Project Report ITAI 1378

Object Detection Using Transfer Learning

Authors: Williane Yarro, Neida Guzman, Niki Vahdati

1. Objective

The goal of this project is to enhance the performance of a baseline object detection model by leveraging transfer learning techniques, optimizing data preparation strategies, and implementing innovative methodologies within computational constraints.

2. Dataset

2.1 Dataset Selection

• **Dataset Name**: [PASCAL VOC 2007 / COCO-minitrain / Custom Dataset]

• **Source**: [http://host.robots.ox.ac.uk/pascal/VOC/voc2007/]

• **Categories**: 20 categories: person, bird, cat, cow, dog, horse, sheep, airplane, bicycle, boat, bus, car, motorbike, train, bottle, chair, dining table, potted plant, sofa, tv/monitor.

• **Size**: Total number of images is 9,963 but only a subset of 10 to 20% was used. An estimate of 996 to 2000 images were used.

2.2 Dataset Preparation

1. **Cleaning**:
    a. Removed low-quality or irrelevant images to ensure data quality.
    b. Ensured consistent annotation format.
    c. Low quality or irrelevant images were removed.
2. **Normalization**:
    a. Rescaled pixel values to the range [0, 1].

3. **Augmentation**:
    a. Applied techniques such as rotation, flipping, and brightness adjustment.
    b. Increased the diversity of training samples.

3. Methodology

3.1 Pre-trained Models

Three pre-trained models were evaluated:

1. **VGG16**:
    a. Strength: Simplicity and effectiveness on small datasets.
    b. Limitation: High computational cost and risk of overfitting.
2. **ResNet50**:
    a. Strength: Robustness in extracting hierarchical features using skip connections.
    b. Limitation: Moderate computational demands.

    -This model was chosen as the base model due to the balance of accuracy and computational efficiency.

3. **MobileNet**:
    a. Strength: Lightweight and fast inference suitable for resource-constrained environments.
    b. Limitation: Lower accuracy on complex datasets.

3.2 Model Architecture

Base Model: **ResNet50** (Best performing).

• Custom Layers:
    o Global Average Pooling.
    o Fully connected dense layers.
    o Output layer with SoftMax activation for classification.

3.3 Training Strategies

1. **Fine-Tuning**:
    a. Experimented with freezing/unfreezing different layers of the pre-trained models.

2. **Data Augmentation**:
    a. Augmented data in real-time during training using ImageDataGenerator.

3. **Hyperparameter Optimization**:
    a. Explored learning rates of 0.01, 0.001, and 0.0001.

4. **Callbacks**:
    a. ModelCheckpoint: Saved the best model based on validation loss.
    b. EarlyStopping: Halted training when no improvement was observed.

4. Results

4.1 Evaluation Metrics

- **Accuracy**: Improved from **75% (baseline)** to **85%** with ResNet50.
- **Precision**: [9%]
- **Recall**: [23%]
- **F1-Score**: [75%]

4.2 Confusion Matrix

The confusion matrix highlights class-wise performance. Common errors occurred between visually similar classes.

Confusion Matrix Visualization:

4.3 Comparison of Models

Model	Accuracy	Inference Time	Memory Usage
VGG16	80%	High	High
ResNet50	85%	Moderate	Moderate
MobileNet	78%	Low	Low

4.3 Comparison of Models

Model Accuracy Inference Time Memory Usage

VGG16 80% High
ResNet50 85% Moderate
MobileNet 78% Low

5. Discussion

5.1 Insights

• **ResNet50** achieved the best balance between accuracy and computational cost.
• Fine-tuning deeper layers significantly improved accuracy.
• Class imbalance and limited training data led to errors in certain classes.

5.2 Challenges

1. Computational limits on free-tier Collab.
2. Class imbalance in the dataset.
3. Overfitting on small datasets.

- Limitations of computational resources, particularly with the use of the free-tier Colab, could have restricted the scope of experiments and model training. Potential data issues, such as inconsistencies in annotations or the presence of irrelevant images, could have impacted model performance. The risk of overfitting, especially given the use of a subset of the PASCAL VOC 2007 dataset, was acknowledged.

5.3 Risk Mitigation

• Used lightweight models (MobileNet) for initial experiments.
• Applied class-weighted training to address imbalance.
• Implemented augmentation to increase data diversity.

- To manage computational constraints, the team started with lightweight models like MobileNet for initial experiments and carefully considered the computational cost when selecting the base model. A thorough data cleaning process was implemented to ensure data quality, removing irrelevant or low-quality images and ensuring consistent annotation formats. Data augmentation techniques were employed to address potential overfitting by increasing data diversity and providing the model with a wider range of training examples

6. Future Work

1. **Ensemble Learning**:
   a. Combine predictions from multiple models for higher accuracy. 2.
Advanced Augmentation:
   a. Explore GAN-based synthetic data generation.
3. **Model Optimization**:

    a. Implement quantization to improve inference speed.
4. **Larger Datasets**:
    a. Incorporate additional annotated datasets for better generalization.

7. Reflection

Team Member Learnings

- **Willaine Yarro**: Gained insights into fine-tuning transfer learning models.
- **Neida Guzman**: Learned about data preprocessing and augmentation strategies.
- **Niki Vahdati**: Explored techniques for hyperparameter optimization.

8. Conclusion

This project demonstrated the effectiveness of transfer learning in improving object detection performance. ResNet50 emerged as the most suitable model, achieving a significant improvement in accuracy. Future work will focus on further optimizing the model and exploring ensemble methods.

9. References

1. PASCAL VOC Dataset
2. TensorFlow Documentation
3. Towards Data Science: Transfer Learning

Deliverables

- **Code**: Available in the src/ directory and Jupyter notebooks in notebooks.
  ITIA_1378_Midterm_2024_Group3.ipynb

- **Reports**: Located in the reports/ directory.
- **Presentation**: Slides summarizing the project are in presentation.