Enhanced DETR for Yellow Leaf Disease Detection

Xiaoning Li
lixiaoni@student.chalmers.se

Yaochen Song
yaochen@student.chalmers.se

***Abstract*—This project explores the improvemnet of the standard DEtection TRansformer (DETR) model for enhanced yellow leaf disease detection by replacing its ResNet50 backbone with EfficientNet and MobileNet respectively. These modifications are evaluated against the original DETR model using metrics like Mean Average Precision (MAP) and the Precision-Recall curve. Preliminary findings indicate that the modified DETR models significantly improve detection performance, showcasing their potential in advancing crop disease detection efforts.**

I. Introduction

Crop diseases, particularly yellow leaf disease, pose significant threats to agricultural productivity. Efficient and accurate detection of such diseases is paramount for timely intervention and management. Recent advancements in machine learning, especially MobileNet and EfficientNet, have shown remarkable capabilities in image-based disease detection with numerous studies illustrating their extensive application in the domain of crop disease identification. However, the standard DEtection TRansformer (DETR) model, equipped with the ResNet50 network architecture, exhibits weaker performance in detecting crop diseases, highlighting a notable research gap.

This gap prompts an exploration into enhancing the DETR model's capability by leveraging the strengths of EfficientNet and MobileNet. In this endeavor, this project introduces two modified versions of the DETR model: one where EfficientNet replaces ResNet50, and the other where MobileNet substitutes for ResNet50. These newly introduced models are deployed to detect yellow leaf disease, with a comparative analysis conducted against the original DETR model to evaluate their detection performance. Through metrics such as Mean Average Precision (MAP) and the Precision-Recall curve, this project aims to assess the efficacy of the modified DETR models in addressing crop disease detection, thereby contributing to the broader initiative of employing advanced machine learning architectures for agricultural disease detection.

II. Background Theory

Historically, the focus was on evaluating the individual performance of the DETR model, EfficientNet, and MobileNet for crop disease detection. Recent studies have investigated the capabilities of these architectures in the agricultural context.

Suherman et al. (2023) demonstrated the effectiveness of ResNet-50 in the DETR model for image recognition tasks [1]. Borhani et al. (2022) used EfficientNet-L2 with the ViT structure for real-time crop disease classification [2]. Kumar et al. (2023) employed an ensemble including EfficientNet and ViT for cassava leaf disease detection [3]. Fu et al. (2023) proposed a transformer-based model using ViT for crop pest recognition [4]. Rajeena et al. (2023) explored EfficientNet's capability for corn leaf disease detection [5]. Kotwal et al. (2023) presented an EfficientNet approach for plant leaf disease classification [6].

Recent trends focus on integrating these models. For instance, DETR integrated with ViT is known as DETR (ViT) [7]. Another study showed that replacing ResNet-50 with EfficientNet-B3 in EfficientDet improved accuracy by 3% and reduced computation by 20% [8]. This illustrates the combined potential of various network architectures in enhancing disease detection.

III. Method and Implications

A. Dataset

This project analyzes small data sets. We use the Yellow Leaf Disease Dataset :COCO standard data set format; 378 train images; 98 val images; 28 test images.

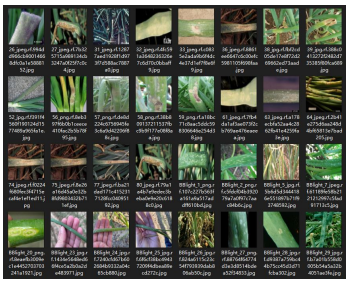

Fig. 1: Part of the images in the training set

This project analyzes small datasets using the Yellow Leaf Disease Dataset, which adheres to the COCO standard dataset format. It comprises 378 training images, 98 validation images, and 28 test images. We employ this dataset to compare the advantages, disadvantages, and unique features of three different network architectures in image detection, particularly focusing on the detection of crop diseases such as yellow leaf disease.

B. Backbone

1) ResNet-50

In the original DEtection TRansformer (DETR) model, ResNet-50 are used as the backbone network architecture for the convolutional layers. The backbone is responsible for extracting feature maps from the input images, which are then processed by the transformer layers in DETR for object detection tasks. [9]

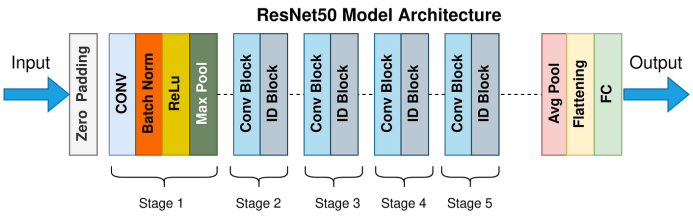

Fig. 2: ResNet-50 Network Architecture

2) MobileNet

MobileNet, on the other hand, is designed for mobile and embedded vision applications. It utilizes depth-wise separable convolutions, breaking down the traditional convolution into a depth-wise convolution followed by a point-wise convolution, reducing computational cost and making it lightweight. [10]

