HUST

ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.

GRADUATION THESIS

ĐẠI HỌC
BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY
OF SCIENCE AND TECHNOLOGY

GRADUATION THESIS: APPLYING DEEP LEARNING FOR MULTI –CLASS SEGMENTATION IN DRONE IMAGERY

Student: Nguyen Ha Son - 20194660
Supervisor: Dr. Tran Nguyen Ngoc

ONE LOVE. ONE FUTURE.

TABLE OF CONTENTS

HUST

hust.edu.vn fb.com/dhbkhn

TABLE OF CONTENTS

1.1 Background

Background

- Nowadays, image segmentation has been applied widely in various applications
- Deep learning models can achieve high segmentation result

Challenges

- Lack of extensive datasets tailored to Vietnamese terrains and cultural features
- Limited number of experiments and studies focused on Vietnamese datasets
- AI models trained on datasets from different regions may not be well-suited for the Vietnamese landscape's context

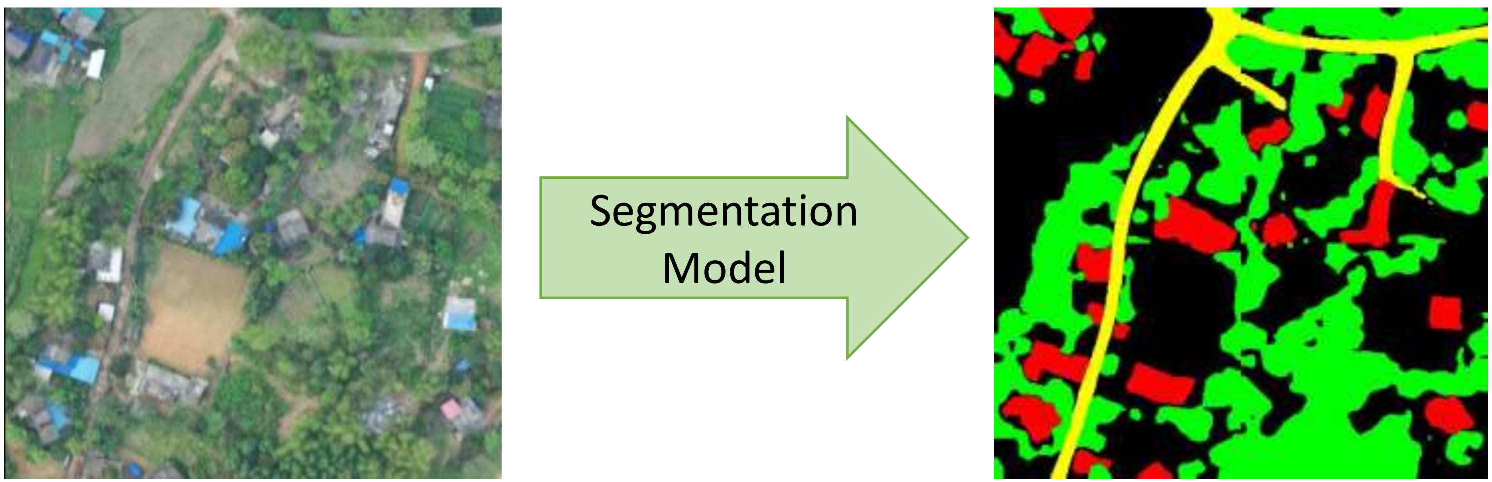
Drone image segmentation

→ This research aims to bridge this gap by developing deep-learning based segmentation process tailored to the specific characteristics of Vietnamese landscapes

1.2 Objective

Thesis Objectives

- Create a complete segmentation pipeline tailored to Vietnamese drone dataset
- Address a high-quality segmentation models that best suits the given problem
- The result could be used as a reference for future related works, and solve practical planning problems in the region



Original Image

Segmented Image

TABLE OF CONTENTS

HUST

hust.edu.vn fb.com/dhbkhn

2.1 Data Labeling

Dataset

- Includes scenes taken from a height of 300 meters, captured various regions in Vietnam
- Needs to be labeled to create the ground truth masks for the segmentation process, contains 5 classes: background, road, water, tree, building
- The labeling process uses Anylabeling tool

