UCC AI Quest 2023

Infection Team

Outline

1. Introduction (About the team)
2. Problem Investigation (Describe your understanding of the problem)
3. Challenges
4. Solution (Architecture / AI Model / Validation / Testing)
5. Discussions & Insights (What is good about your solution)
6. Conclusion

Introduction - Our Team

Minh-Khoi Pham

PhD student in DCU

Van-Tu Ninh

PhD student in DCU

Nam Trinh

PhD student in DCU

Tu-Khiem Le

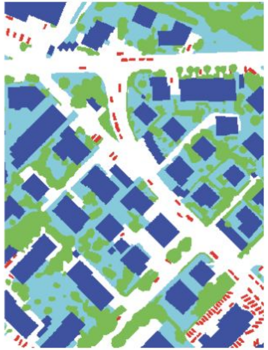
PhD student in DCU

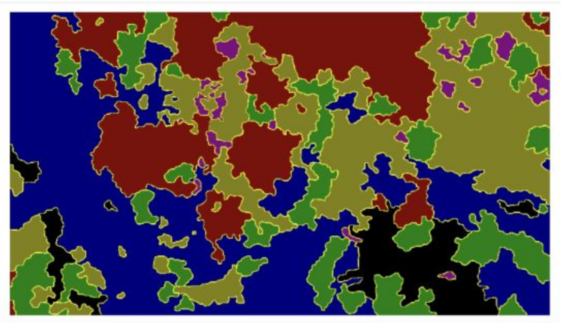
Uyen Nguyen

Undergraduate student in TCD

Problem Investigation

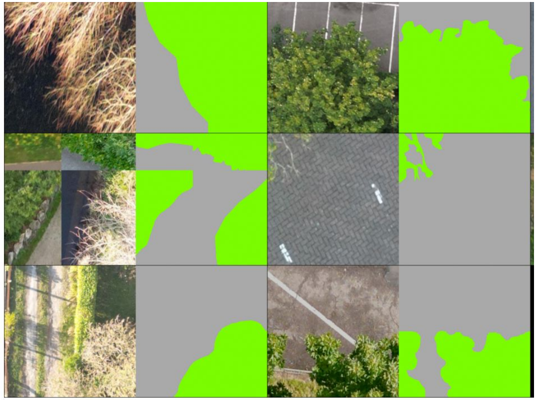
- This is a new problem, limited literature investigation
- Similar to remote sensing image segmentation, Aerial and Optical Images-Based Plant Species Segmentation





Challenges

- The small-scaled of the dataset:
  - **Public**: 4322 images in Train set & 689 images for Valid Set
  - **Warmup + Public**: 4658
  - **Private**: 783



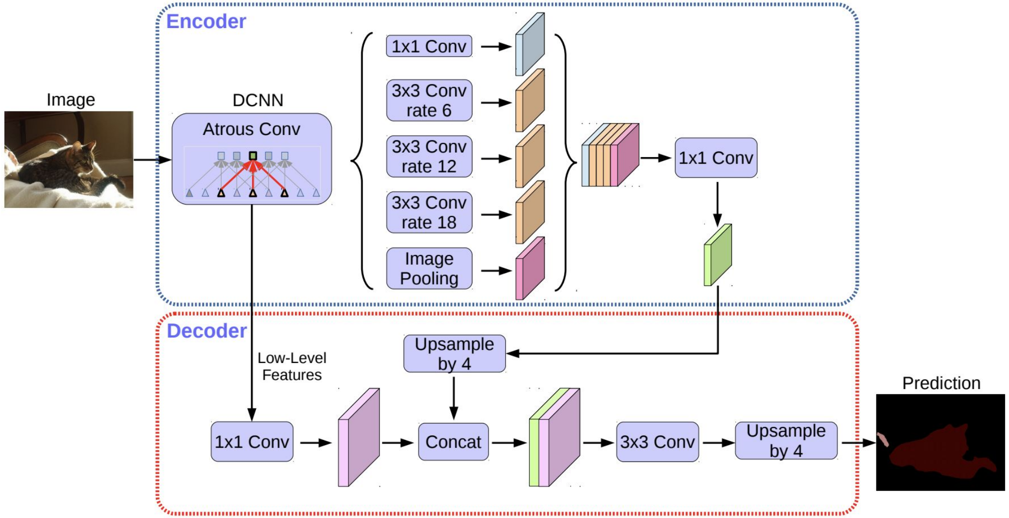
Some input examples

Results on Private Test

Models	High Vegetation IoU	High Vegetation Acc
Ours (DinoV2 & UNet++ & DeepLabv3+)	**86.67**	**93.97**
Ours (DinoV2)	86.5	94.00
Ours (UNet++ & DeepLabv3+)	84.6	92.43
vantuan5644	85.61	92.85
philip1	83.35	89.60

