Semantic Segmentation for Agriculture

Guide : Dr. UMAPATHY

110117039 Kamalesh P

110117073 Rupp Santos

110117084 S Ganesaan

110117087 Sanjay Kumar

Introduction

Agricultural lands are important for the economy of any country especially for a country like India which relies heavily on agriculture. Understanding the yield of an agricultural land is a very strenuous task, due to several predictable factors. By using machine learning we can improve the predictions of the yield.

Weed, standing water, planters skip are some of the most significant factors that can reduce crop yield. Computer vision combined with image processing is an effective method for weed detection, and site-specific weed management has become an effective tool for weed control. The use of an encoder-decoder deep learning network for pixel-wise semantic segmentation of crop and weed was explored in this study.

Objective

→ The objective of our project is to use semantic segmentation to segment different regions of farmland to obtain the type. This can be used to predict the yield of a particular region.

Motivation

India has made several advances pertaining to the field of agricultural sciences and with the effective supply chain management has led to food wastage being reduced, but there is a lot of potential for AL/ML tools to be used to for assisting the farmers.

Existing algorithms use convolution or graph-based networks which do not give good performance. CNNs are difficult to train because they require a lot of samples to gain a good performance on image datasets. Attention mechanism has recently shown that we can se simple inductive biases to understand the features from the images. We use recent advances in attention-based models such transformers to improve the accuracy of models on agriculture datasets.

PROCESS

Literature review

Test baseline model

Choosing the CNN backbone

Write the DETR Code

Generate the masked outputs

Literature Review

Reducing the feature divergence of RGB and near-infrared images using Switchable Normalization

- The dataset used in the agriculture vision task contains RGB and Near-Infrared images. Most of the existing works use all these images together as a single input to the CNN model.
- But when tested using KL divergence, it was found out that the RGB features and the Near-Infrared(NI) features were divergent. By ensuring that the features are aligned properly, we can improve the mAP score of existing methods by around 10%.
- The paper proposes using a modified version of Instance Normalization to achieve this.

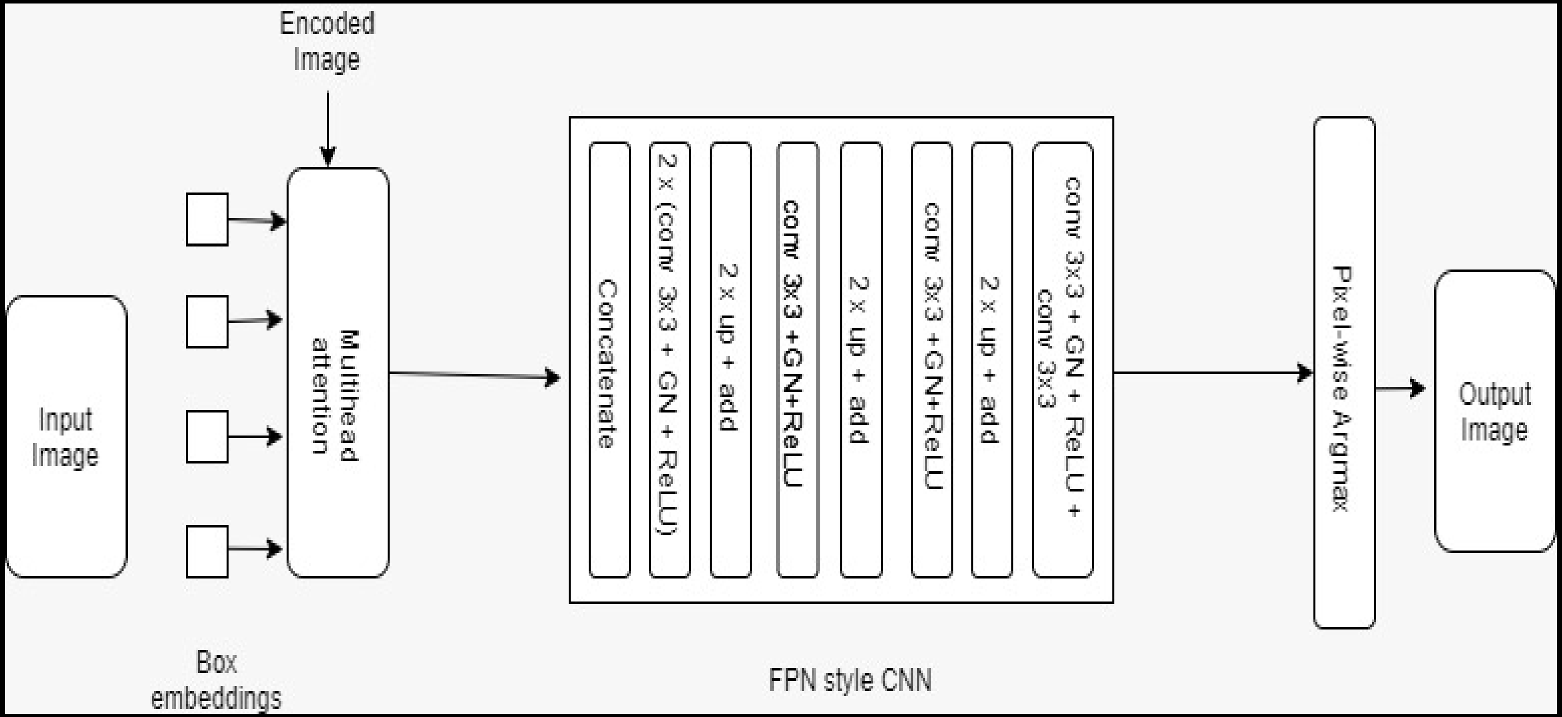
Multi-view Self-Constructing Graph Convolutional Networks with Adaptive Class Weighting Loss for Semantic Segmentation