Labeling Process

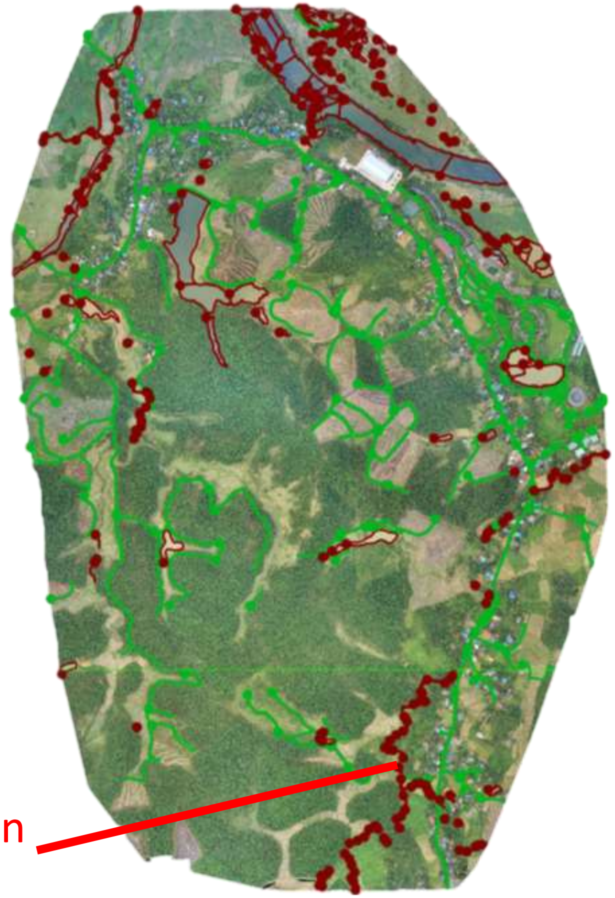
Original Scene

Zoom in to perform data labeling

Labeling Process

Original Scene

Labeled Scene

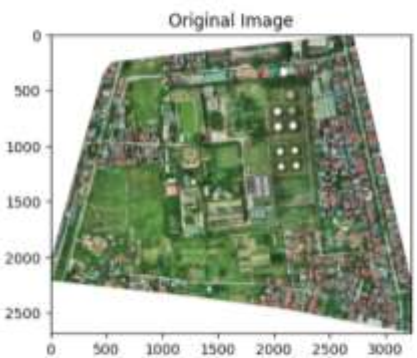


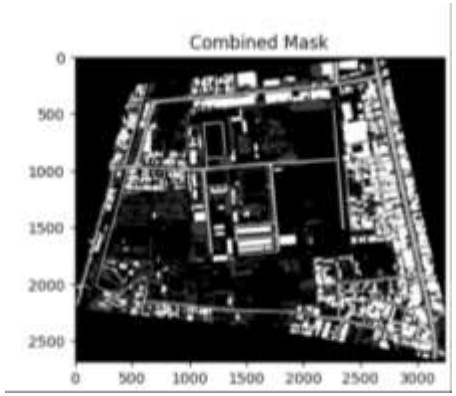
Label stored in JSON format

2.2 Data Preprocessing

Data Preprocessing Steps

Step 1: Convert JSON file to combined mask image. Mask consists of 5 classes, with label value from 0 to 4, then crop a large scene -> small patches





Cropped Patches with size 256x256

Step 2: Remove meaningless patch (contains only 0 pixel)

Step 3: Split data into train/validation/test with 8/1/1 ratio

Step 4: Perform augmentation on train set, combine augmented data with the original data

→ 3114 train (with 3114 augmented patches), 389 validation, 390 test patches

2.3 Models Implementation

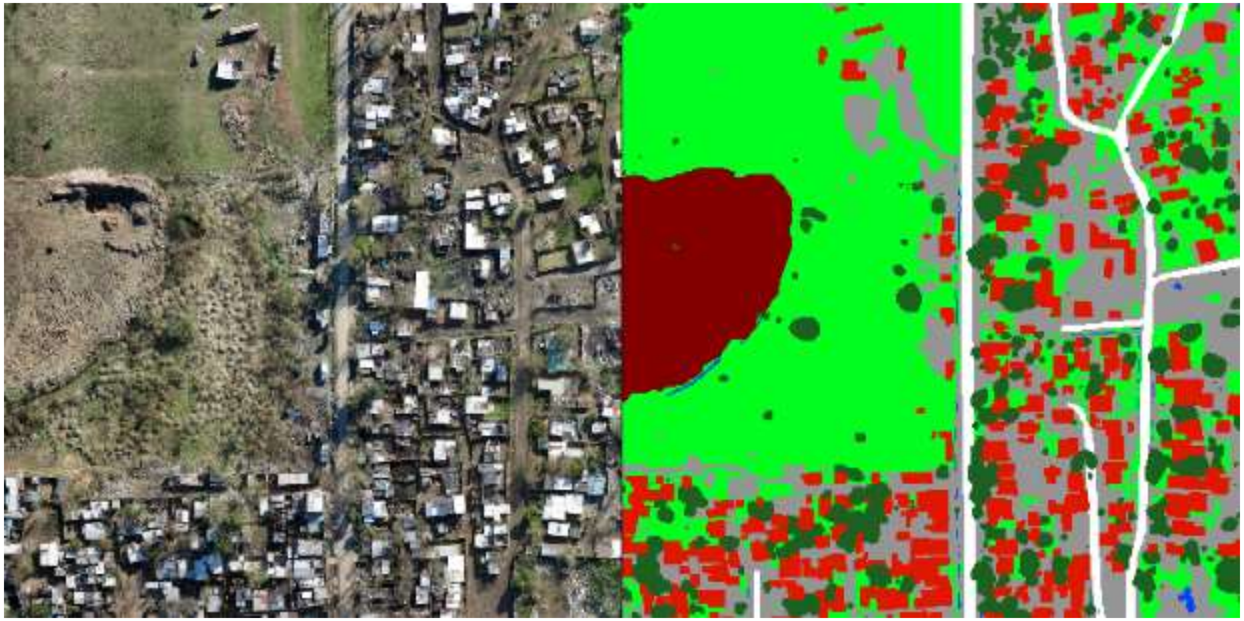
Baseline Models Implementation Strategy:

Applying multiple models, from older to newer models, with different complexity

Implemented Model	Type/Backbone	Total Params (million)
PSPNet	ResNet18	2.1
SegFormer	B0	3.7
	B3	47.2
	B5	84.6
FPN	ResNet50	26.9
Unet	ResNet50	32.5
UperNet	ConvNeXt Base	101.9
	ConvNeXt Tiny	37.0

Training with additional data

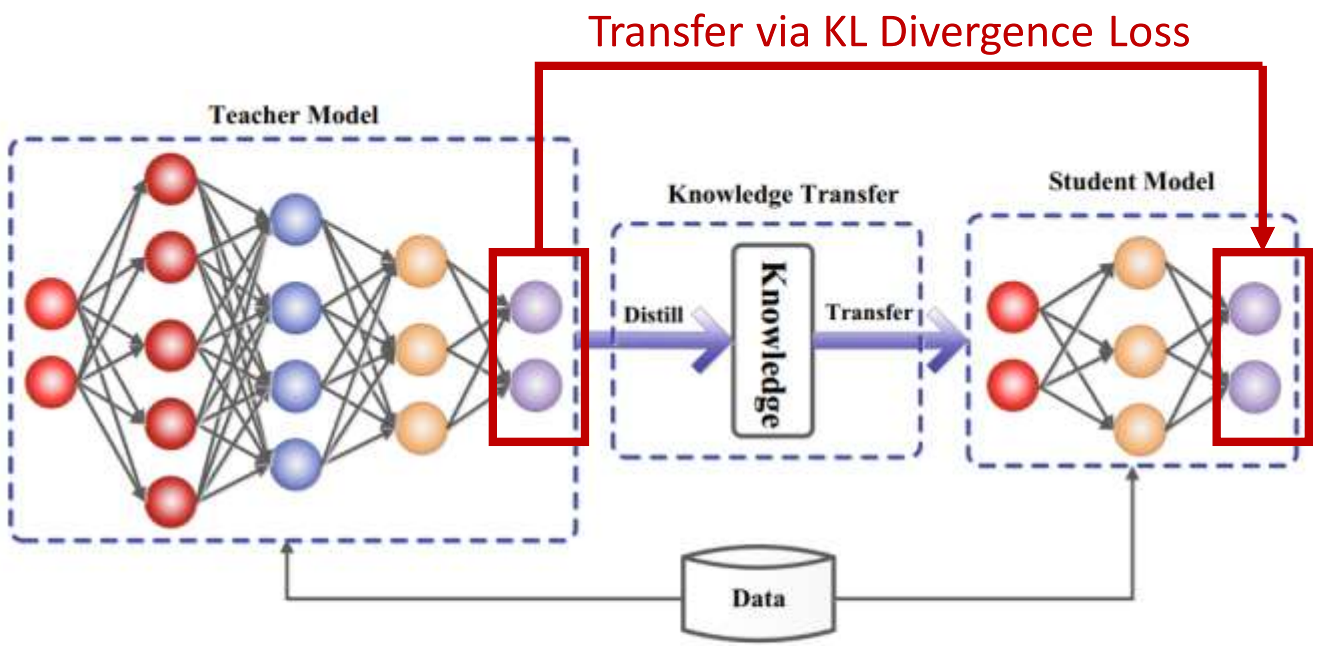
- ❖ Pretrain dataset: Open Earth Map, generalized dataset
- ❖ Train on pre-train dataset, then train with the original dataset afterward
- ❖ Consists of 8 classes -> convert 8 classes into 5 classes like in the original dataset



