Semantic Segmentation of the aerial/satellite imagery

Team – Space Applications Centre, ISRO

Problem Statement

Problem statement: Earth Observation Applications

Title: Semantic Segmentation of the aerial imagery

Technique Details: **Semantic segmentation technique using Artificial Intelligence** is the process of classifying each pixel in an image belonging to a certain class and hence can be thought of as a classification problem per pixel using AI methods. This technique is used here to classify each pixel in aerial imagery to a certain class.

Dataset Details

- Orthorectified Images (ORI)
  - 4 Channels: RGB and alpha
- Digital Elevation Model (DEM)
- Class Boundaries in GeoDataBase (.gdb)
  - Major classes are Vegetation, Roof/BuiltUp, Open Area, Road etc.

Hardware Details and Technology Stack

GPU: Nvidia Quadro RTX 8000
RAM: 512 GB
Storage: 1 TB SSD

Solution Technology: Artificial Intelligence (Deep Learning)
Language: Python
Major Libraries:
- Streamlit: To develop Web Interface
- Rasterio: To handle ORI images as GIS data
- Keras, tensorflow: To develop DL models

Tools: QGIS, ArcGIS, Anaconda3

Dataset Preparation

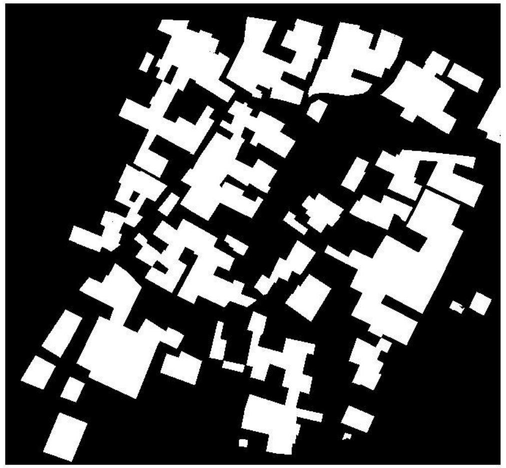
Operations on ORI Data

1. Removal of alpha band
2. DEM image resampling and stacked with ORI data
3. Clipped stacked data with extent of 'roof/builtup' gdb.

4 Channels ORI
(R, G, B, DEM)

Operations on Class Boundaries

1. Vector data cleaning
2. Conversion of vector data to image mask keeping ORI data resolution.

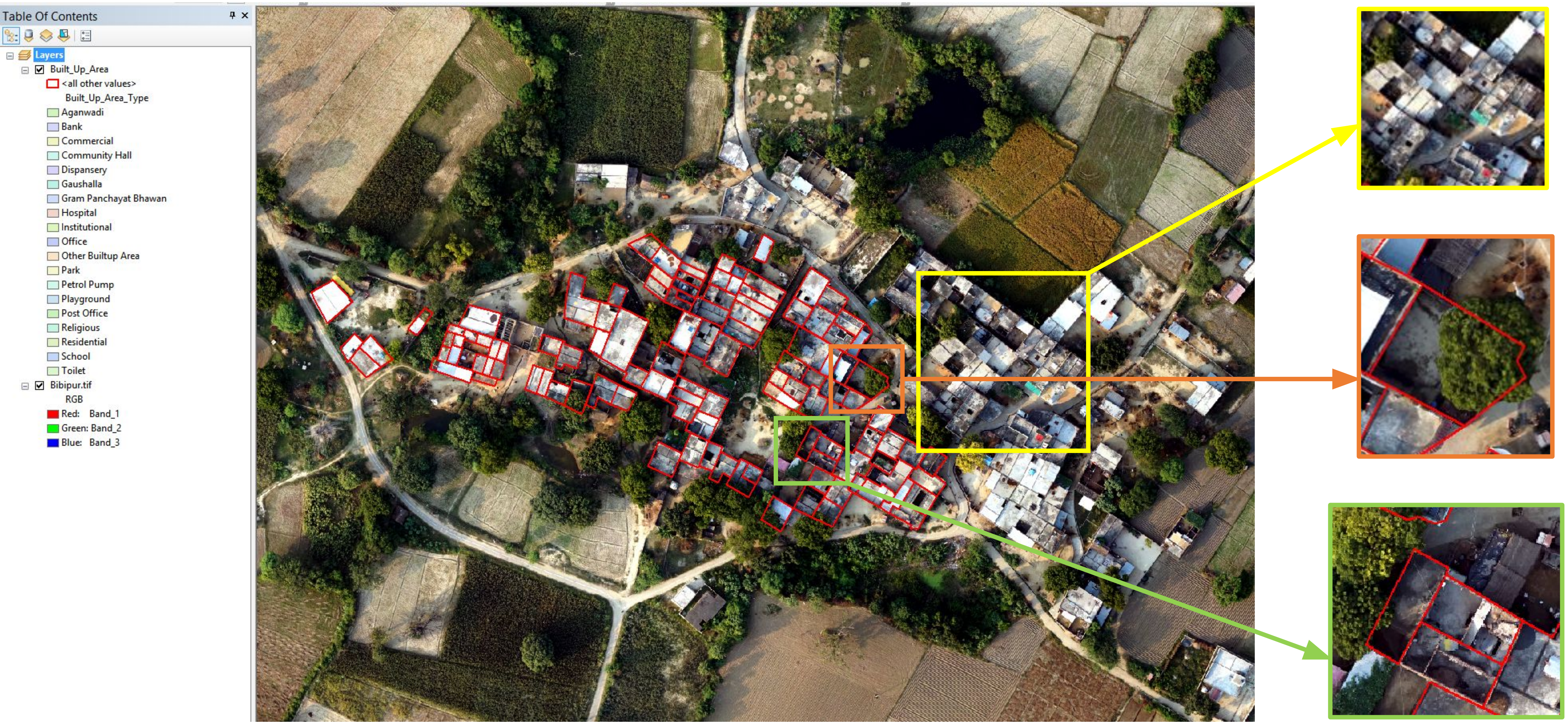


Roof/BuildUp Mask

Challenges during dataset preparation

Challenges	Actions taken
Though dataset had multiple classes but only rooftop/ built-up were the only classes that were common across multiple datasets.	Semantic segmentation scope was reduced to produce semantic map for rooftop/ built-up class.
Rooftop/ built-up class was not properly labelled: 1. Whole dataset was not labelled. 2. Overlapping over other classes such as open-area, vegetation, trees etc.	1. Digitization was done for the rooftop/ built-up vector data extent. 2. Removal of Overlapping of classes to some extent.
Datasets 1. ORI Images and some of the provided vector data (GDB) were in different projection (ALUPUR, CHIBRAMAU, KANNAUJ(146874)) 2. ORI images had either 3 or 4 (+alpha) channels 3. Though sufficient ORI images were provided but associated vector data was incomplete.	1. Re-projection has been done. 2. Alpha channel was removed from ORI images. Integrated DEM as one of the channel in input images. 3. Data volume increased for training through augmentation techniques

