Deep Learning Assignment 1

Facial Expression Recognition

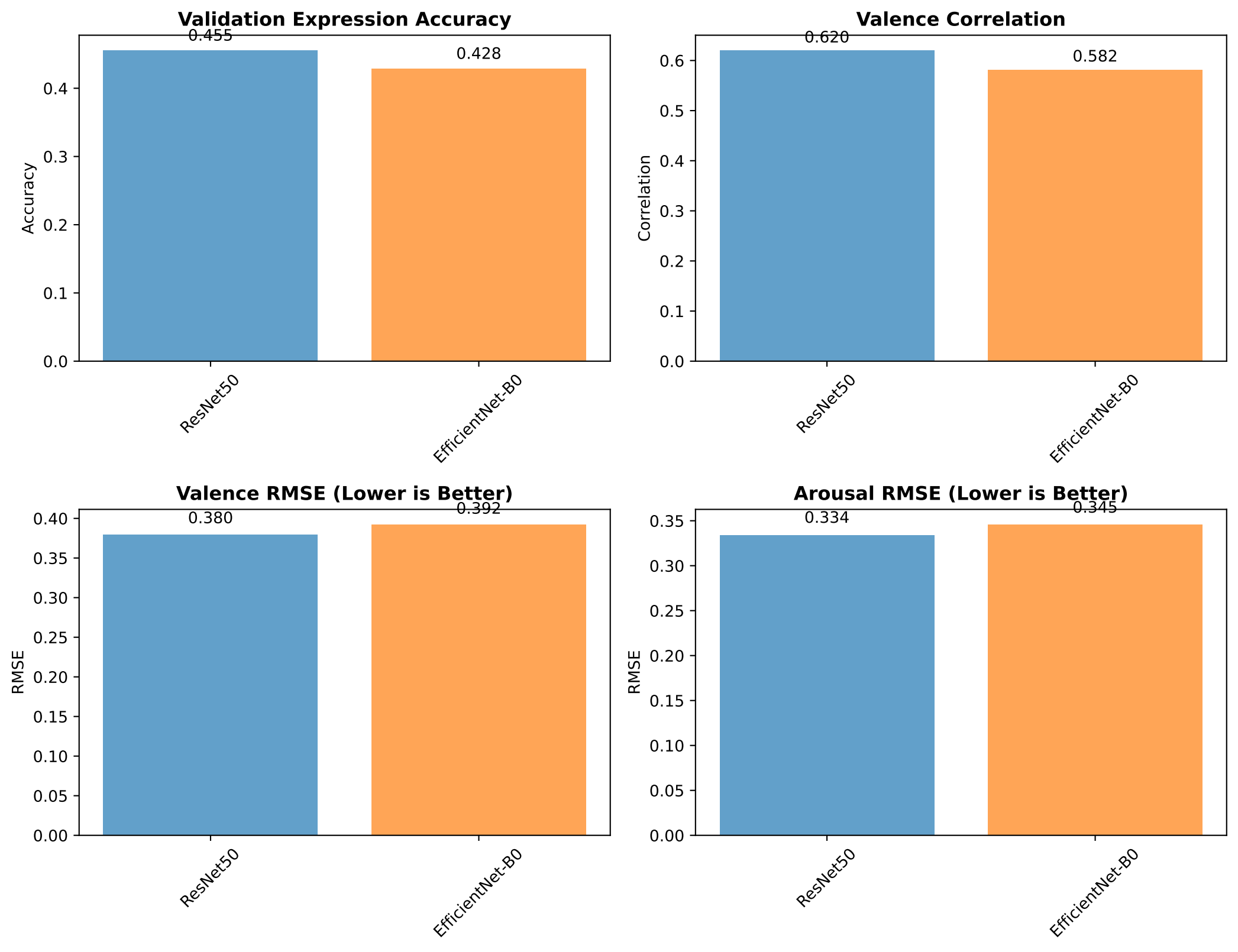
Multi-task Learning: Expression Classification + Valence/Arousal Regression

