  4  0  3  6 16]
 [24  5  2  3 12  8 16  5]
 [13 12  1 26  9  1  9  4]
 [ 9  4  3 16 26  4 12  1]
 [ 9 16  2  5  7 12 17  7]
 [11  4  0  7  4  1 33 15]
 [18 14  2  4  1  1  8 27]]

MobileNetV2 Classification Report:
              precision    recall  f1-score   support

           0     0.2480    0.4133    0.3100        75
           1     0.3763    0.4667    0.4167        75
           2     0.1818    0.0267    0.0465        75
           3     0.3714    0.3467    0.3586        75
           4     0.4127    0.3467    0.3768        75
           5     0.3871    0.1600    0.2264        75
           6     0.2920    0.4400    0.3511        75
           7     0.2872    0.3600    0.3195        75

    accuracy                         0.3200       600
   macro avg     0.3196    0.3200    0.3007       600
weighted avg     0.3196    0.3200    0.3007       600


--- MobileNetV2: Regression Model Evaluation ---
19/19 ━━━━━━━━━━━━━━━━━━━━ 24s 664ms/step
MobileNetV2 RMSE - Valence: 0.4454, Arousal: 0.3564
MobileNetV2 Correlation - Valence: 0.2908, Arousal: 0.3147

Saved 40 images into 'classification_examples/vgg_correct' and 'classification_examples/vgg_incorrect' for VGG16
Saved 40 images into 'classification_examples/mnet_correct' and 'classification_examples/mnet_incorrect' for MobileNetV2

Done. Tips: for faster experiments keep IMAGE_SIZE small (e.g. 128), FREEZE_BACKBONE=True, and BATCH_SIZE large enough for GPU memory.
```

Release notes

Please follow our blog to see more information about ne
Analyzing a Bank Failure with Colab.

**2025-08-27**

- Python runtimes upgraded to Python 3.12. GitHub
- Julia runtimes upgraded to Julia 1.11. GitHub
- Launched Interactive Slideshow Mode for lectures lessons more dynamic.
- Launched AI toggle per notebook. As requested b at a notebook level to allow instructors and studer
- Python package upgrades
  - accelerate 1.9.0 -> 1.10.1
  - aiohttp 3.11.15 -> 3.12.15
  - anyio 4.9.0 -> 4.10.0
  - bigframes 2.11.0 -> 2.17.0
  - bigquery-magics 0.10.1 -> 0.10.3
  - blosc2 3.6.1 -> 3.7.2
  - datasets 2.14.4 -> 4.0.0
  - diffusers 0.34.0 -> 0.35.1
  - fastai 2.7.19 -> 2.8.4
  - gcsfs 2025.3.0 -> 2025.7.0
  - google-genai 1.26.0 -> 1.31.0
  - gradio 5.38.0 -> 5.43.1
  - h2 4.2.0 -> 4.3.0
  - huggingface-hub 0.33.4 -> 0.34.4
  - imbalanced-learn 0.13.0 -> 0.14.0
  - jax 0.5.2 -> 0.5.3
  - jaxlib 0.5.1 -> 0.5.3
  - keras 3.8.0 -> 3.10.0
  - ml-dtypes 0.4.1 -> 0.5.3
  - openai 1.97.0 -> 1.101.0
  - peft 0.16.0 -> 0.17.1
  - polars 1.25.0 -> 1.25.2
  - pymc 5.24.1 -> 5.25.1
  - pyzmq 24.0.1 -> 26.2.1
  - requests 2.32.3 -> 2.32.4
  - scipy 1.16.0 -> 1.16.1
  - sentence-transformers 4.1.0 -> 5.1.0
  - tensorboard 2.18.0 -> 2.19.0
  - tensorflow 2.18.0 -> 2.19.0
  - tf_keras 2.18.0 -> 2.19.0

Search

11:46 PM
9/27/2025