Challenges during dataset preparation

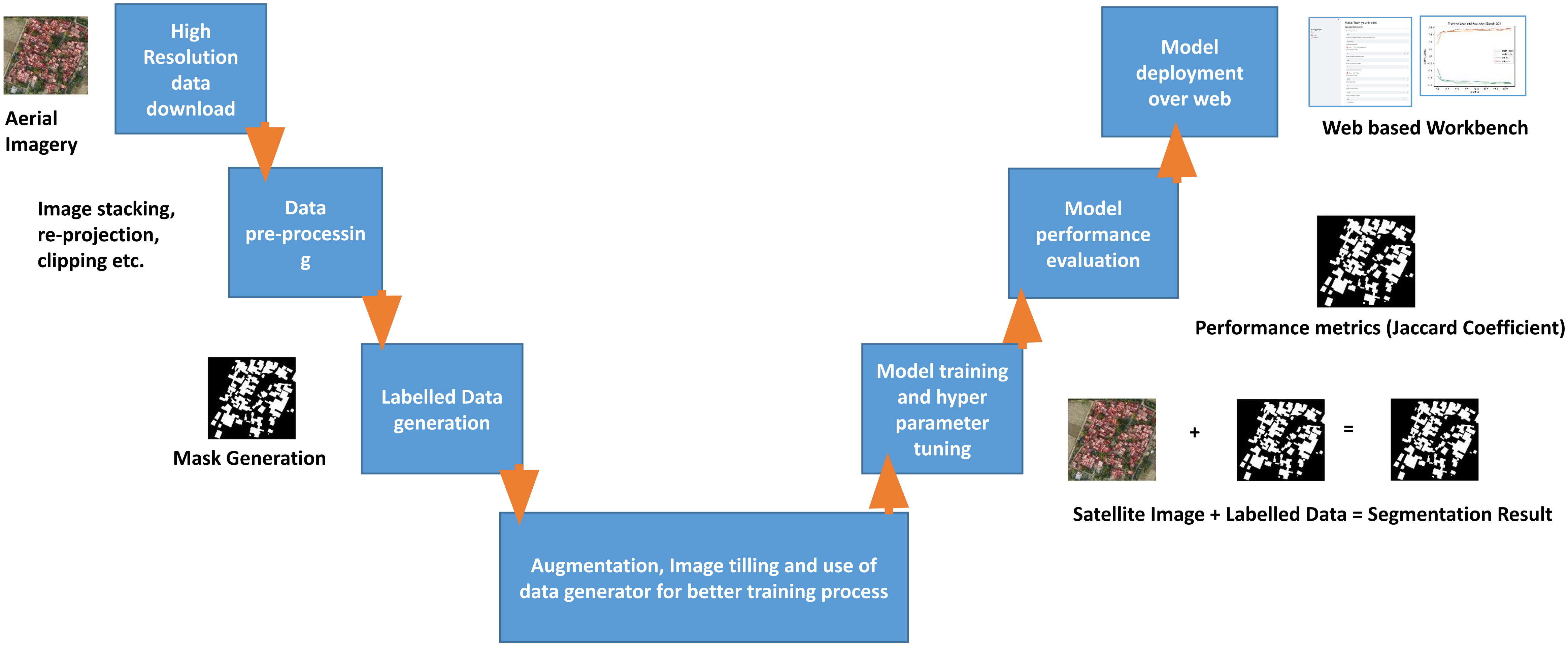


Un-labelled / Non-Digitized Data

Improper/ Cross Class Labelling

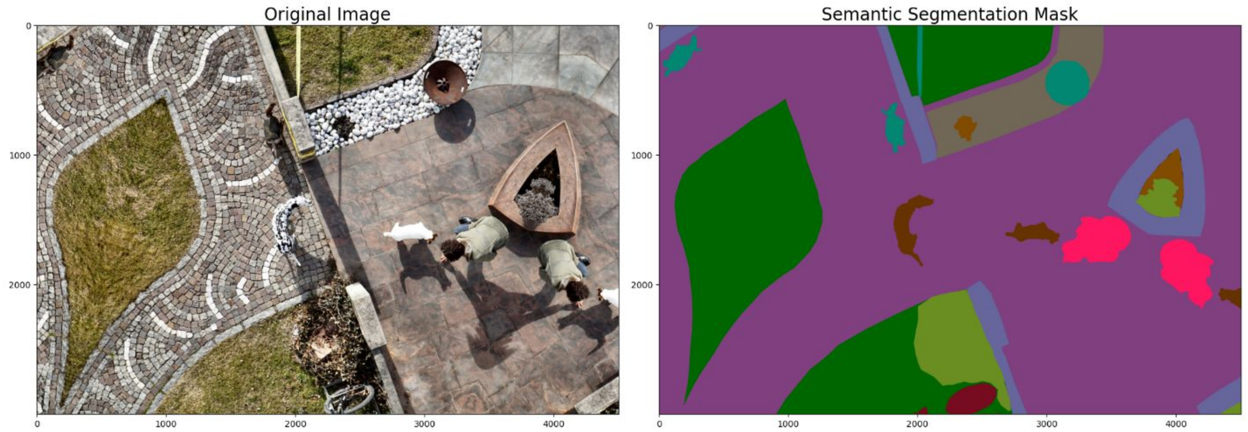
Open Area marked as Built-Up

Block Diagram of Proposed Solution



Explored Approaches

- Developed prototype segmentation model on publically available data (semantic drone dataset).



24 Classes:
paved-area, dirt, grass, gravel, water, rocks, pool, vegetation, roof, wall, window, door, fence, fence-pole, person, dog, car, bicycle, tree, bald-tree, ar-marker, obstacle, conflicting and unlabeled class.

  - Model didn’t performed well on SoI data because of:
    - Different resolution of train (public dataset) and test (SoI dataset) datasets.
    - Different Sensor sensitivity resulting into difference in tone, texture and context.

- Tried transfer learning using VGG-16 and ResNet pretrained models. The pretrained models are trained on 3 channels RGB images only and hence cannot be scaled for more channels.