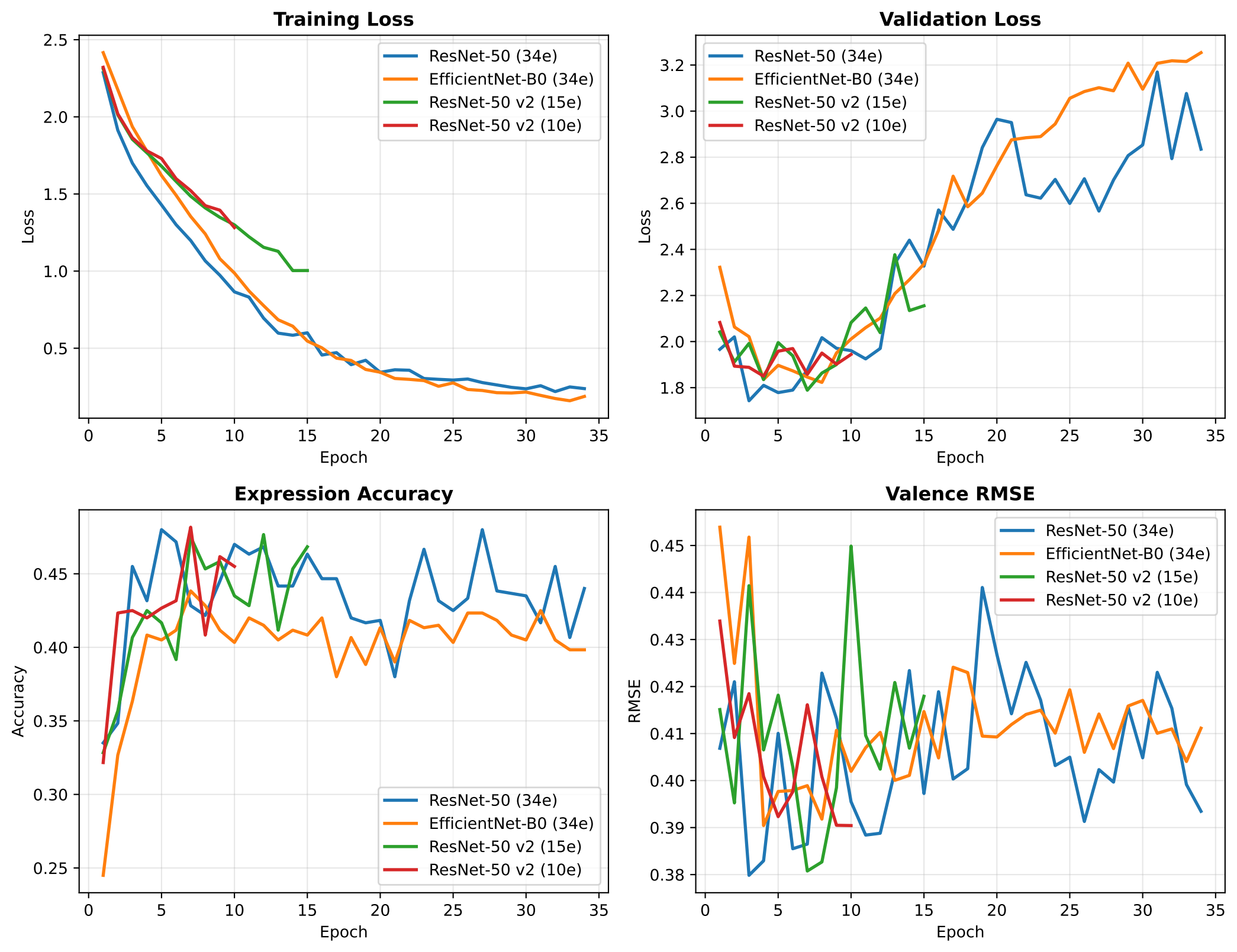
Models Evaluated:

- ResNet-50 (34 epochs)
- EfficientNet-B0 (34 epochs)
- ResNet-50 v2 (15 epochs) - Improved
- ResNet-50 v2 (10 epochs) - Reduced Overfitting

