Muhammad Subhan
22I2024
DSC

Assignment 1 Report
Deep Learning

Emotion Recognition Using CNNs: EfficientNetB0 vs ResNet50

1. Network Details

1.1 EfficientNetB0

- **Architecture:** EfficientNetB0 backbone (pretrained on ImageNet)
- **Output heads:**
    - Expression classification (softmax, 7 classes)
    - Valence regression (linear)
    - Arousal regression (linear)
- **Parameters:** ~5.3M
- **Training configuration:**
    - Phase 1 (heads only): 15 epochs, LR = 0.00015
    - Phase 2 (fine-tuning backbone + heads): 12 epochs, LR = 3e-5
    - Loss weights: Expression = 2.5, Valence = 1.0, Arousal = 1.0

1.2 ResNet50

- **Architecture:** ResNet50 backbone (pretrained on ImageNet)
- **Output heads:** Same as EfficientNetB0
- **Parameters:** ~23M
- **Training configuration:**
    - Phase 1 (heads only): 12 epochs, LR = 0.00015

- Phase 2 (fine-tuning backbone + heads): 12 epochs, LR = 2e-5
- Loss weights: Expression = 2.0, Valence = 1.0, Arousal = 1.0

---

2. Dataset Splits

- **Total samples:** 3,999
- **Training set:** 80% (~3,199 samples)
- **Validation set:** 20% (~800 samples)
- **Test set:** 800 samples

---

3. Training Graphs

(Summary of trends from logs)

- **EfficientNetB0:**
  - Expression accuracy plateaued ~0.15
  - Valence and arousal losses stable after LR reduction
- **ResNet50:**
  - Expression accuracy steadily improved to ~0.28
  - Valence and arousal losses steadily decreased, showing better convergence than EfficientNetB0

Graphs can be plotted using `matplotlib` for loss vs epochs and accuracy vs epochs.

---

4. Performance Measures

Model	Expression ACC	Expression F1	Valence RMSE	Valence Corr	Arousal RMSE	Arousal Corr
EfficientNetB0	0.1588	0.1507	0.5173	0.1283	0.4212	0.0303
ResNet50	0.2812	0.2635	0.4528	0.3652	0.3915	0.1790

Additional metrics:

- **EfficientNetB0:** Cohen's Kappa = 0.0386, Krippendorff's alpha = 0.0386, Expression AUC = 0.5421, PR-AUC = 0.1583
- **ResNet50:** Cohen's Kappa = 0.1786, Krippendorff's alpha = 0.1786, Expression AUC = 0.7262, PR-AUC = 0.3032
- **SAGR (Sign Agreement):**
    - EfficientNetB0: Valence = 0.5537, Arousal = 0.7462
    - ResNet50: Valence = 0.7125, Arousal = 0.7700
- **CCC (Concordance Correlation Coefficient):**
    - EfficientNetB0: Valence = 0.0758, Arousal = 0.0197
    - ResNet50: Valence = 0.3224, Arousal = 0.1395

---

5. Performance Comparison

Metric	EfficientNetB0	ResNet50
Expression Accuracy	0.1588	0.2812
Expression F1	0.1507	0.2635
Valence RMSE	0.5173	0.4528
Valence Correlation	0.1283	0.3652

Arousal RMSE	0.4212	0.3915
Arousal Correlation	0.0303	0.1790
Training Time per Epoch	~43s	~110s

Observations:

- ResNet50 outperforms EfficientNetB0 across all metrics.
- EfficientNetB0 is faster but less accurate.
- Multi-output training was stable in both models, with ResNet50 showing superior regression performance.

---

6. Summary

- **Best Expression Classification:** ResNet50 (Accuracy: 0.2812, F1: 0.2635)
- **Best Valence Prediction:** ResNet50 (Correlation: 0.3652)
- **Best Arousal Prediction:** ResNet50 (Correlation: 0.1790)
- ResNet50 is recommended for multi-task emotion recognition despite longer training time.
- EfficientNetB0 may be preferable for resource-constrained scenarios.