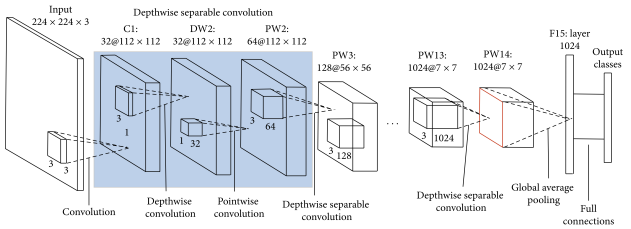

Fig. 3: MobileNet Network Architecture

3) EfficientNet

EfficientNet scales the network width, depth, and resolution with a set of fixed scaling coefficients, optimizing for both accuracy and efficiency. It introduces a compound scaling method that uses a mix of depthwise separable convolutions and traditional convolutions, resulting in models that perform comparably or even better than other architectures while being computationally efficient.

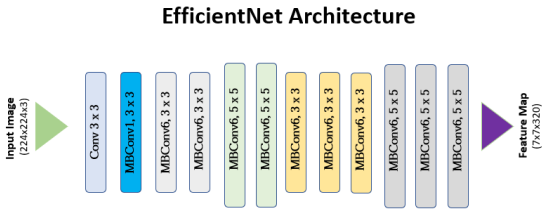

Fig. 4: EfficiencyNet Network Architecture

C. Hardware and platform

- OS: Windows
- Graphics card: GTX1080 Ti
- Training memory: 16G
- Video memory: 11G

D. Performance Metrics

1) Precision Recall

When making classification predictions, 4 outcomes are possible:

- **True positives (TP)**: Prediction → Sick ——*Actually* → Sick.
- **True negatives (TN)**: Prediction → Healthy ——*Actually* → Healthy.
- **False positives (FP)**: Prediction → Sick ——*Actually* → Healthy.
- **False negatives (FN)**: Prediction → Healthy ——*Actually* → Sick.

It can be concluded as *Confusion Matrix*.

Predict \ Actually	Actually 1	Actually 0
Predict 1	TP (True Positive)	FP (False Positive)—Type I Error
Predict 0	FN (False Negative)—Type II Error	TN (True Negative)

Significant indicators: **Precision**, **Recall** and **Score** (result of harmonic mean of recall and precision) are:

$$\text{Precision} = \frac{TP}{TP + FP} \qquad \text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

Based on the above formula, we plotted the Precision/Recall Scores/Recall images of different backbones.These curves show how recall changes as the confidence score threshold and precision threshold change.

2) AP

AP is the area under the Precision-Recall curve. It provides a single value that summarizes the Precision-Recall curve. In other words, it averages the precision values at different recall levels.

In addition to the basic AP curve, we also generate other AP related curves

- AP Small AP Medium AP Large : Performance on Small Medium and large objects.
- AP 50 AP 75: These metrics evaluates Average Precision at an Intersection over Union (IoU) threshold of 0.50 and 0.75.It's a stricter metric compared to AP 50, and a high score here indicates that the model is not only detecting objects correctly but also providing highly accurate bounding box predictions.

IV. Results and Discussion

In this project, we analyze the three backbones by comparing accuracy and resource usage.

A. Resource occupation

epoch	ResNet-50	MobileNet	EfficientNet
50epochs	48min	38min	59min
100epochs	1h4min	1h17min	2h16min