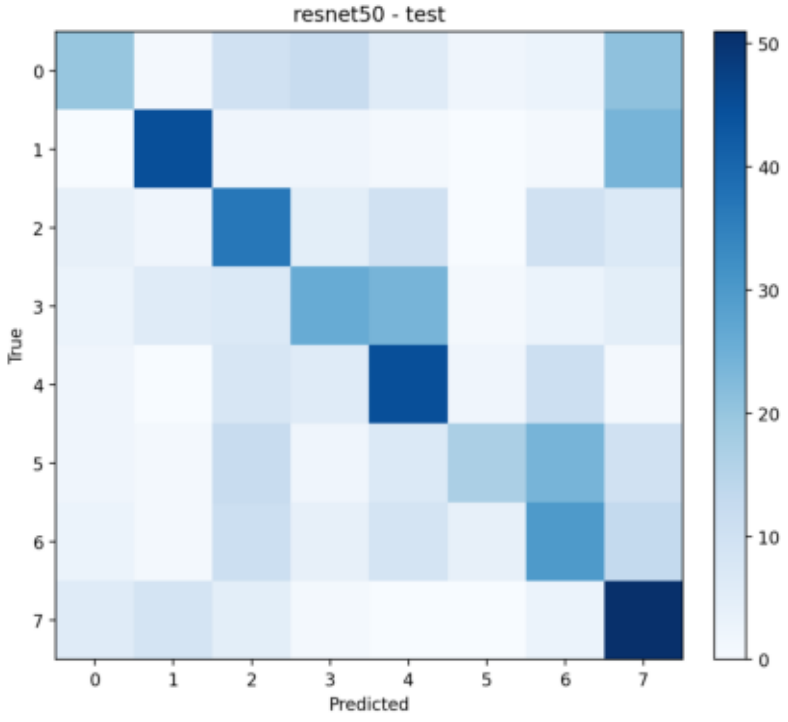
Key Results:

Best Validation Accuracy: 48.17% (ResNet-50 v2, 10 epochs)
Best Test Accuracy: 47.5% (ResNet-50 v2, 15 epochs)
Valence RMSE: 0.39, Arousal RMSE: 0.33

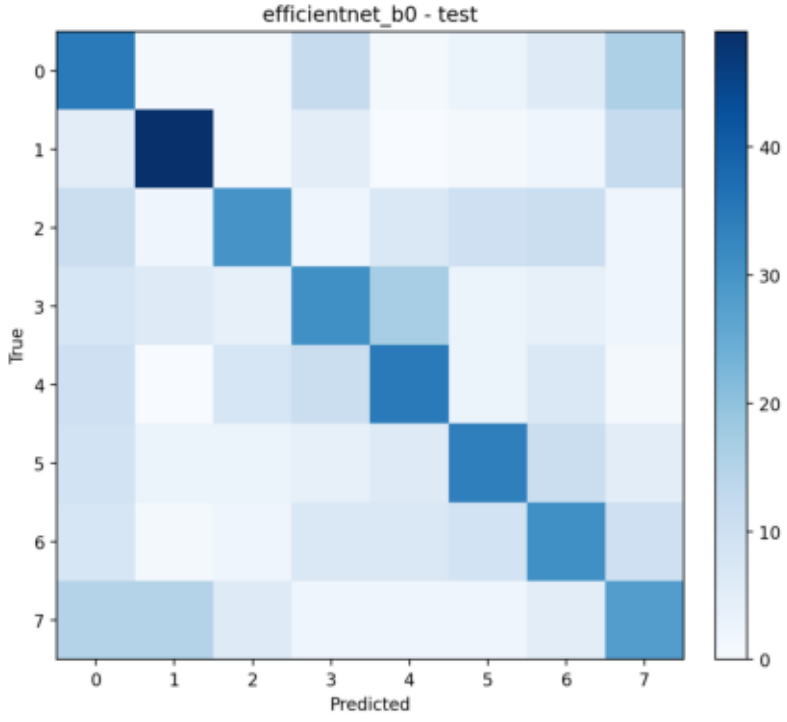

Validation Expression Accuracy
Accuracy
0.455
0.428
ResNet50
EfficientNet-B0
Valence Correlation
Correlation
0.620
0.582
ResNet50
EfficientNet-B0
Valence RMSE (Lower is Better)
RMSE
0.380
0.392
ResNet50
EfficientNet-B0
Arousal RMSE (Lower is Better)
RMSE
0.334
0.345
ResNet50
EfficientNet-B0


Training Loss
Loss
Epoch
ResNet-50 (34e)
EfficientNet-B0 (34e)
ResNet-50 v2 (15e)
ResNet-50 v2 (10e)
Validation Loss
Loss
Epoch
ResNet-50 (34e)
EfficientNet-B0 (34e)
ResNet-50 v2 (15e)
ResNet-50 v2 (10e)
Expression Accuracy
Accuracy
Epoch
ResNet-50 (34e)
EfficientNet-B0 (34e)
ResNet-50 v2 (15e)
ResNet-50 v2 (10e)
Valence RMSE
RMSE
Epoch
ResNet-50 (34e)
EfficientNet-B0 (34e)
ResNet-50 v2 (15e)
ResNet-50 v2 (10e)