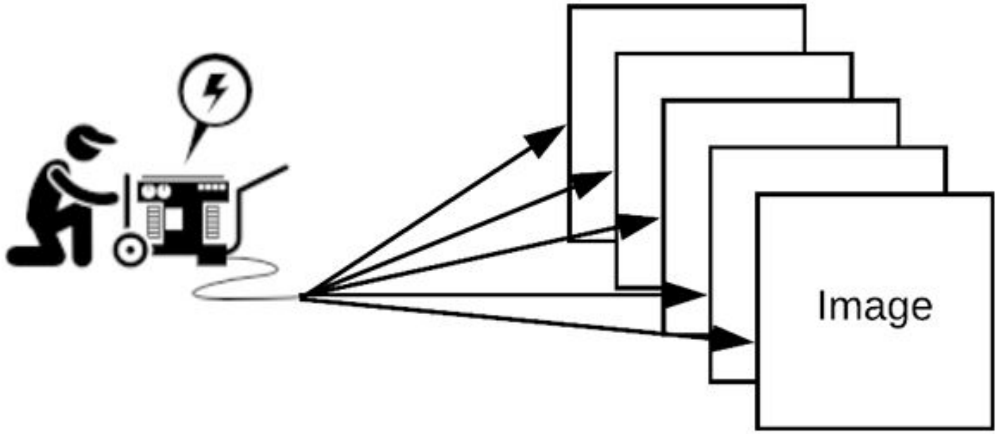
Explored Approaches for Training & Prediction

- Adopted strategy involved creating train datasets from provided SOI data. Limited datasets were augmented and tiled to increase data volume.

(512, 512, 4)

(512, 512, 1)

- Use of Custom Data Generator to efficient utilization of memory resources.



Architectures - UNET

- UNET is used for performing semantic segmentation of satellite imagery.
- The architecture contains both encoder and decoder paths.
- First path is the contraction path (also called as the encoder) which is used to capture the context in the image. The encoder is stack of convolutional and max pooling layers.
- The second path is the symmetric expanding path (also called as the decoder) which is used to provide precise localization using transposed convolutions. It is an end-to-end fully convolutional network (FCN).
- Labelled data is generated using satellite imagery for providing it to model with images for training.

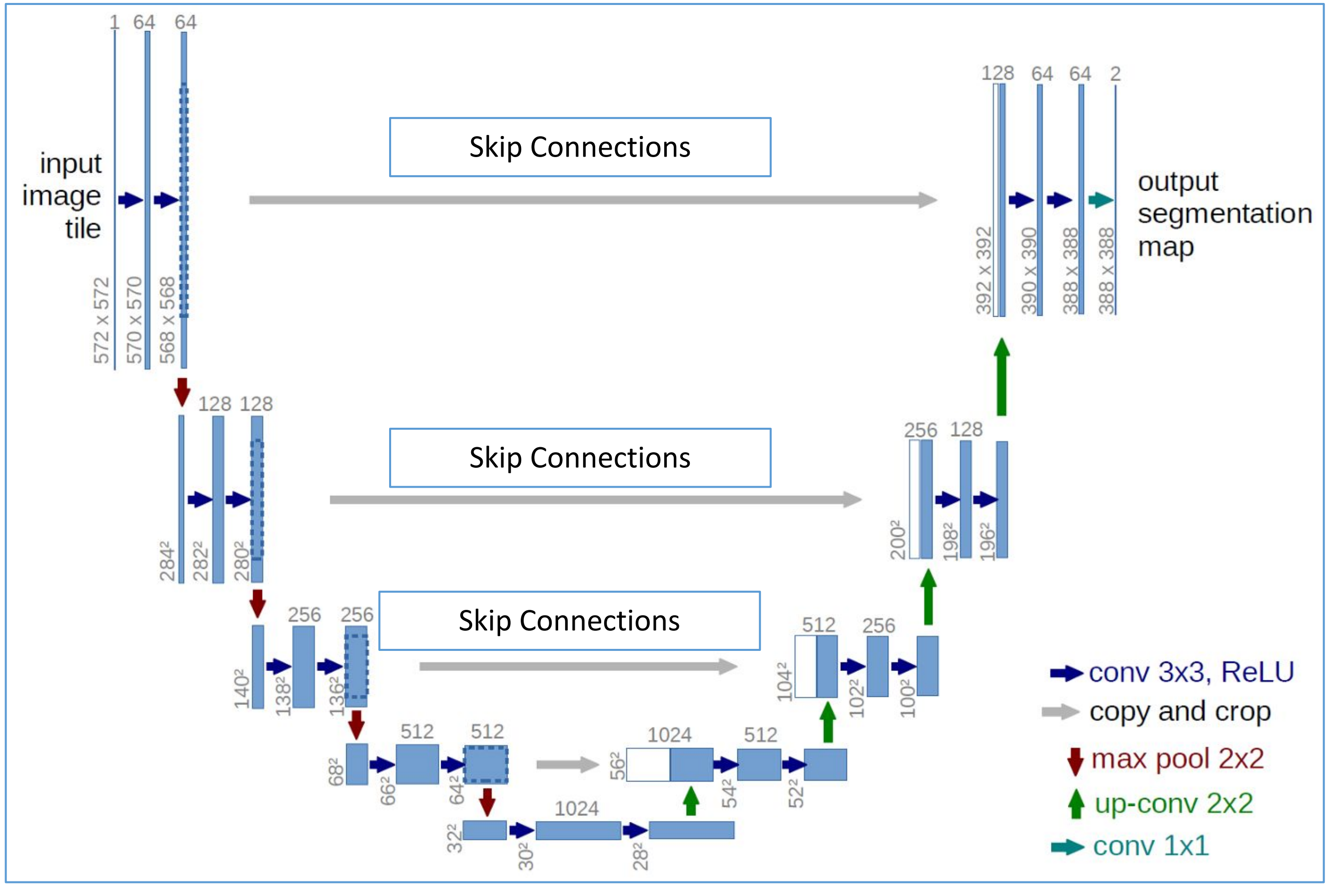


Fig.: Illustration of UNET Architecture with encoder and decoder path

Architectures – UNET-ASPP or Atrous Spatial Pyramid Pooling

UNET - Atrous Spatial Pyramid Pooling

The Atrous Spatial Pyramid Pooling os ASPP layer is applied at bottleneck layer i.e. between encoder and decoder part of the UNET. This layer captures multi-scale features by applying multiple parallel filters at different dilation rate, thus increasing effective Field of View (EFoV). These filters are later passed to decoder path after concatenating and convoluting.