TABLE I: Training time

Fig. 5: Resource occupation

B. Accuracy

1) Precision/Score Recall

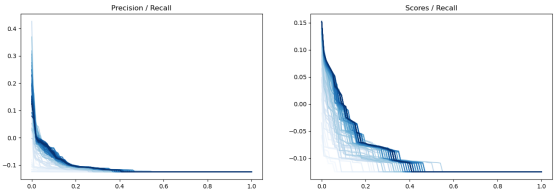

Fig. 6: EfficiencyNet

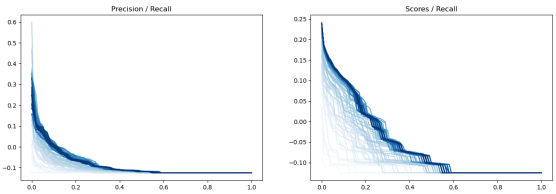

Fig. 7: MobileNet

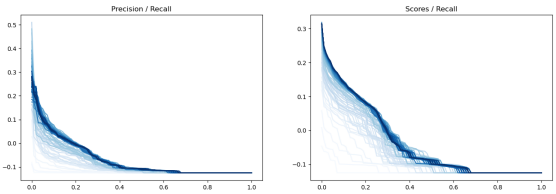

Fig. 8: ResNet-50

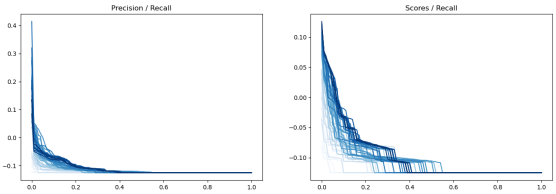

Fig. 9: EfficiencyNet

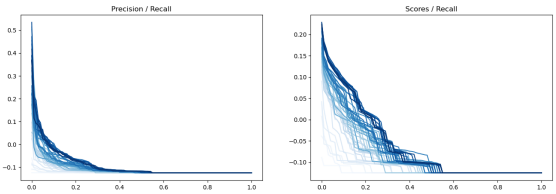

Fig. 10: MobileNet

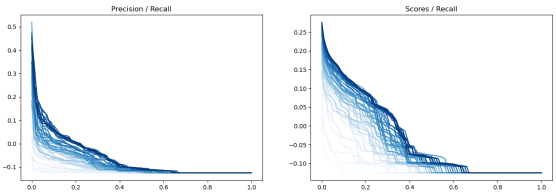

Fig. 11: ResNet-50

2) AP

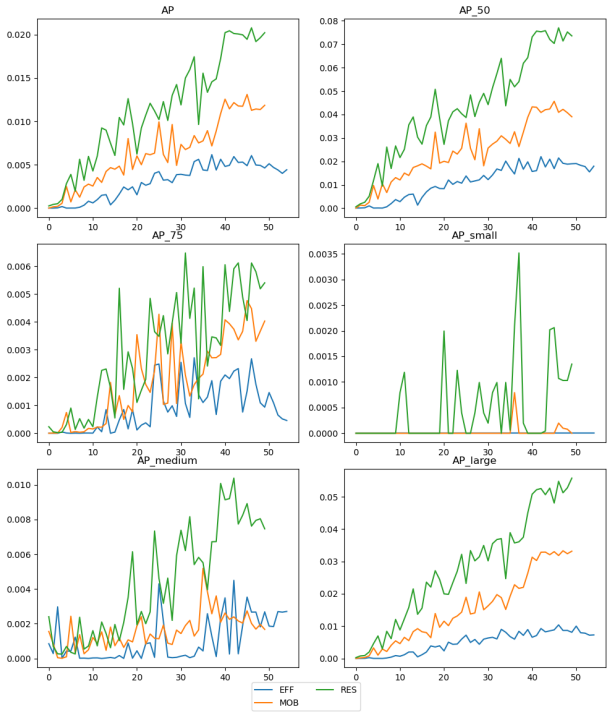

Fig. 12: AP 50 epochs

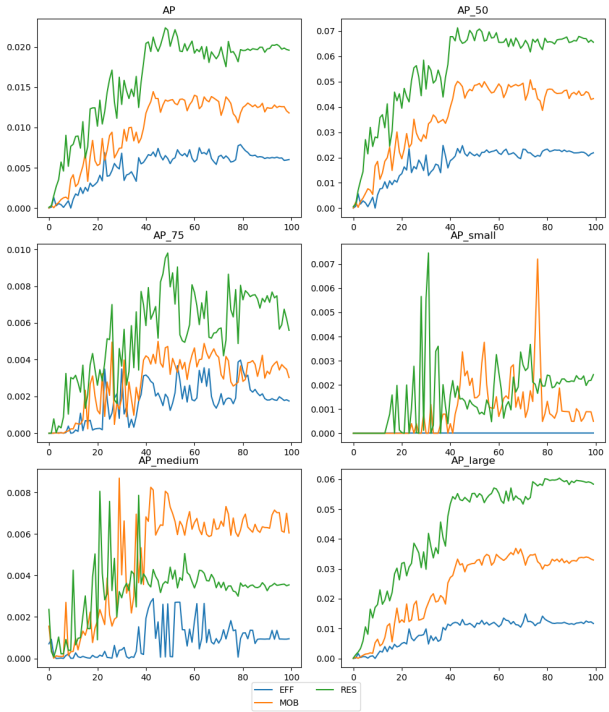

Fig. 13: AP 100 epochs

V. Discussion

A. Precision/Score and Precision/Recall analyze

From the figure in Precision/Score and Precision/Recall both in 50 epochs and 100 epochs, we find that ResNet-50 perform best.Because Precision/Recall and Precision/Score curves shows hightest in ResNet/50 model.The right chart, representing scores vs. recall, shows overlapping curves that peak early and then descend, suggesting variability in the scores associated with the predicted bounding boxes.