A scene in the Open Earth Map Dataset and its corresponding mask data

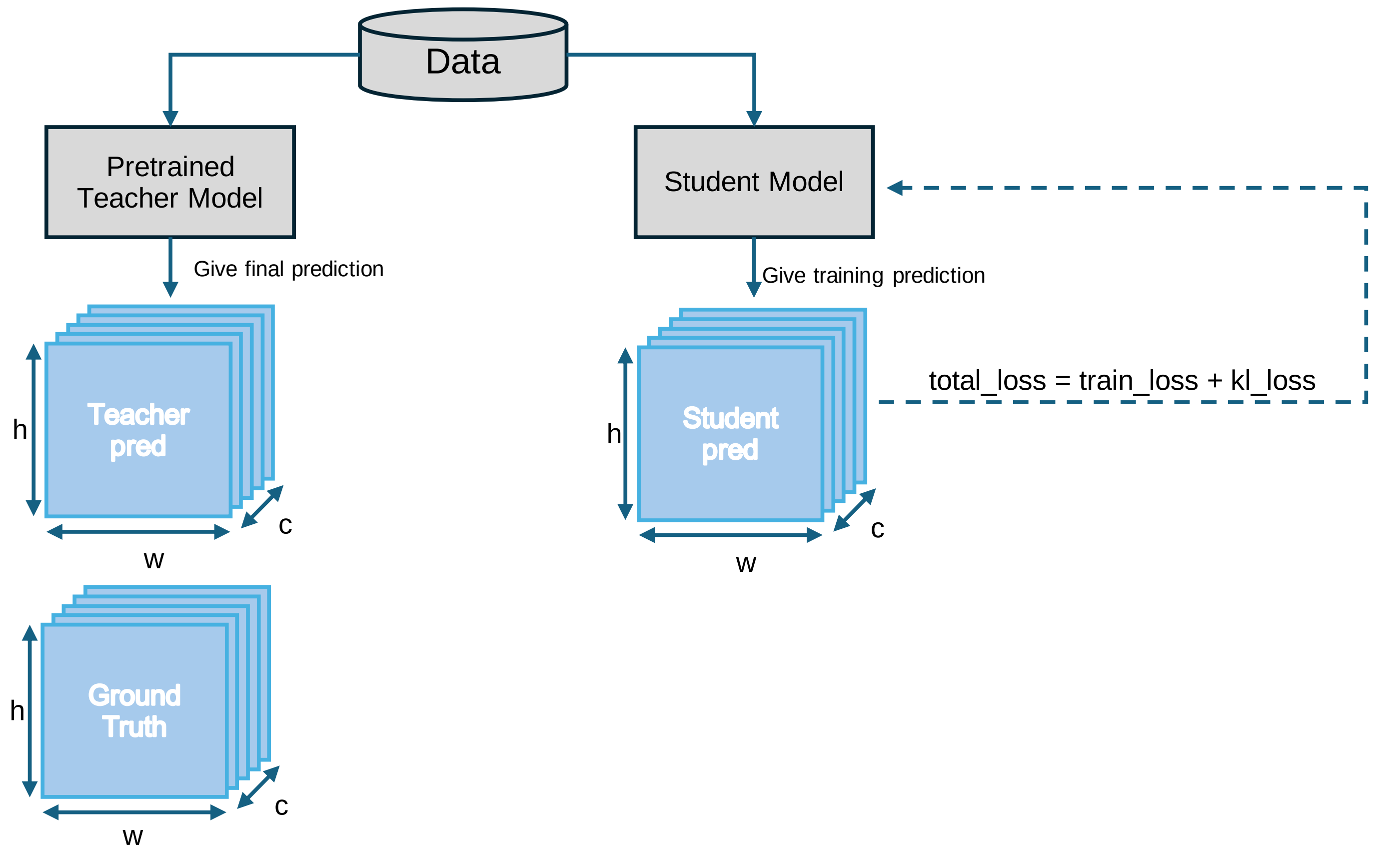
Offline Knowledge Distillation

- Used to improve light weight model's performance
- Suitable for the given problem
- Distill knowledge by comparing the between pretrained teacher's prediction and model's prediction via KL divergence loss



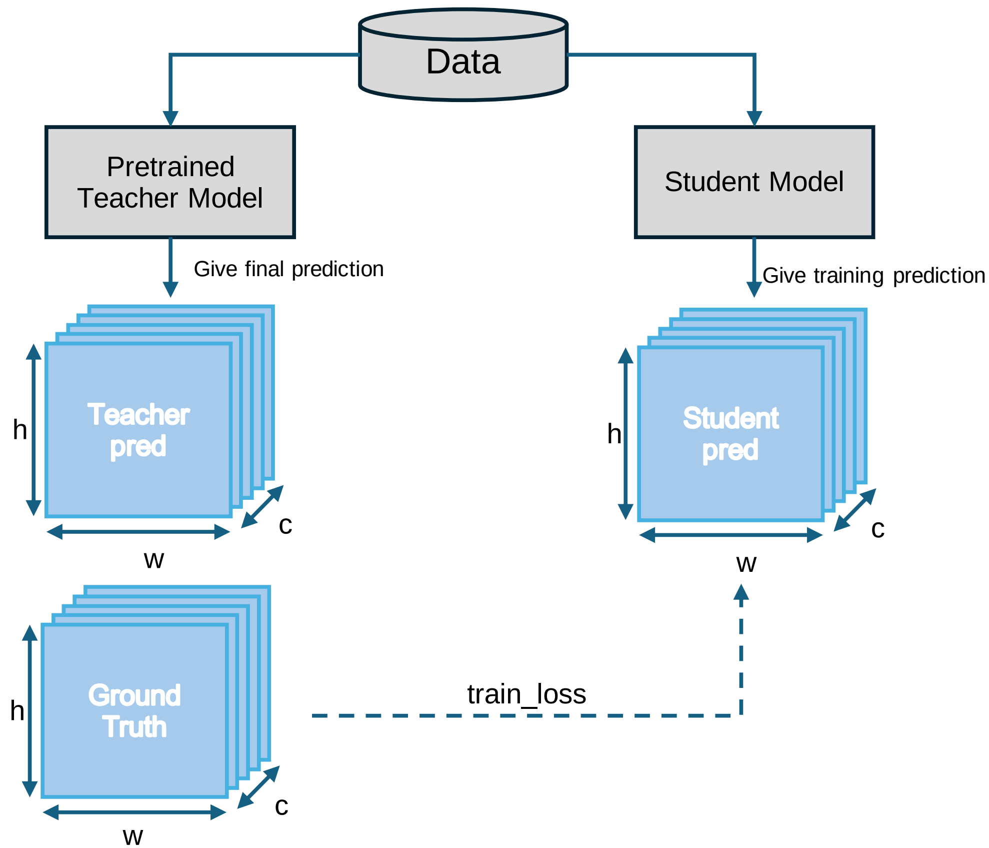
Offline Distillation Scheme For Classification Task

Offline Distillation For Segmentation Problem



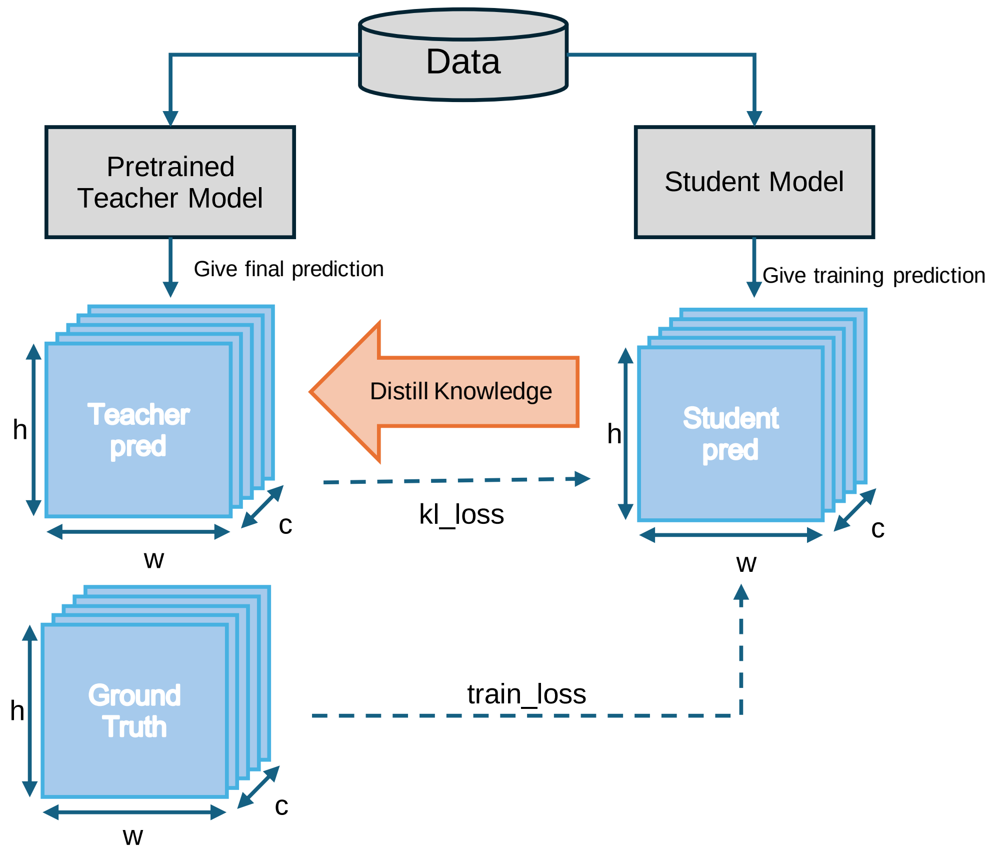
Offline Distillation Scheme For Segmentation Problem By Pixel-wise Comparison

Offline Distillation For Segmentation Problem



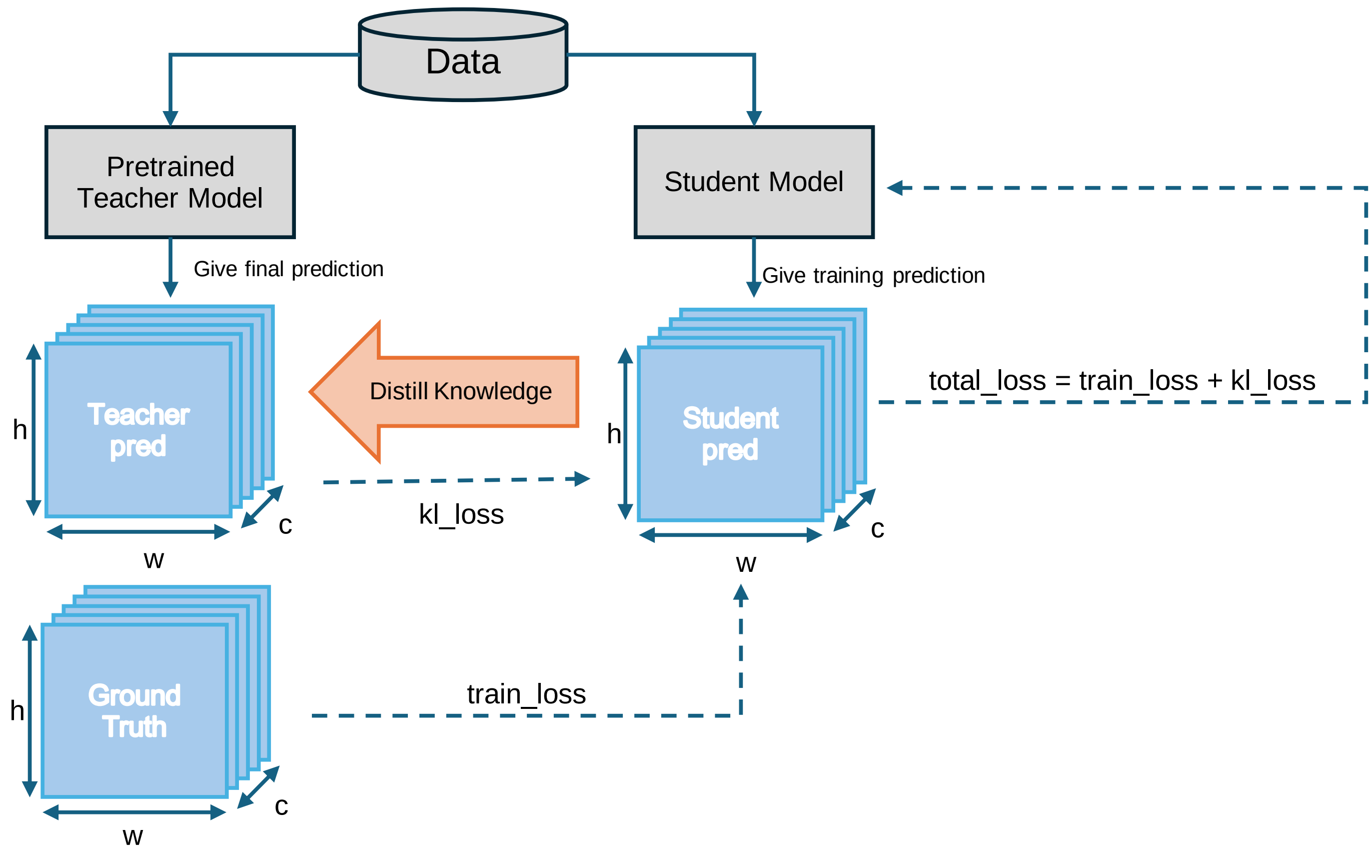
Offline Distillation Scheme For Segmentation Problem By Pixel-wise Comparison

Offline Distillation For Segmentation Problem



Offline Distillation Scheme For Segmentation Problem By Pixel-wise Comparison

Offline Distillation For Segmentation Problem



Offline Distillation Scheme For Segmentation Problem By Pixel-wise Comparison

TABLE OF CONTENTS

hust.edu.vn fb.com/dhbkhn

3.1 Evaluate result on Test Set

Training Setting

- Train all models for 200 epochs. After that, the models were trained with reduced learning rate until convergence
- Evaluate the result with mean IoU metric on the test set

Model	Type/Backbone	Params (million)	mIoU (%)
PSPNet	ResNet18	2.1	77.98
SegFormer	B0	3.7	**79.73**
	B3	47.2	80.25
	B5	84.6	80.22
FPN	ResNet50	26.9	81.60
Unet	ResNet50	32.5	80.66
UperNet	ConvNeXt Base	101.9	81.23
	ConvNeXt Tiny	37.0	**81.91**

3.2 Training with additional data

Training with additional data setting

- Choose the model with best performance (UperNet with ConvNeXt Tiny backbone), and the best light weight model (SegFormer B0) for training with additional data
- Train with pretrain-data for 200 epochs, after that train the model with the original dataset

Model	Additional Data	mIoU (%)
UperNet (ConvNeXt Tiny)	×	81.91
	√	**82.49**
SegFormer B0	×	79.73
	√	**80.65**

→ Both models benefit significantly from additional training data

3.3 Distilled vs Vanilla Result

Result after the first 200 epochs

- Experiment with PSPNet (ResNet18) and SegFormer B0 as student, and UperNet as teacher
- All models were trained for 200 epochs with Adam LR of 0.001 and batch size of 16, and then evaluate the results

Model	mIoU (%)
Vanilla PSPNet	77.98
Distilled PSPNet	**79.28**
Vanilla SegFormer	78.40
Distilled SegFormer	**79.38**

✓ Training models with distilled setting gives better performance and models also tend to converge faster

