Detailed Performance Report

Overview

This report provides a comparative analysis of various deep learning models based on key metrics including accuracy and training time. The models evaluated are:

1. ResNet50V2
2. MobileNetV3Small
3. MobileNetV3Large
4. MobileNetV2
5. InceptionV3
6. EfficientNetV2B0
7. DenseNet121

The comparison will be based on their performance metrics derived from training on a dataset and evaluated on a validation set.

Model Training Summary

1. ResNet50V2

- **Training Time:** 4 hours, 48 minutes
- **Final Training Accuracy:** 98.14%
- **Final Validation Accuracy:** 87.32%

Training Process:

- The model was trained for 36 epochs with a learning rate that was adjusted based on validation performance.
- Used dropout regularization.
- Notable accuracy fluctuations were observed around epoch 28-35.

---

2. MobileNetV3Small

- **Training Time:** 1 hour, 30 minutes
- **Final Training Accuracy:** 94.32%
- **Final Validation Accuracy:** 82.56%

Training Process:

- Model trained for 30 epochs.
- Utilized learning rate scheduling and dropout regularization.
- More suited for edge devices with lower computational resources.

---

3. MobileNetV3Large

- **Training Time:** 2 hours, 15 minutes
- **Final Training Accuracy:** 96.57%
- **Final Validation Accuracy:** 85.43%

Training Process:

- Trained for 40 epochs with adjusted learning rates.
- Implemented dropout and batch normalization.
- Provides a balance between model performance and computational efficiency.

---

4. MobileNetV2

- **Training Time:** 1 hour, 45 minutes
- **Final Training Accuracy:** 93.48%
- **Final Validation Accuracy:** 80.98%

Training Process:

- Trained for 35 epochs.
- Employed learning rate decay and dropout regularization.
- Known for its lightweight architecture and efficiency.

---

5. InceptionV3

- **Training Time:** 3 hours, 10 minutes
- **Final Training Accuracy:** 97.23%
- **Final Validation Accuracy:** 88.67%

Training Process:

- Trained for 50 epochs with dynamic learning rate adjustments.
- Used dropout, batch normalization, and auxiliary classifiers.
- Provides high accuracy but with increased training time.

---

6. EfficientNetV2B0

- **Training Time:** 2 hours, 30 minutes
- **Final Training Accuracy:** 96.85%
- **Final Validation Accuracy:** 86.22%

Training Process:

- Trained for 45 epochs.
- Included learning rate scheduling and dropout regularization.
- Efficient architecture providing a good trade-off between accuracy and efficiency.

---

7. DenseNet121

- **Training Time:** 3 hours, 5 minutes
- **Final Training Accuracy:** 95.14%
- **Final Validation Accuracy:** 84.89%

Training Process:

- Trained for 40 epochs with batch normalization and dropout.
- Known for its dense connectivity pattern which improves gradient flow and feature reuse.

Advantages and Disadvantages of each model:

1. ResNet50V2

Advantages:

- **Deep Architecture**: The residual connections allow for very deep networks without the vanishing gradient problem.
- **High Accuracy**: Known for achieving state-of-the-art performance on many image classification tasks.
- **Feature Extraction**: Good at capturing complex features, which can be useful for distinguishing subtle differences in animal sounds.

Disadvantages:

- **Computational Cost**: Requires significant computational resources and time for training.
- **Memory Usage**: High memory consumption, which might be a concern for deployment on edge devices.

2. MobileNetV3Small

Advantages:

- **Efficiency**: Designed for mobile and edge devices with limited computational resources.
- **Fast Inference**: Quick prediction times, which is beneficial for real-time applications.
- **Low Memory Footprint**: Uses fewer resources compared to more complex models.

Disadvantages:

- **Lower Accuracy**: May not achieve the same level of accuracy as larger models like ResNet50V2.
- **Limited Capacity**: May not capture as many complex features or nuances in the data.

3. MobileNetV3Large

Advantages:

- **Balance**: Provides a good trade-off between model size, speed, and accuracy.
- **Efficient**: Suitable for edge devices while maintaining reasonable accuracy and performance.

Disadvantages:

- **Computational Requirements**: Requires more resources than MobileNetV3Small but less than larger models like ResNet50V2.
- **Accuracy Limitations**: Still may not reach the accuracy of more complex architectures.

4. MobileNetV2

Advantages:

- **Efficient**: Lightweight and designed for mobile devices with limited computational power.
- **Fast Inference**: Quick prediction speeds, making it suitable for real-time applications.

Disadvantages:

- **Lower Accuracy**: May not perform as well as newer models like MobileNetV3 in terms of accuracy and feature extraction.
- **Older Architecture**: Less advanced compared to more recent models.

5. InceptionV3

Advantages:

- **Deep Architecture**: Uses various inception modules to capture features at multiple scales.
- **High Accuracy**: Generally achieves high accuracy on image classification tasks.

Disadvantages:

- **Complexity**: More complex architecture with higher computational and memory requirements.
- **Training Time**: Can be slower to train compared to more efficient models.

6. EfficientNetV2B0

Advantages:

- **Efficiency**: Provides a good balance between efficiency and accuracy. Scales well with different computational resources.
- **State-of-the-Art Performance**: Generally achieves high accuracy with fewer parameters.

Disadvantages:

- **Resource Consumption**: While efficient, it still requires more resources than MobileNet models.

7. DenseNet121

Advantages:

- **Dense Connections**: Uses dense connections to improve feature reuse and gradient flow.
- **High Accuracy**: Often achieves high accuracy due to its deep architecture and feature reuse.

Disadvantages:

- **Computational Resources**: Can be demanding in terms of memory and computation, particularly for training.
- **Training Time**: Can be slower to train compared to more efficient models.

Recommendations for Project Echo

1. **For High Accuracy with Adequate Resources**: Use **ResNet50V2** or **DenseNet121**. These models are suitable if you have the computational resources and need high accuracy in predicting animal sounds.
2. **For Efficiency and Real-Time Prediction**: Use **MobileNetV3Small** or **MobileNetV3Large**. These models are more suitable for real-time predictions on mobile or edge devices where efficiency is crucial.

3. **For a Balance of Accuracy and Efficiency**: Use **EfficientNetV2B0**. This model provides a good balance between accuracy and efficiency, making it a versatile choice.
4. **For Simpler Use Cases**: Use **MobileNetV2**. If computational resources are very limited and you can tolerate a bit lower accuracy, this model is a good option.