B. AP analyze

Resnet-50 (Green Curve): - Demonstrates a consistently increasing trend in the AP and AP 50 metrics in both epochs, indicating stable performance improvement over time. - For size-specific AP metrics (AP small, AP medium, and AP large), performance fluctuates, especially in AP small and AP large, showing some volatility in recognizing different sized objects.

Efficientnet (Blue Curve): - Shows an upward trend in the AP and AP 50 metrics, but the growth rate is slower compared to Resnet-50, especially in the 100-epoch chart. - In size-specific AP metrics, it exhibits more stability than Mobilenet but has greater fluctuations than Resnet-50, especially in AP small and AP medium in the 100-epoch chart.

Mobilenet (Orange Curve): - Performance in AP and AP 50 metrics seems to plateau and even decline after a certain point, especially evident in the 100-epoch chart. - Displays significant fluctuations in size-specific AP metrics across both epochs, with pronounced spikes and dips, indicating inconsistent performance in detecting differently sized objects.

3. Epoch-to-Epoch Performance:

- From 50 epochs to 100 epochs, most curves exhibit more fluctuations, indicating increased challenges or potential overfitting as training progresses. - The Mobilenet curve in particular displays more pronounced fluctuations in the 100-epoch chart, suggesting that its performance might be degrading or becoming more unstable over time.

C. Resource Analyze

Resnet50 and Efficientnet have similar GPU power usage. Meanwhile, Mobilenet has the lowest GPU power usage. Resnet50 provides the best performance in GPU allocation, whereas Efficientnet consumes a considerable amount of memory. Both Efficientnet and Resnet50 have similar memory access time, and Mobilenet outperforms them.

VI. Conclusions

Through the above experimental data, we can evaluate the three backbones.

1) ResNet-50

- **Pros**:
  - Established architecture with well-understood behavior.
  - High accuracy rates in various crop disease detection tasks.
  - Extensive community support and pre-trained models available.
- **Cons**:
  - Relatively high computational resources and memory requirements.
  - Might be overkill for simpler tasks or smaller datasets.

2) MobileNet

- **Pros**:
  - Highly efficient with fewer parameters, suitable for mobile or edge computing.
  - Faster inference times compared to ResNet-50 and EfficientNet.
  - Variants like MobileNetV2 and MobileNetV3 offer improved performance.
- **Cons**:
  - May sacrifice some accuracy for speed and size.
  - Might not perform as well as other architectures on complex tasks or very large datasets.

3) EfficientNet

- **Pros**:
  - Designed for efficiency with a balance between accuracy and computational resources.
  - Scalable architecture allowing for performance tuning.
- **Cons**:
  - Newer architecture with potentially less community support.
  - Might require more fine-tuning for specific tasks.

VII. Future Work

- Further Optimization: Further fine-tuning and optimization of the Mobile and Efficient backbones might lead to improvements in accuracy and efficiency.
- Expanding the Dataset: Incorporating more diverse and extensive datasets can lead to improved generalizability of the model.

References

[1] E. Suherman, B. Rahman, D. Hindarto, and H. Santoso, "Implementation of resnet-50 on end-to-end object detection (detr) on objects," *Sinkron: jurnal dan penelitian teknik informatika*, vol. 8, no. 2, pp. 1085–1096, 2023.

[2] Y. Borhani, J. Khoramdel, and E. Najafi, "A deep learning based approach for automated plant disease classification using vision transformer," *Scientific Reports*, vol. 12, no. 1, p. 11554, 2022.

[3] H. Kumar, S. Velu, A. Lokesh, K. Suman, and S. Chebrolu, "Cassava leaf disease detection using ensembling of efficientnet, seresnext, vit, deit and mobilenetv3 models," in *Proceedings of the International Conference on Paradigms of Computing, Communication and Data Sciences: PCCDS 2022*. Springer, 2023, pp. 183–193.

[4] X. Fu, Q. Ma, F. Yang, C. Zhang, X. Zhao, F. Chang, and L. Han, "Crop pest image recognition based on the improved vit method," *Information Processing in Agriculture*, 2023.

[5] F. Rajeena PP, A. SU, M. A. Moustafa, and M. A. Ali, "Detecting plant disease in corn leaf using efficientnet architecture—an analytical approach," *Electronics*, vol. 12, no. 8, p. 1938, 2023.

[6] J. G. Kotwal, R. Kashyap, and P. M. Shafi, "Artificial driving based efficientnet for automatic plant leaf disease classification," *Multimedia Tools and Applications*, pp. 1–32, 2023.

[7] H. Song, D. Sun, S. Chun, V. Jampani, D. Han, B. Heo, W. Kim, and M.-H. Yang, "Vidt: An efficient and effective fully transformer-based object detector," *arXiv preprint arXiv:2110.03921*, 2021.
[8] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10 781–10 790.
[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition."
[10] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications."