

Cloud and Shadow Segmentation using Swin UNet

Team Registration ID: NRCC251128

Team Lead: Rachit Gupta

1. Objective

The goal of this project is to develop an automated deep learning-based segmentation pipeline to detect clouds and cloud shadows from satellite imagery. The output must be georeferenced TIFF files that overlay accurately on the original input images. This solution is intended to contribute to real-world applications such as agriculture, forestry, and disaster monitoring.

2. Dataset

- Training Data: 20 zip folders containing RGB satellite images and labeled masks.
- Test Data: 13 folders (TestData-Cloud-Shadow) provided by NRSC.

Preprocessing Steps:

- Extracted RGB BAND.jpg from each training zip folder.
- Converted provided masks into pseudo-color masks.
- Assigned class values:
 - 0: No Cloud
 - 1: Cloud
 - 2: Shadow
- Resized all images and masks to 512×512 for compatibility with Swin UNet.
- Normalized input images with ImageNet mean and standard deviation.

3. Model Architecture

We used Swin UNet, a combination of Swin Transformer and UNet, which captures both global context and fine local details.

- Encoder: Swin Transformer (tu-swin_base_patch4_window7_224)

Cloud and Shadow Segmentation using Swin UNet

- Decoder: UNet-style decoder from segmentation_models_pytorch
- Output: 1-channel mask converted into 3-class final mask via post-processing.

Loss Functions Used:

- Primary: Binary Cross Entropy with Logits (BCEWithLogitsLoss)
- Optional: Combined loss of BCE + Dice for better overlap during training

4. Training Methodology

- Epochs: 35
- Batch Size: 1
- Learning Rate: 1e-4
- Optimizer: AdamW
- Scheduler: ReduceLROnPlateau
- Split: 80% training, 20% validation
- Frameworks: PyTorch, MONAI, TIMM, Torchvision
- Environment: Trained locally using VS Code on CPU
(Macbook Pro M3 Chip)

5. Evaluation Metrics

We evaluated the model using:

- Binary Cross Entropy Loss
- Dice Coefficient
- Intersection over Union (IoU)

Best Metrics Achieved:

- Dice Score: ~0.55
- IoU Score: ~0.42
- Validation Loss: ~0.31

6. Inference and Output Generation

Cloud and Shadow Segmentation using Swin UNet

- Inference was performed on the provided 13 test folders.
- Outputs were thresholded, class-mapped, and resized to original resolution.
- Each folder contains:
 - cloud_mask.tif: 8-bit 3-class georeferenced GeoTIFF mask
 - predicted_mask.png: visualization of prediction
 - .cld and .sdw: placeholder files for standard format compatibility

7. Deliverables Summary

- Report.pdf - This report explaining methodology
- Training_Labeled_Data.zip - RGB inputs and labeled masks
- RegistrationID_Training_Inference.csv - Performance metrics and metadata
- RegistrationID_Inference_code.zip - All scripts (training, inference, utils)
- Model.zip - Trained model weights (.pth)
- Masks.zip - Final test predictions in required format

8. Innovation and Impact

- Used Transformer-based Swin UNet to boost segmentation accuracy.
- Designed a robust local training and prediction pipeline from scratch.
- Contributed toward automation in cloud and shadow detection for scalable EO applications.