3.3 Distilled vs Vanilla Result

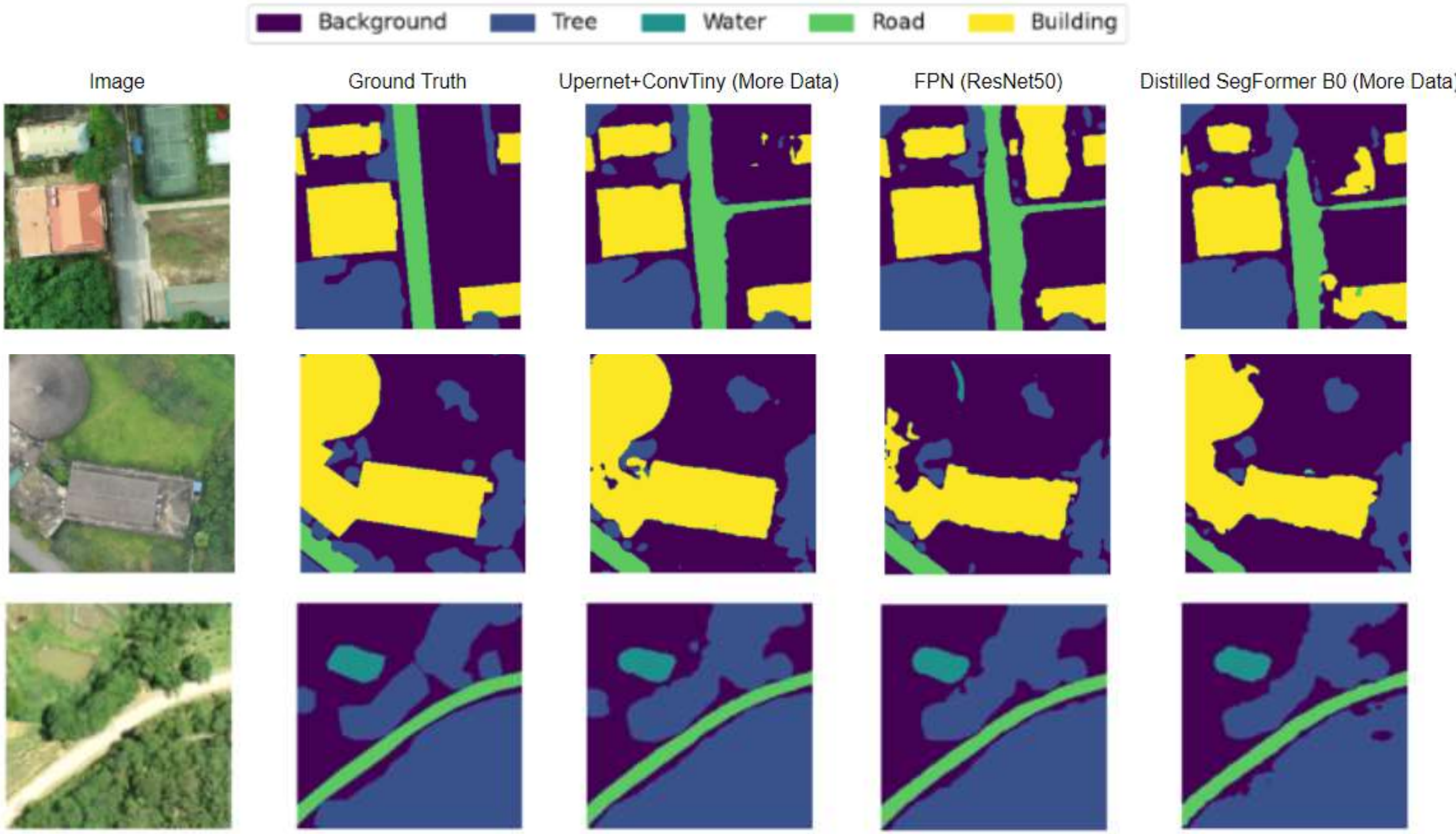
Combine knowledge distillation with pretrain-data

Applying both fine tuning & knowledge distillation for SegFormer B0, with fine tuned UperNet with ConvNeXt Tiny as teacher model. Train all model until convergence

Model	Distilled	Additional Data	mIoU (%)
SegFormer B0	×	×	79.73
	√	×	80.65
	√	√	**80.94**

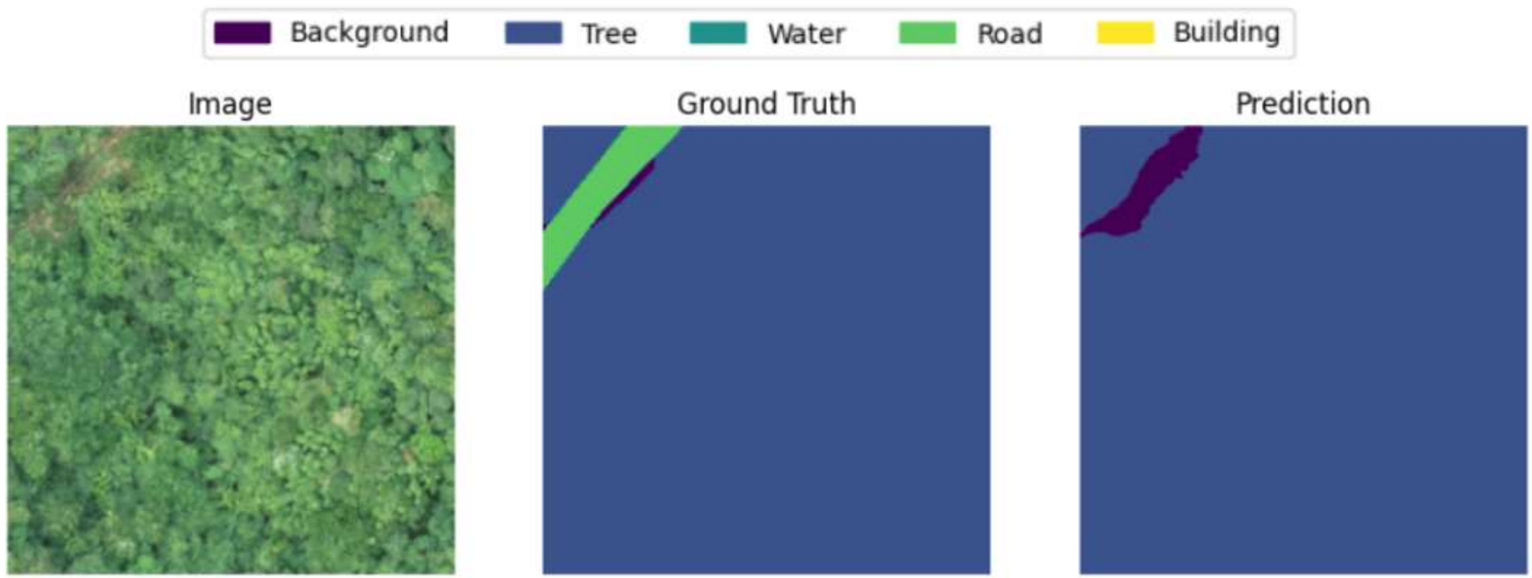
- ✓ The newly created model has performance better then its B3 and B5 counterpart
- ✓ Could have better result if a better teacher is available