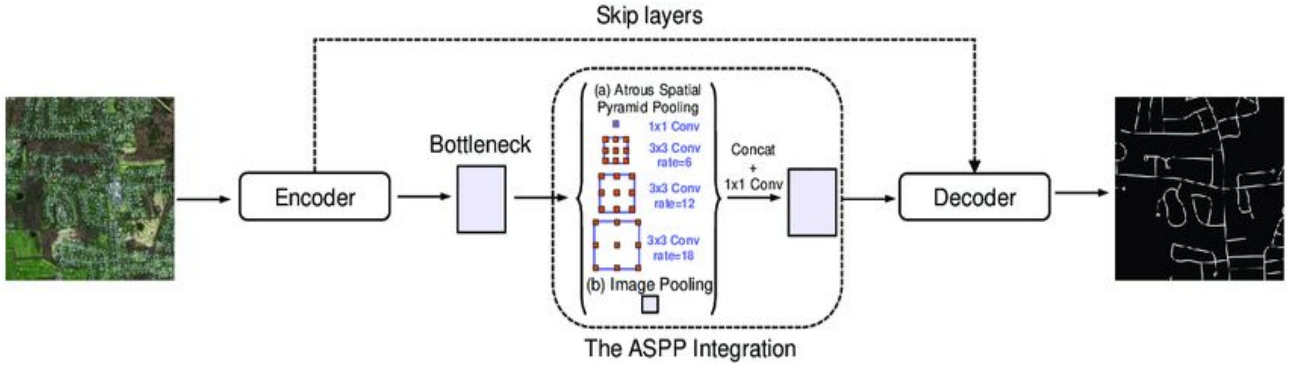


Fig.: Illustration of ASPP Integration in UNET architecture

Architectures – UNET Attention

Attention

- While traditional UNET takes skip connections directly as input, UNET-Attention or Attention aware UNET applies attention as ‘weights based on importance’ over skip connections.
- The skip connections are later on concatenated with output layers at each depth in decoder path.

The Jaccard Coefficient (JC) is used as model loss (negative of JC) and performance metric in proposed solution. It is defined as:

$$JC(A,B) = \frac{A \cap B}{(A \cup B) - (A \cap B)}$$

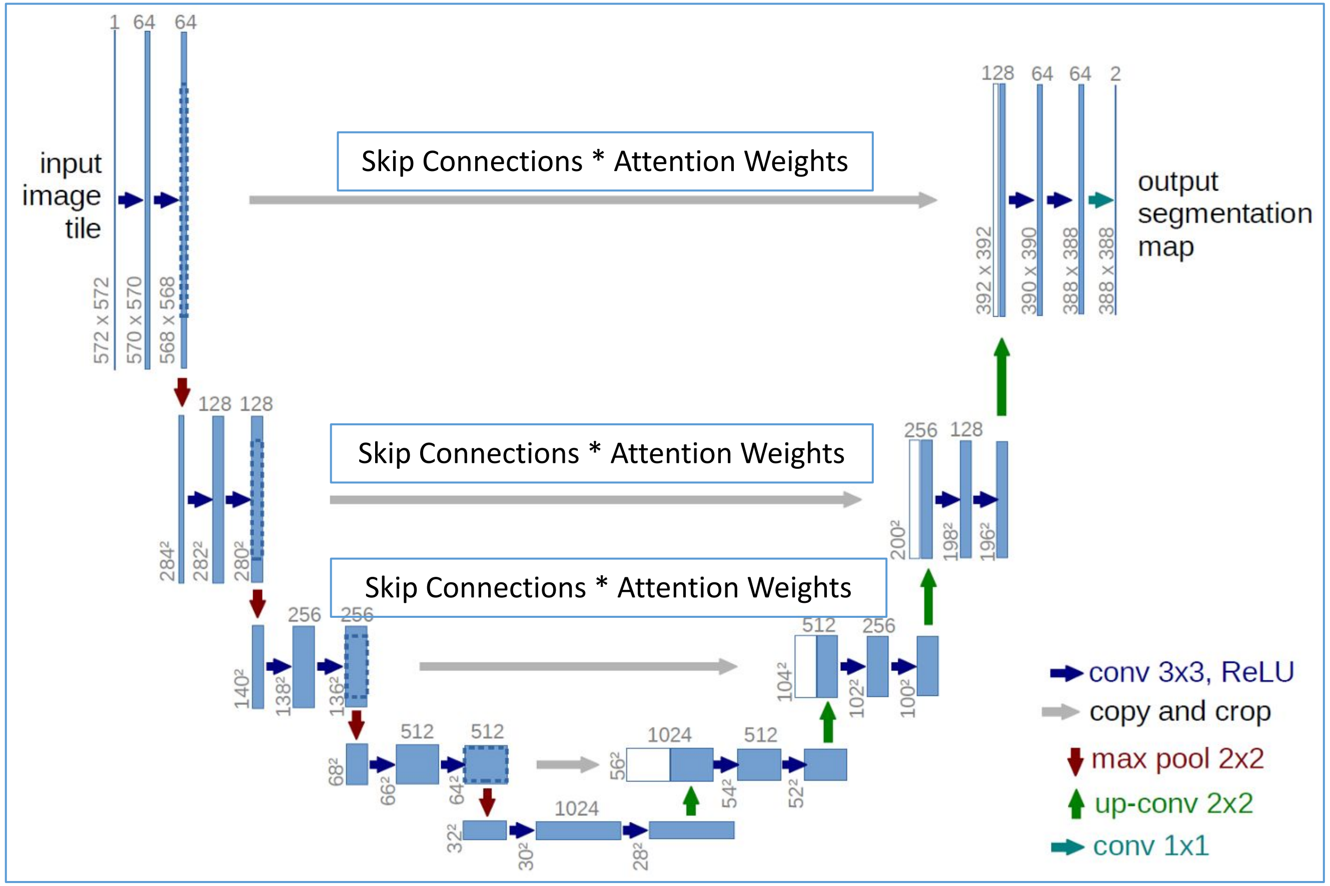
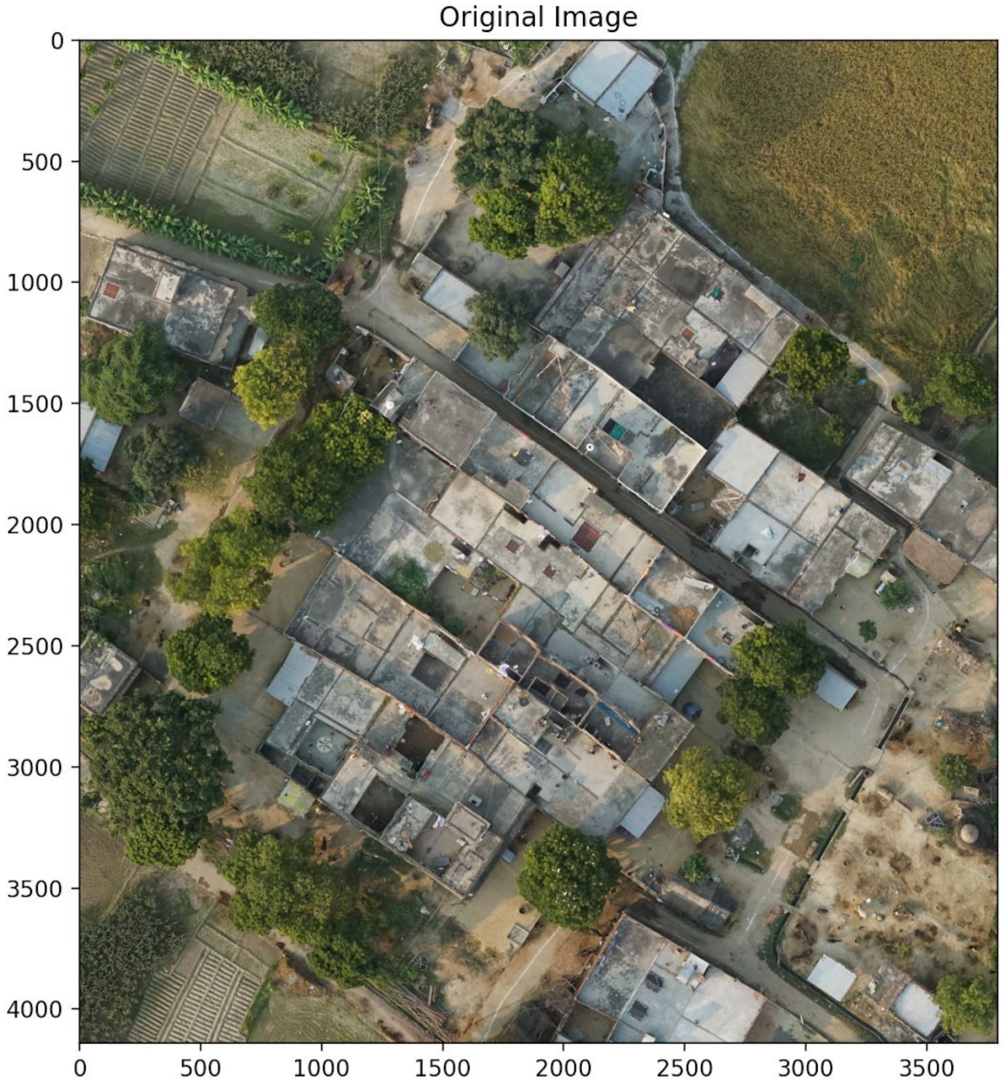
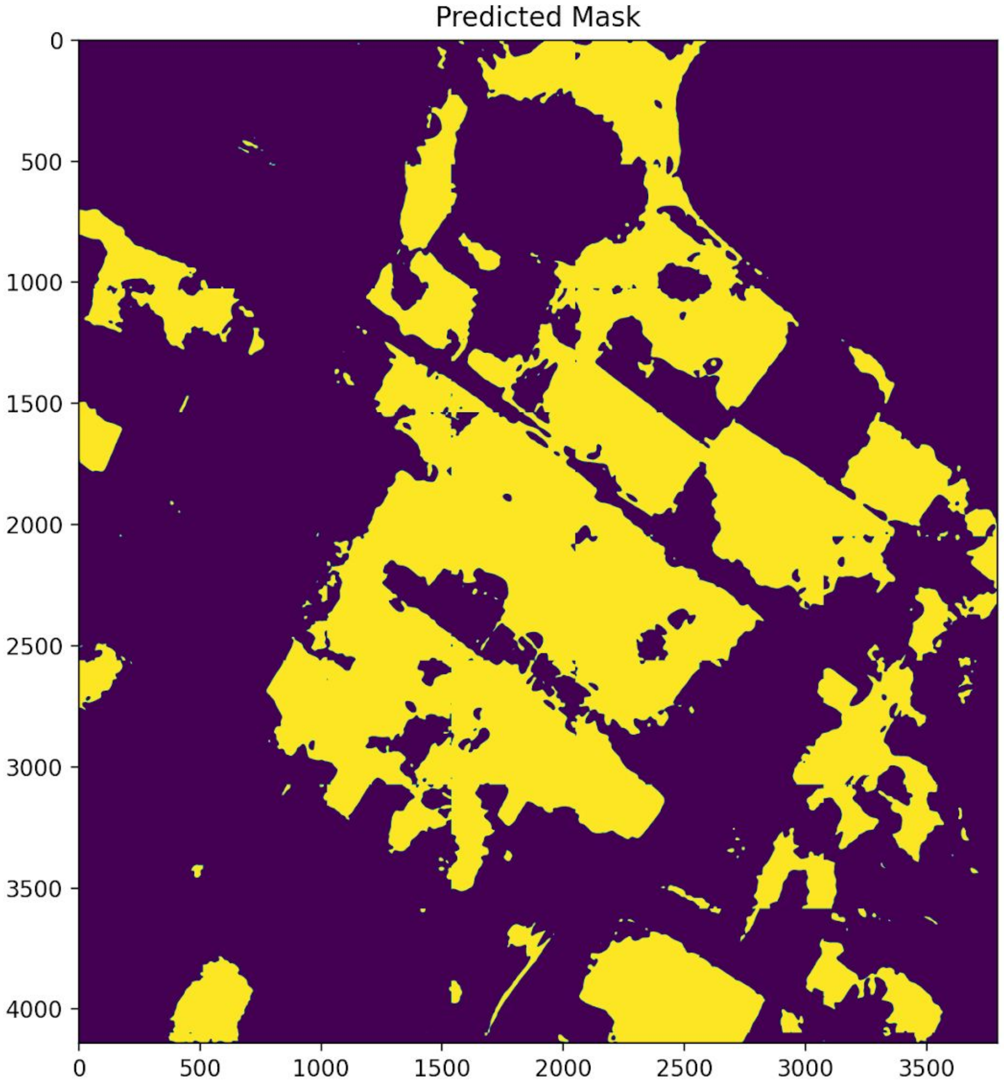


Fig.: Illustration of UNET Architecture with encoder and decoder path

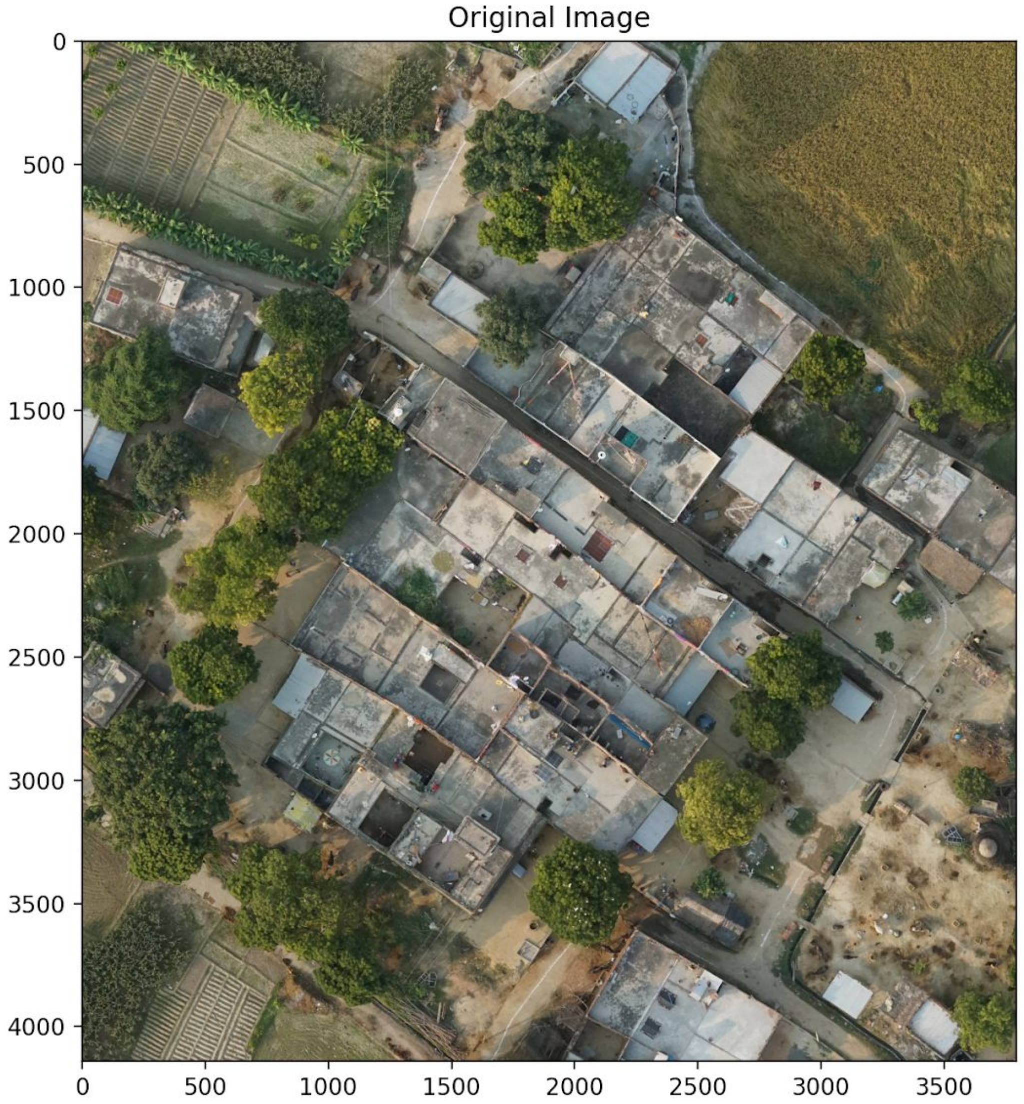
Results

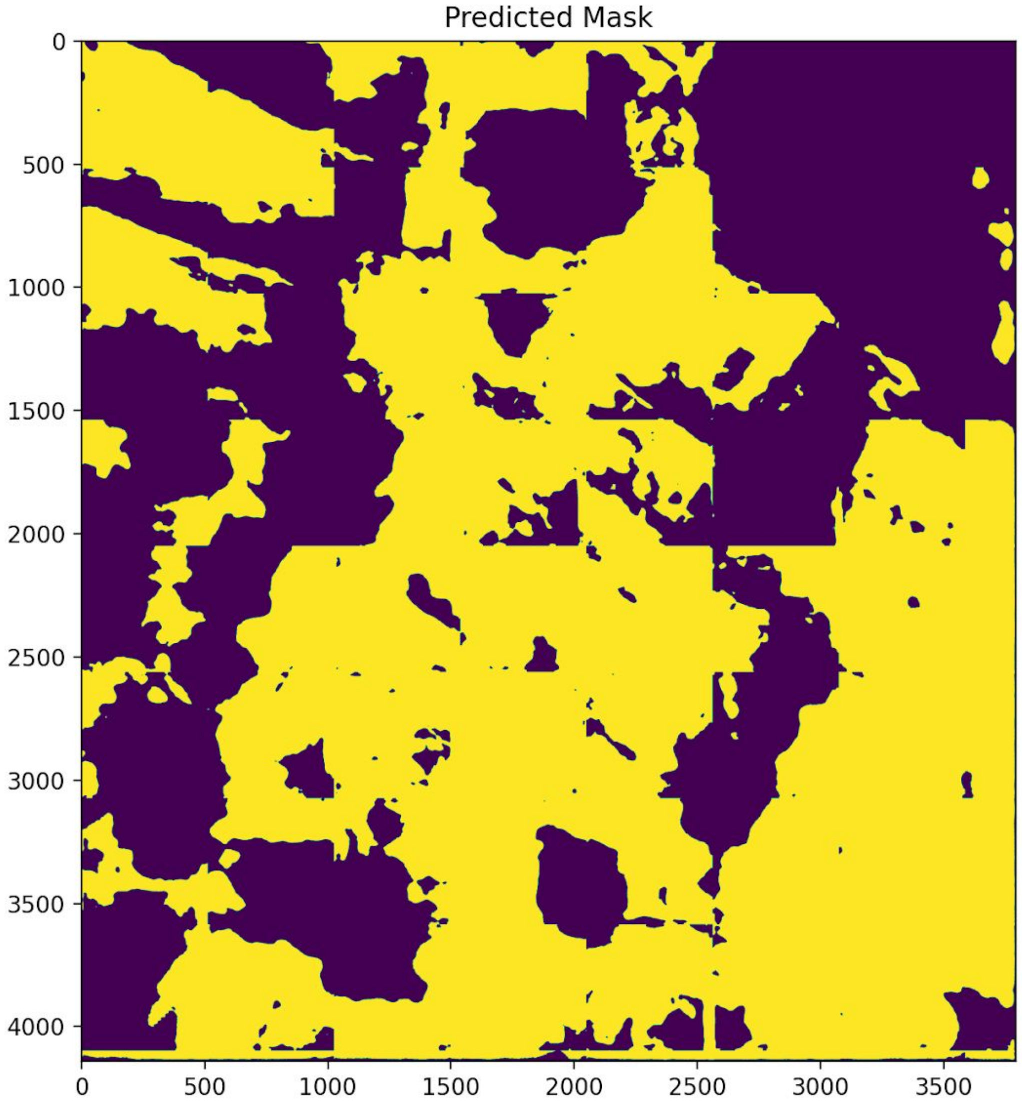




Location: Bibipur
Model: UNET (Depth=4, Filters=32)
Jaccard Coef (Val) in % = 76.26%