Solution - Models

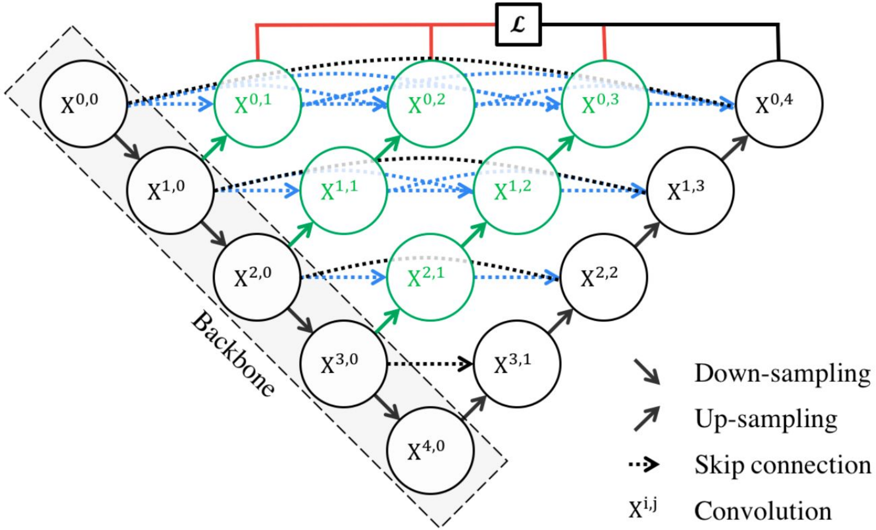
- Two **DeepLabV3+** with EfficientNet-b4 & EfficientNet-b5 backbone



Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation (Chen; 2018)

Solution - Models

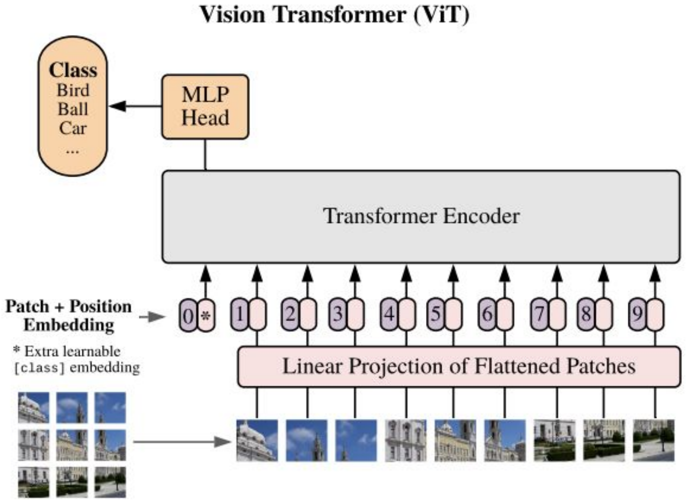
- One **UNet++** model with EfficientNet-b4 backbone



UNet++: A Nested U-Net Architecture for Medical Image Segmentation (Zhou; 2018)

Solution - Models

- Use ViT-b14 of DinoV2 as backbone, **add two Fully Convolutional Network Layers**, then re-train the model



DINOv2: Learning Robust Visual Features without Supervision (Oquab; 2023)

Solution – Loss function & Data Augmentation

- Our Loss function is the summation of **Dice Loss** and **Bootstrapping Cross Entropy Loss** (OHEM)
- For data augmentation, we employed **Mosaic image augmentation** with **spatial & color augmentation** including Random Crop, Horizontal & Vertical flip, Hue Saturation on the train set.

Solution – Training Methods

- **Challenges: Small-scaled dataset → Large-scaled model would be hard to converge during training**
- **Training tricks for fast convergence:**
  1. Freeze the encoder backbone layers, only train/fine-tune the segmentation head/decoder layers.
  2. Unfreeze all the layers of the network, fine-tune the whole model with 10 times smaller learning rate

Solution – Models Ensemble

- We choose top 5 models that have highest metric score on our validation set:
  - two deeplabv3+ (efficientnet B4&B5)
  - one unet++ (efficientnet B4)
  - and one DinoV2 (base);
- We gathered all the probability masks predicted by the models and average them to get the final segmentation mask for the private set

Results on Private Test

Models	High Vegetation IoU	High Vegetation Acc
Ours (DinoV2 & UNet++ & DeepLabv3+)	**86.67**	**93.97**
Ours (DinoV2)	86.5	94.00
Ours (UNet++ & DeepLabv3+)	84.6	92.43
vantuan5644	85.61	92.85
philip1	83.35	89.60

Discussion and Insights

Ideas:

- Talk about what we have tried? What worked and what didn’t?
    - ViT as backbone but didn't improve the results
    - Try to “smoothen” the boundary of the segment in predictions but didn't improve the results
    - YOLOv8 but data conversion is complicated, requires high-quality masks.
- How to improve in the future?
    - Implement hyperparameter tuning more thoroughly
    - Replace two-layer FCNs with more complex segmentation head (for example, Mask2Former head)

Conclusion

Our main contributions:

- Training a **DinoV2-ViTB14** with a **customized two-layer FCN** for semantic segmentation task.
- Combining both region-based and class-based loss functions as objective function, namely **Dice loss and OHEMCE loss**
- Applying **Mosaic augmentation** to generate variety of complex data scenarios to enhance models' training
- Introducing simple yet effective **technique to finetune** small-scaled dataset on such large state-of-the-art model
- Employing **ensemble method** that further boost the precision of predicted masks

Thank you