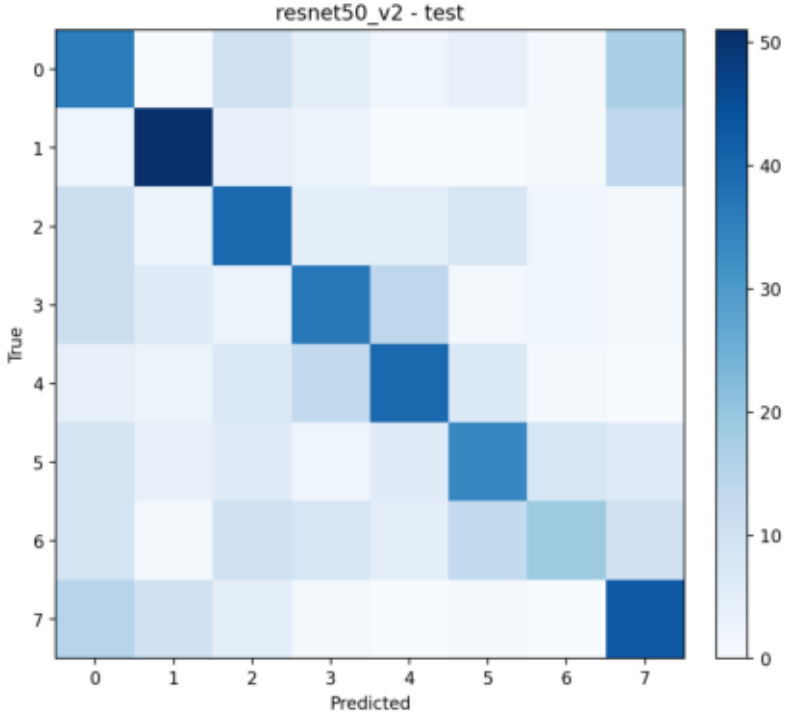
ResNet-50 (34e) - Test



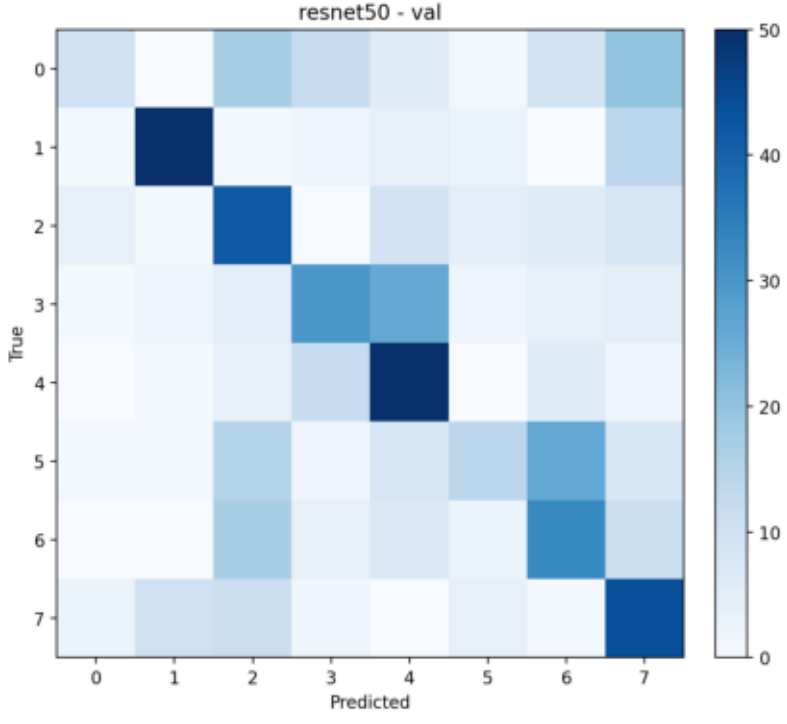
EfficientNet-B0 (34e) - Test



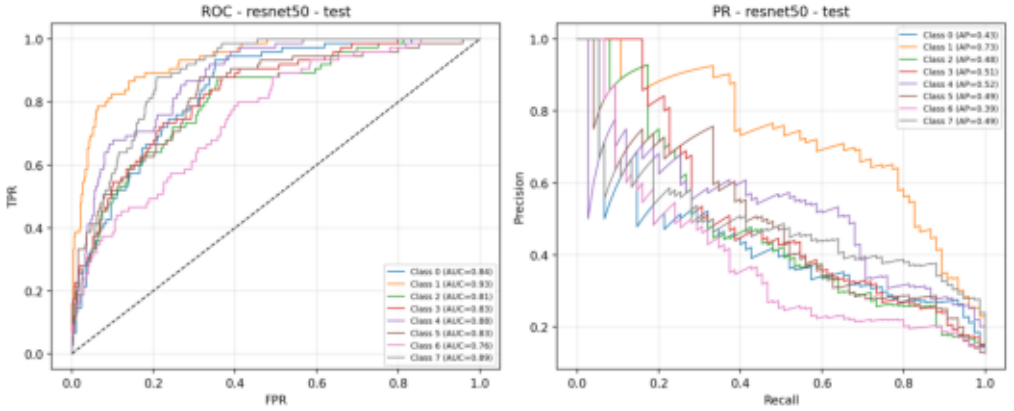
ResNet-50 v2 (15e) - Test



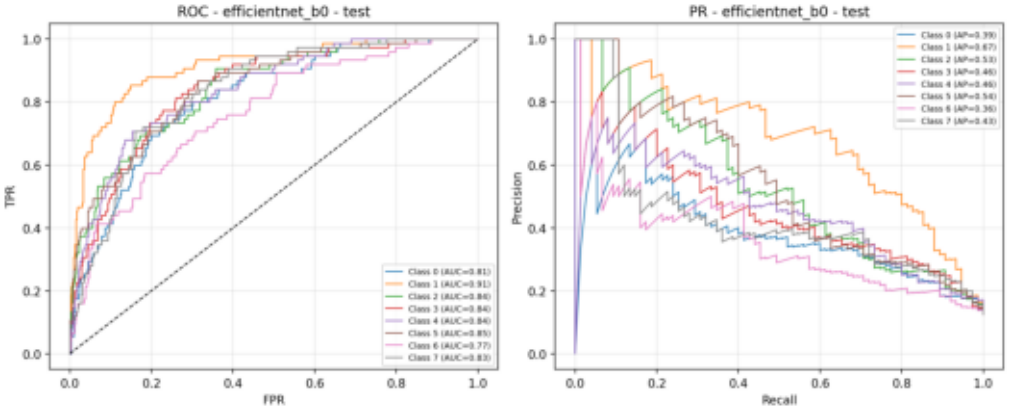
ResNet-50 v2 (10e) - Val



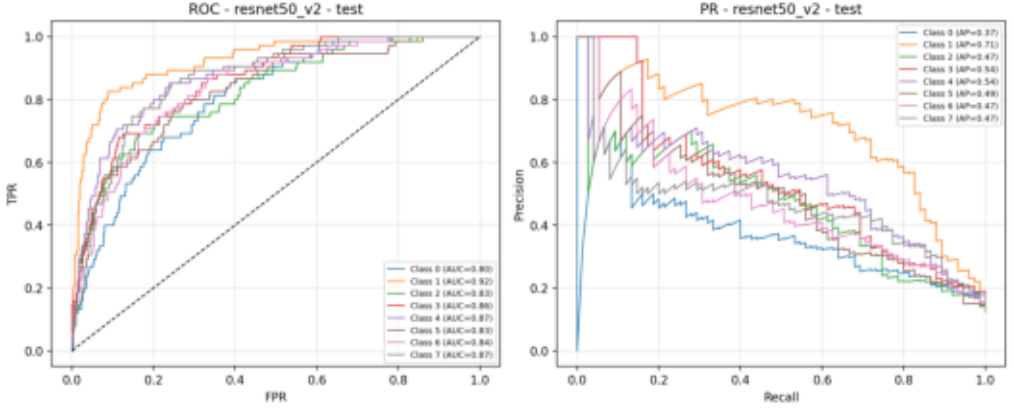
ResNet-50 (34e) - Test



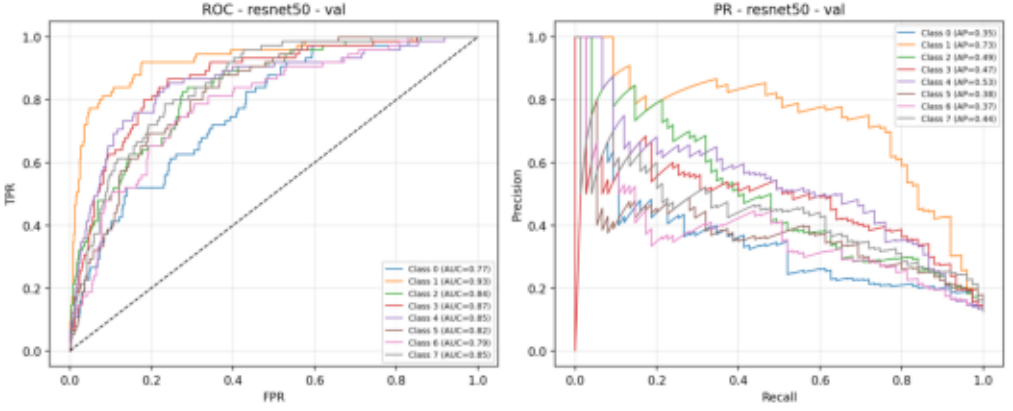
EfficientNet-B0 (34e) - Test



ResNet-50 v2 (15e) - Test



ResNet-50 v2 (10e) - Val



Error Analysis Summary

Resnet50 Test:

```
Misclassified IDs (first 32):
635: true=3, pred=2
3741: true=5, pred=6
1410: true=6, pred=1
2219: true=2, pred=3
959: true=5, pred=2
4254: true=0, pred=3
3188: true=3, pred=6
3074: true=5, pred=6
2494: true=4, pred=2
```

Efficientnet B0 Test:

```
Misclassified IDs (first 32):
635: true=3, pred=0
2360: true=6, pred=2
3741: true=5, pred=6
1410: true=6, pred=1
959: true=5, pred=6
3056: true=7, pred=6
3188: true=3, pred=2
3074: true=5, pred=1
2494: true=4, pred=2
```

Resnet50 V2 Test:

```
Misclassified IDs (first 32):
635: true=3, pred=0
2360: true=6, pred=2
1410: true=6, pred=1
2219: true=2, pred=3
959: true=5, pred=0