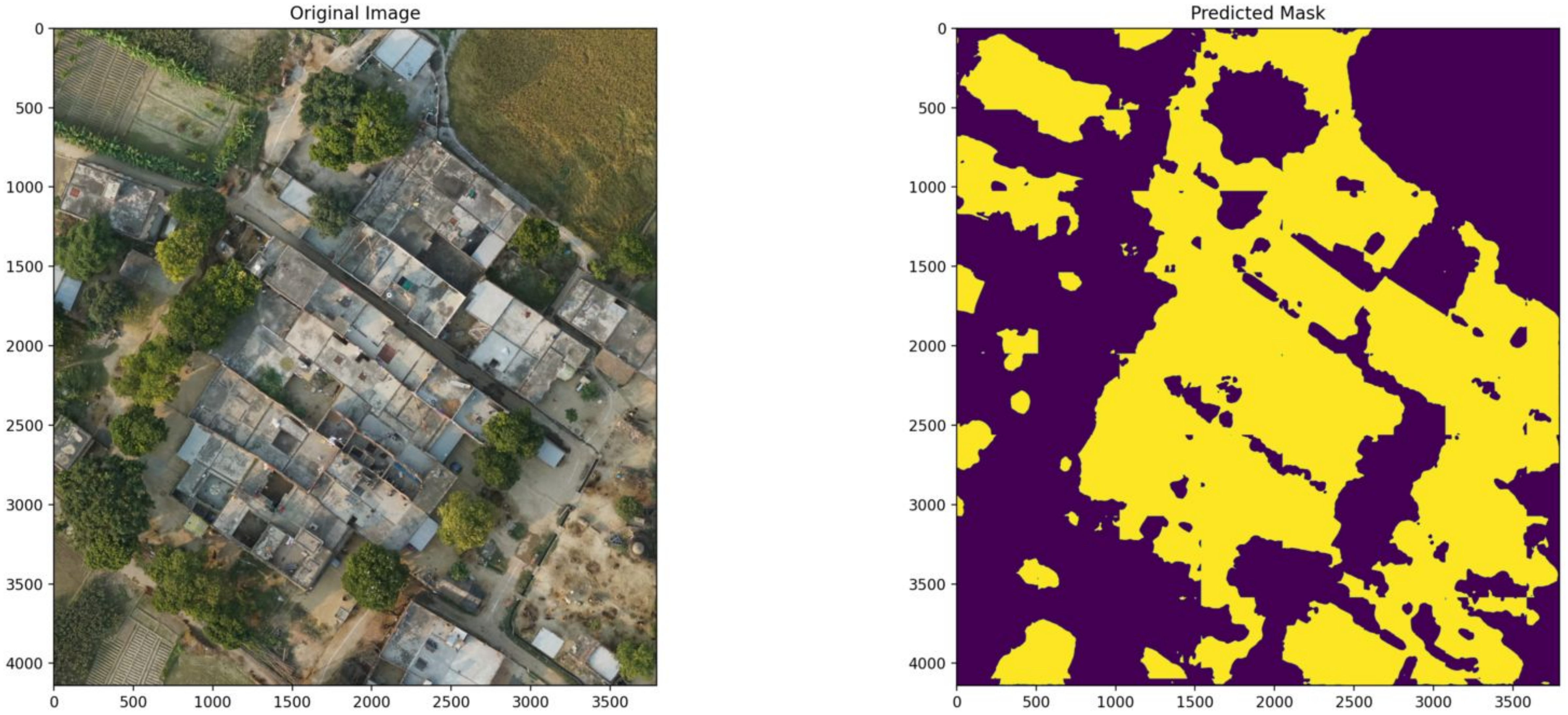
Results





Location: Bibipur
Model: UNET-ASPP (Depth=4, Filters=32)
Jaccard Coef (Val) in % = 73.54%

Results



Location: Bibipur
Model: UNET-Attention (Depth=4, Filters=32)
Jaccard Coef (Val) in % = 73.81%