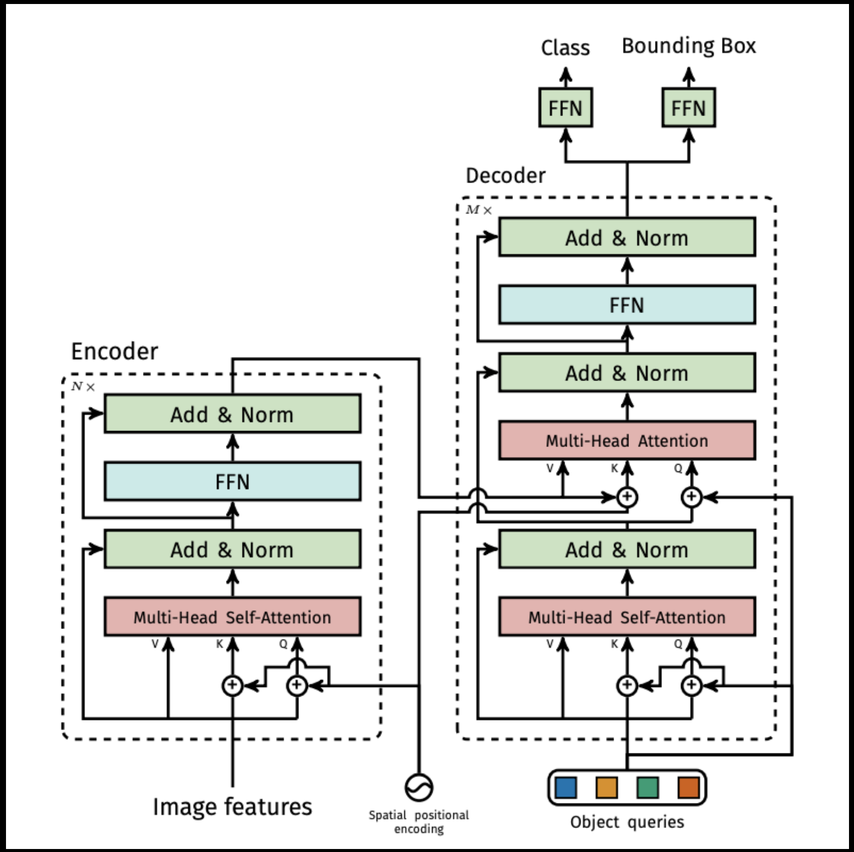
- This paper proposes multiple novelties, one in the form of a graph neural network that is used to extract the features and the other is a loss function that takes care of the class imbalance problem in each image. This modification improves the results by 5% over the existing state-of-the-art.

Dataset Used

- The dataset used in this thesis is a subset of the Agriculture-Vision dataset. The dataset contains 21,061 aerial farmland images captured throughout 2019 across the US. Each image consists of four 512x512 colour channels, which are RGB and Near Infra-red (NIR).
- Each image also has a boundary map and a mask. The boundary map indicates the region of the farmland, and the mask indicates valid pixels in the image. Regions outside of either the boundary map or the mask are not evaluated.
- This dataset contains six types of annotations: Cloud shadow, Double plant, Planter skip, Standing Water, Waterway and Weed cluster. These types of field anomalies have great impacts on the potential yield of farmlands, therefore it is extremely important to accurately locate them. In the Agriculture-Vision dataset, these six patterns are stored separately as binary masks due to potential overlaps between patterns.

DETER Model Used for Semantic segmentation





Input Mode for DETR

- → ResNet Backbone: The resnet model is used to extract the features from the images which is given as input to the transformer model. ResNet follows VGG’s full 33 convolutional layer design.
- → Patch Embeddings: In this type, we reshape the image into a sequence of flattened 2D patches where (H;W) is the resolution of the original image, C is the number of channels, (P; P) is the resolution of each image patch, and N = HW/P^2 is the resulting number of patches, which also serves as the effective input sequence length for the Transformer.