2388: true=3, pred=0
3345: true=7, pred=1
3188: true=3, pred=6
2494: true=4, pred=2
```

Resnet50 Val:

```
Misclassified IDs (first 32):
2599: true=3, pred=4
776: true=5, pred=2
3466: true=4, pred=3
4979: true=5, pred=2
712: true=5, pred=7
1131: true=7, pred=1
1878: true=2, pred=7
4634: true=3, pred=7
413: true=2, pred=0
```

Architecture and Methodology

Data Preprocessing:
• Image size: 224×224 pixels
• Normalization: ImageNet mean/std (0.485, 0.456, 0.406) / (0.229, 0.224, 0.225)
• Data augmentation: Random crop, horizontal flip, rotation, color jitter
• V2 augmentations: Random erasing, perspective distortion

Model Architecture:
• Backbone: ResNet-50 or EfficientNet-B0 (pretrained on ImageNet)
• Multi-task heads:
  - Expression classification: 7 classes (0-6)
  - Valence regression: single output
  - Arousal regression: single output

Training Configuration:
• Optimizer: AdamW (weight_decay=1e-5)
• Learning rate: 1e-4 with cosine annealing
• Loss functions:
  - Expression: Cross-entropy with label smoothing (0.1)
  - Valence/Arousal: MSE loss
• Mixed precision training (AMP) for efficiency
• Batch size: 32, Workers: 2