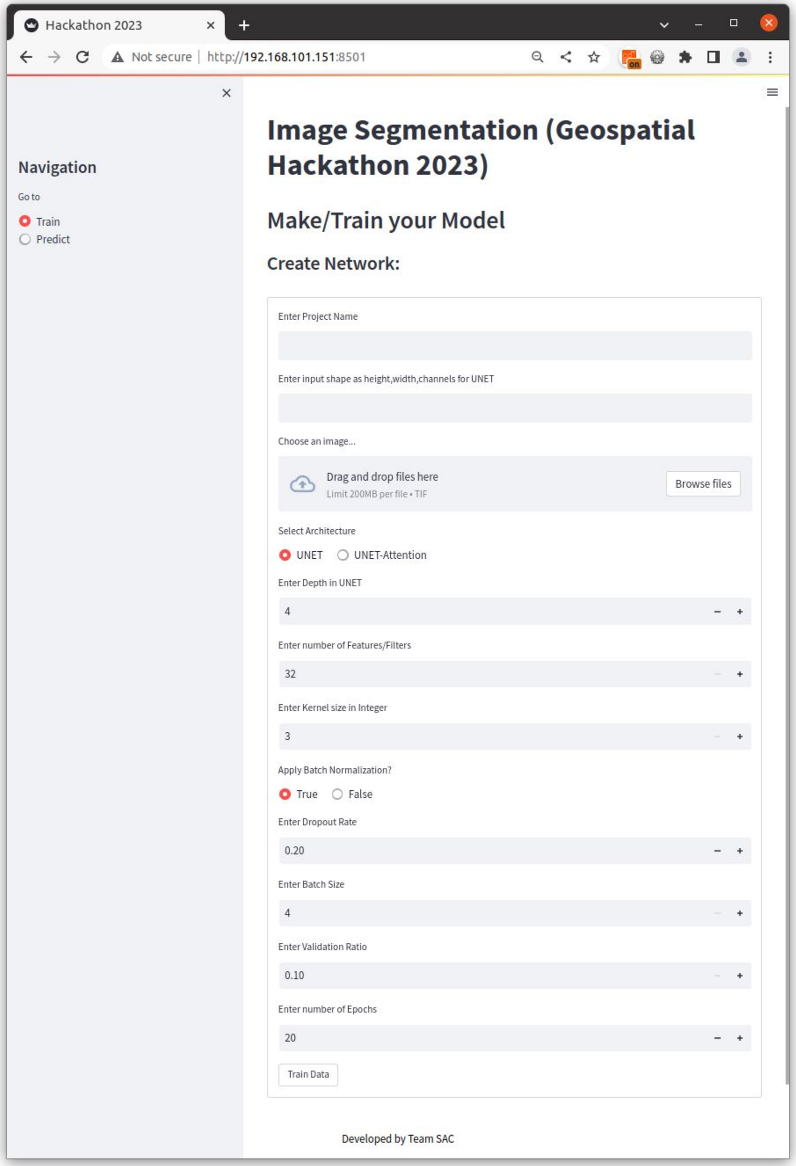
Observations and Discussions in Model Development

- Use of DEM channel improved Jaccard Coefficient (JC) by ~10%.
- Rooftop/Built-Up class is misclassified in ‘road’ and ‘open area’ classes.
  - This could be due to improper labelling of pixels in input mask.
  - Some houses has open roofs exposing underneath floor. Such floors and roads has similar tone and texture which makes it hard to distinguish between them provided RGB and DEM bands.
- The UNET architecture outperformed UNET-Attention architecture in various experiments conducted by varying hyperparameters.

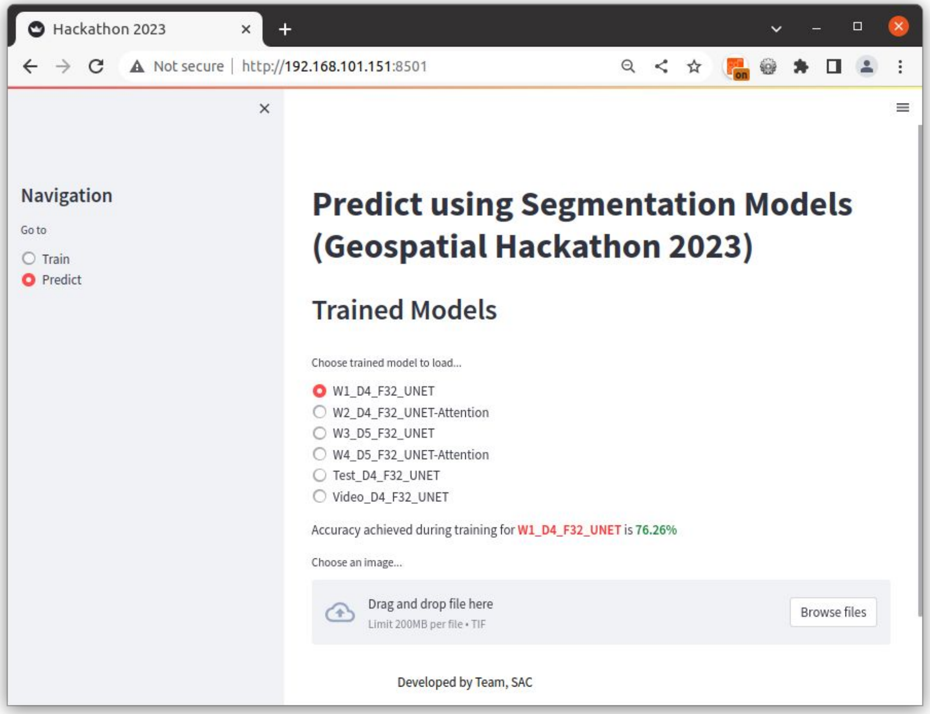
Web Based Workbench

- Allows users to exploit available architectures for various segmentation problems by tuning hyperparameters and uploading data.
- Salient Features:
  - Train
    - Data Preprocessing
      - Automatic Data Normalization
      - Augmentation
      - Image tiling
      - Random data splitting in train & validation sets.
      - Custom data generators
    - Live feedback from training process via plots
    - Callbacks to save model history and training metrics per epoch.
    - Automatic saving best model
  - Prediction
    - Selection of Trained Models with training accuracy.
    - Facility to upload Dataset for on-the-fly prediction.
    - Results visualization

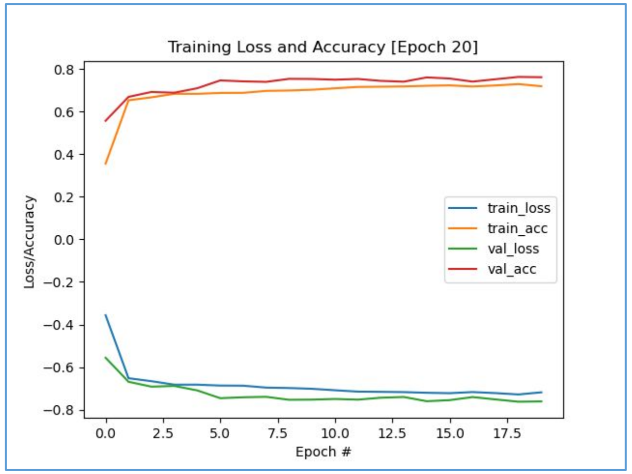
Web Based Workbench



Interface for Training Model



Interface for prediction using saved models



Live feedback of Training process

Demo VideoLink:

Conclusion

- The developed models show satisfactory performance in semantic segmentation problem on SOI high resolution ORI datasets. The highest IoU achieved is 76.26% using UNET architecture.
- Developed web based User Interface (UI) allows SOI users to have abstract controls over segmentation architectures through hyper-parameters tuning in training process and to save the same model for later predictions.
- The web based UI also facilitates prediction using saved models over uploaded image.

!!! Thanks !!!