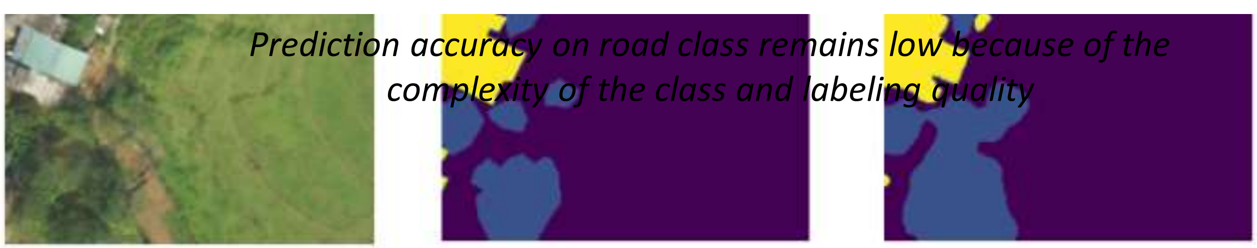
3.4 Prediction on Test Patches



Models give high quality segmentation result. Fine tuned UperNet with ConvNeXt Tiny tend to perform better when it comes to hard to predict region

Prediction on Test Patches





Prediction accuracy on road class remains low because of the complexity of the class and labeling quality

Prediction produced by UperNet with ConvNeXt Tiny could handle mislabeling case

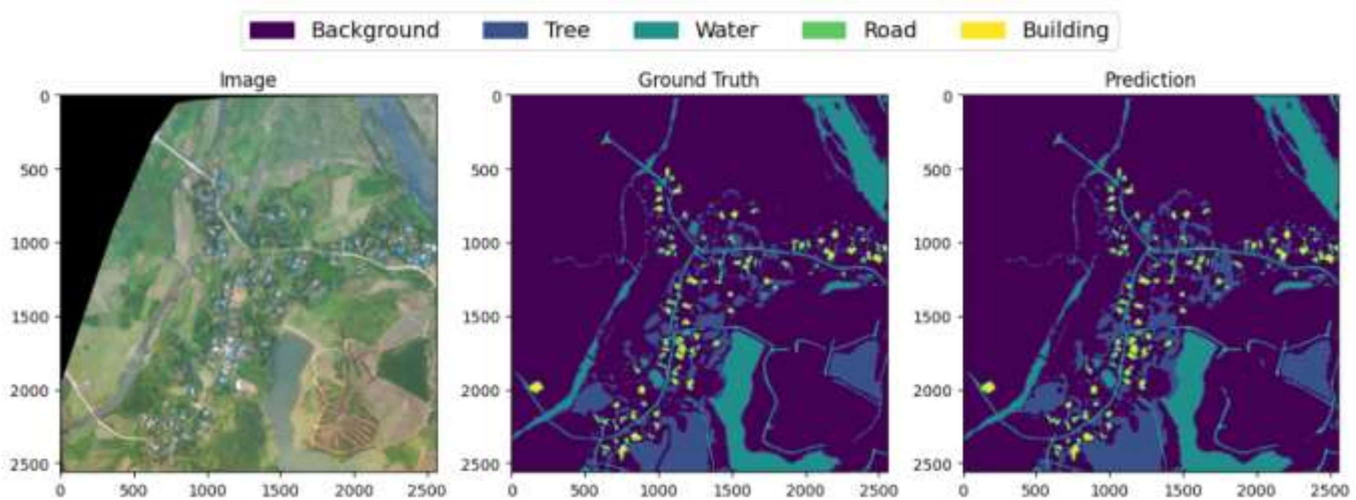
HUST

hust.edu.vn fb.com/dhbkhn

TABLE OF CONTENTS

Obtained Result

- ✓ Create a complete segmentation pipeline for Vietnamese Drone Imagery Dataset
- ✓ **Best Performance Experiment:** UperNet + ConvNeXt Tiny backbone, reaching **82.49%** mIoU (with additional data)
- ✓ Lightweight model could achieve good performance in the given dataset, with distilled SegFormer B0 reaching **80.94%** mIoU with additional data
- ✓ Effectiveness of Training with more data and Knowledge Distillation



Inference on a large image using fine tuned UperNet with ConvNeXt Tiny Backbone

HUST

hust.edu.vn fb.com/dhbkhn

THANK YOU !