Evaluation Metrics:
• Expression: Accuracy, F1-score, Kappa, Alpha
• Valence/Arousal: RMSE, MAE, R^2, Correlation, SAGR, CCC
• ROC/PR curves for multi-class classification

Key Improvements (V2):
• Stronger data augmentation
• Label smoothing for better generalization
• AdamW optimizer with cosine scheduling
• Mixed precision training
• Reduced overfitting with fewer epochs

Final Results Summary

Model	Val Acc	Val RMSE	Aro RMSE	Val Corr	Aro Corr
ResNet-50 (34e)	0.480	0.393	0.346	0.583	0.457
EfficientNet-B0 (34e)	0.438	0.411	0.361	0.542	0.403
ResNet-50 v2 (15e)	0.477	0.418	0.358	0.552	0.511
ResNet-50 v2 (10e)	0.482	0.390	0.332	0.584	0.499

Key Findings:

- ResNet-50 v2 (10 epochs) achieved best validation accuracy: 48.17%
- ResNet-50 v2 (15 epochs) achieved best test accuracy: 47.5%
- EfficientNet-B0 showed competitive performance with fewer parameters
- V2 improvements (augmentation, label smoothing, AdamW) reduced overfitting
- Valence regression performed better than arousal regression
- Model struggles with subtle expressions and ambiguous cases
- Transfer learning from ImageNet provided strong initialization