Loss Function

→ Further to improve the accuracy of our models, we used different loss functions for semantic segmentation. The distribution of the classes is highly imbalanced in the dataset. To address this problem, most existing methods make use of weighted loss functions with precomputed class weights based on the pixel frequency of the entire training data to scale the loss for each class-pixel according to the fixed weight before computing gradients. In this work, we use a novel class weighting method based on iterative batch-wise class rectification, instead of pre-computing the fixed weights over the whole dataset.

$$f_j^t = \frac{\hat{f}_j^t + (t-1) * f_j^{t-1}}{t} .$$

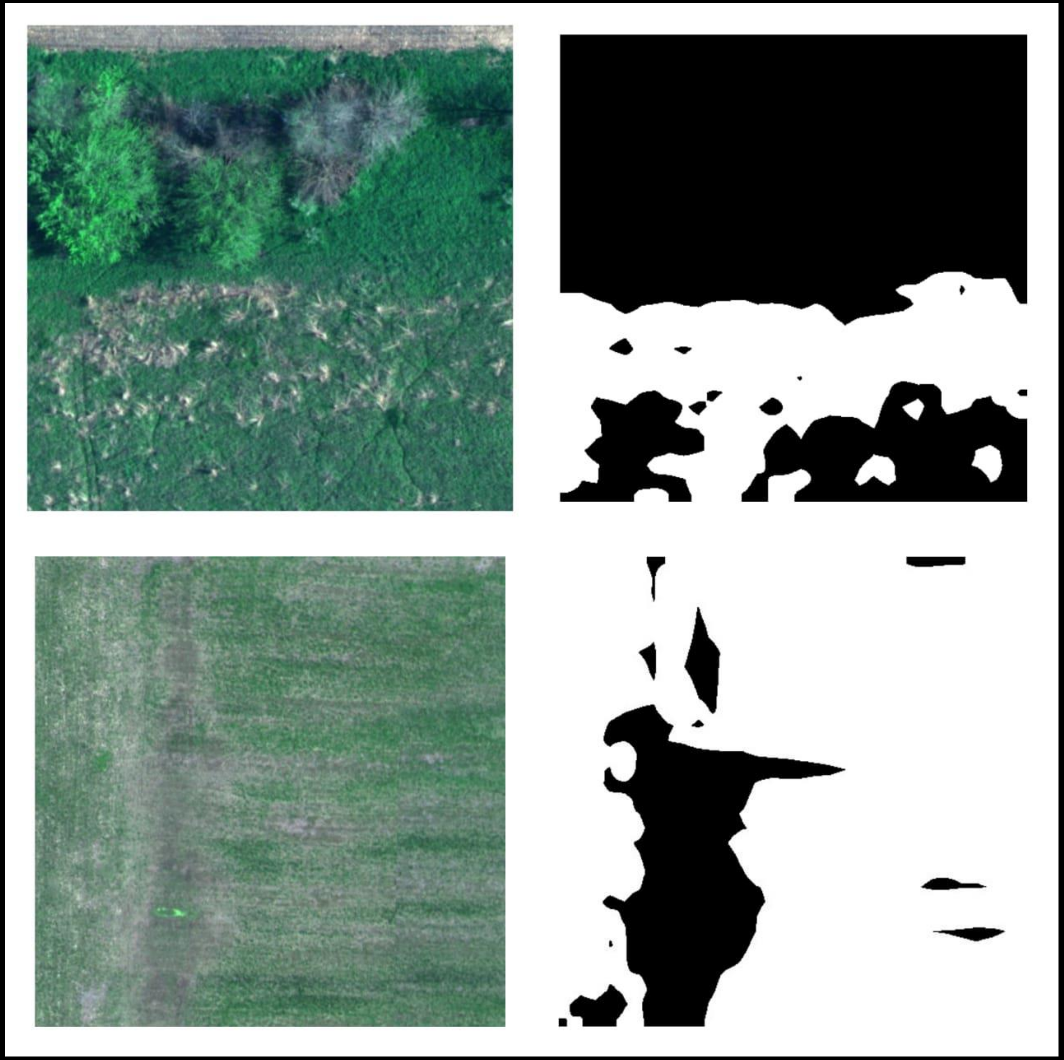
$$\mathcal{L}_{acw} = \frac{1}{|Y|} \sum_{i \in Y} \sum_{j \in C} \tilde{w}_{ij} * p_{ij} - \log\left(\mathrm{MEAN}\{d_j | j \in C\}\right) ,$$

Hyperparameters

→ We used grid search to find the learning rate for the model. The results of the grid search showed us that a learning rate of 0.001 with the Adam Optimizer gave the best results. We trained our models for over 50 epochs using free GPU on google colab. Further, for the decoder we needed to choose the right number of layers and found out of that six layers of convolution followed by upsampling layers gave the best results.

Result

Model	mIOU(%)
MSCG-Net - ResNet - 50	55.0%
MSCG-Net - Ensemble	66.0%
MSCG-Net - ResNext - 50	54.8%
DETR-ResNext 50	**57.1%**
DETR-ResNet101	56.1%



Future Work

1. **Vision Transformer** - We are planning to replace the CNN baseline with vision transformer, as this model has proved to be better at Image Classification in recent years.