Automated Classification of Video Capsule Endoscopy Abnormalities Using DINOv2

Avala Sri Karthik [a]

[a] Vellore Institute of Technology, Chennai

Email: srikarthikavala@gmail.com.com

Abstract

This project develops a deep learning model utilizing the DINOv2 architecture to automatically classify abnormalities detected in video capsule endoscopy (VCE) frames. The focus is on ten specific abnormalities, assisting healthcare professionals in diagnosing gastrointestinal disorders efficiently.

Keywords - Automated Classification, Video Capsule Endoscopy, Abnormalities Detection, Deep Learning, Medical Imaging, Gastrointestinal Disorders, Machine Learning, DINOv2 Model, Feature Extraction, Multi-class Classification

1 Introduction

The aim of this project is to create an automated classification system for video capsule endoscopy abnormalities, addressing the increasing need for efficient diagnostic tools in gastrointestinal healthcare.

2 Methods

The model is implemented using the DINOv2 architecture, designed for efficient feature extraction from the VCE frames.

2.1 Preprocessing Steps

1. **Image Resizing**: All images are resized to 280x280 pixels.
2. **Data Augmentation**: Techniques such as random rotations, flips, and brightness adjustments are applied.

2.2 Model Architecture

The model architecture combines vision transformers for feature extraction and fully connected layers for multi-class classification.

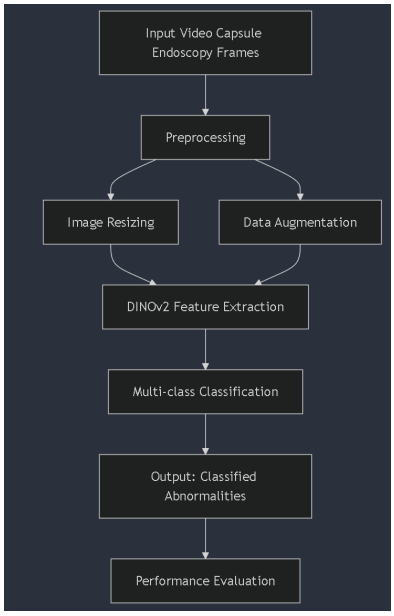

Figure 1: Block diagram of the developed pipeline.

```
class DinoVisionTransformerClassifier(nn.Module):
    def __init__(self):
        super(DinoVisionTransformerClassifier, self).__init__()
        self.transformer = dinov2_vits14
        self.classifier = nn.Sequential(
            nn.Linear(384, 256),
            nn.ReLU(),
            nn.Linear(256, 10)  # Assuming 10 classes
        )

    def forward(self, x):
        x = self.transformer(x)
        x = self.transformer.norm(x)
        x = self.classifier(x)
        return x
```

3 Results

3.1 Achieved results on the validation dataset

Results are logged during the training process, including metrics such as accuracy, loss, and area under the curve (AUC). The following table summarizes the comparison of our model against baseline models.

Table 1: Validation results and comparison to the baseline methods reported by the organizing team of Capsule Vision 2024 challenge.

Method	Avg. AUC	Avg. Specificity	Avg. Sensitivity	Avg. F1-score	Avg. Precision
DINOv2 Model	0.9914	0.9847	0.7849	0.8148	0.6876

4 Discussion

The performance of the DINOv2 model indicates its efficacy in classifying the specified abnormalities in video capsule endoscopy frames. Future work could focus on further refining the model and exploring additional datasets.

5 Conclusion

The automated classification system developed in this project shows promising results, demonstrating the potential of deep learning models in aiding gastrointestinal diagnostics.

6 Acknowledgments

As participants in the Capsule Vision 2024 Challenge, we fully comply with the competition's rules as outlined in [1]. Our AI model development is based exclusively on the datasets provided in the official release in [2].

References

[1] 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.

[2] 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. *Fishare*, 7 2024. doi: 10.6084/m9.figshare.26403469.v1. URL `https://figshare.com/articles/dataset/Training_and_Validation_Dataset_of_Capsule_Vision_2024_Challenge/26403469`.