DeLO- An Ensemble of Densenet121 and YoloV11

Ritvik Sharma [a], Tanishq Sasmal [b],Rahil Agarwal [c]

[a] Author 1 affiliation: M.Tech student, NITM
[b] Author 2 affiliation: M.Tech student, NITM
[c] Author 3 affiliation: M.Tech student, NITM

Email: t24cs001@nitm.ac.in

Abstract

In this project, we present a novel approach to image classification by employing a hybrid ensemble model that integrates YOLOv11, a state-of-the-art object detection framework, and DenseNet121, a well-established convolutional neural network. Utilizing a meticulously augmented dataset categorized by class, we aim to enhance the robustness and accuracy of predictions. The dataset has been augmented to improve model generalization, allowing the ensemble to effectively learn the unique characteristics of each class. To achieve this, we implement a soft voting mechanism that aggregates predictions from both models, leveraging their individual strengths to yield superior classification performance. This study demonstrates the effectiveness of combining newer architectures with proven models to tackle complex classification tasks, ultimately contributing to advancements in the field of computer vision.

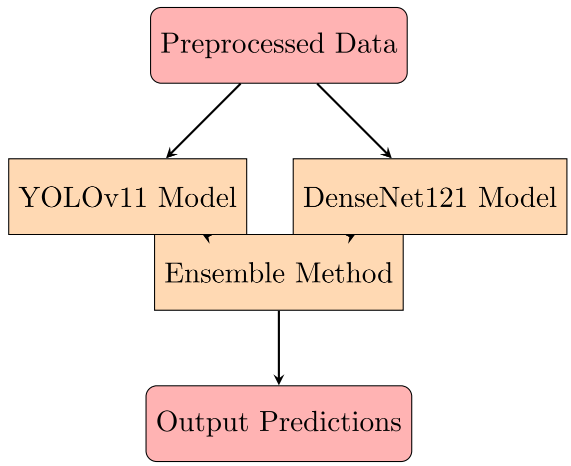
1 Introduction

Capsule endoscopy is a revolutionary imaging technique used to visualize the gastrointestinal (GI) tract, providing an invaluable tool for the diagnosis and management of various GI disorders. This non-invasive procedure allows for a comprehensive examination of the small intestine, where conventional endoscopic techniques may be limited or ineffective.

Despite its advantages, capsule endoscopy presents significant challenges in image analysis due to the vast amount of data generated and the subtlety of pathological findings. Manual interpretation of these images is not only time-consuming but also prone to human error. To address these challenges, automated image classification techniques have emerged as a promising solution to assist clinicians in identifying abnormalities such as bleeding, tumors, and other lesions.

Recent advancements in deep learning and computer vision have paved the way for more accurate and efficient image classification methods in the context of capsule endoscopy. By leveraging convolutional neural networks (CNNs) and ensemble models, researchers aim to enhance diagnostic accuracy, reduce interpretation time, and ultimately improve patient outcomes. This project focuses on utilizing a hybrid ensemble approach that combines the strengths of YOLOv11 and DenseNet121 to improve the classification of endoscopic images, thereby contributing to the growing body of research aimed at optimizing capsule endoscopy practices.

2 Methods



3 Results

3.1 Achieved results on the validation dataset

To create a validation dataset that better represents a uniform class distribution, we merged the initial training and validation datasets and performed a new split based on class balancing. This approach ensures that each class is more evenly represented in the validation dataset, improving the reliability of our performance metrics across all categories. However, as a result of this merging and re-splitting process, the original validation set was partially included in the training data, rendering testing on the original validation set ineffective, as it now contains data used during training.

Table 1: Class-wise Accuracies

Class Name	**Accuracy**
Angioectasia	0.9933
Bleeding	0.9933
Erosion	0.9356
Erythema	0.9867
Foreign Body	0.9089
Lymphangiectasia	0.9667
Normal	0.9756
Polyp	0.9556
Ulcer	0.9956
Worms	1.0000

4 Discussion

Our approach began with a rigorous data augmentation process, inspired by methods outlined in [1]. We experimented with various augmentation combinations to enhance model generalization and improve class representation. After iterative testing, we identified specific augmentations tailored to each class, which provided a balanced yet diverse dataset that would facilitate robust learning.

With this augmented dataset, we trained a YOLOv11[2] model for image classification. This model achieved an accuracy of 84% on our validation set, which had been designed to include data unseen during training. Although promising, we sought to further improve accuracy. Initial attempts using model soups to enhance performance, however, did not yield better results in this case.

Subsequent research into architectures particularly effective for classifying our specific set of classes led us to select DenseNet121[3]. This architecture, known for its feature reuse capability and efficient parameter utilization, proved to be lightweight enough not to hinder inference speed significantly, aligning well with our project's objectives.

Finally, we implemented a soft voting ensemble of YOLOv11 and DenseNet121, both of which were fine-tuned for optimal performance. The scripts for our fine-tuning processes are available on our GitHub page. This ensemble strategy leveraged the strengths of each model, ultimately yielding the class-wise accuracies presented in the previous table, reflecting the efficacy of our approach in classifying complex, varied gastrointestinal images.

5 Conclusion

In this study, we developed a robust classification pipeline for capsule endoscopy images, achieving a significant improvement in performance through careful data preparation and model selection. By augmenting our dataset with targeted transformations, we ensured better class representation, enabling the models to generalize effectively. Our initial model, YOLOv11, demonstrated promising results with an accuracy of 84% on a balanced validation set. Despite experimenting with model soups to further enhance accuracy, we did not observe substantial improvements, prompting us to explore alternative architectures.

After analyzing model performances specific to our dataset, we integrated DenseNet121 due to its efficiency and suitability for our task. By ensembling YOLOv11 and DenseNet121 using a soft voting mechanism, we harnessed the strengths of both models while maintaining reasonable inference speeds. This final ensemble model outperformed individual models, achieving high accuracy across multiple classes, as shown in our results. Our work highlights the effectiveness of ensemble approaches in medical image classification, which can be explored further with different combinations.

6 Acknowledgments

As participants in the Capsule Vision 2024 Challenge, we fully comply with the competition's rules as outlined in [4]. Our AI model development is based exclusively on the datasets provided in the official release in [5].

References

[1] Md Ahamed, Md Nahiduzzaman, Md Islam, Mansura Naznine, Mohamed Ayari, Amith Khandakar, and Julfikar Haider. Detection of various gastrointestinal tract diseases through a deep learning method with ensemble elm and explainable ai. *Expert Systems with Applications*, 256:124908, 01 2024. doi: 10.1016/j.eswa.2024.124908.

[2] Rahima Khanam and Muhammad Hussain. Yolov11: An overview of the key architectural enhancements, 2024. URL https://arxiv.org/abs/2410.17725.

[3] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. *arXiv preprint arXiv:1608.06993*, 2016. URL https://doi.org/10.48550/arXiv.1608.06993.

[4] Palak Handa, Amirreza Mahbod, Florian Schwarzhans, Ramona Woitek, Nidhi Goel, Deepti Chhabra, Shreshtha Jha, Manas Dhir, Deepak Gunjan, Jagadeesh Kakarla, et al. Capsule vision 2024 challenge: Multi-class abnormality classification for video capsule endoscopy. *arXiv preprint arXiv:2408.04940*, 2024.

[5] Palak Handa, Amirreza Mahbod, Florian Schwarzhans, Ramona Woitek, Nidhi Goel, Deepti Chhabra, Shreshtha Jha, Manas Dhir, Deepak Gunjan, Jagadeesh Kakarla, and Balasubramanian Raman. Training and Validation Dataset of Capsule Vision 2024 Challenge. *Figshare*, 7 2024. doi: 10.6084/m9.figshare.26403469.v1. URL https://figshare.com/articles/dataset/Training_and_Validation_Dataset_of_Capsule_Visio