Ahsanullah University of Science and Technology

Department of Computer Science and Engineering

Program: Bachelor of Science in Computer Science and Engineering

Course No: CSE 4108

Course Title: Artificial Intelligence Lab

Project Name: **Satellite Image Classification**

Date of Submission: 6 September 2022

Submitted to:

Ms. Tamanna Tabassum
Lecturer, Department of CSE, AUST.

Mr. Md. Siam Ansary
Lecturer, Department of CSE, AUST.

Submitted by:

Aurnob Sarker Student ID: 18.02.04.137

S. M. Tasnimul Hasan Student ID: 18.02.04.142

Introduction

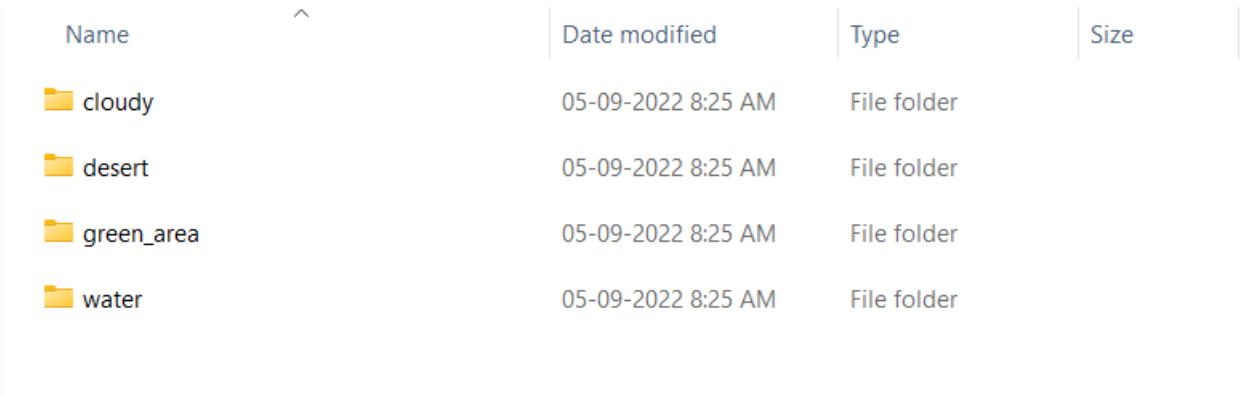
We are trying to predict the classes of Satellite Images using Convolutional Neural Network (CNN) and Transformer. This will help to classify the images of different categories taken by satellite. The proposed system used Resnet-34, VGG-16, DenseNet 161, and Swin Transformer Small.

Dataset

The whole dataset has 5631 images in jpg format. There are 4 classes (Cloudy, Dessert, Green Area, and Water). 80% of instances are used as train data and 20% of instances are used to test the models. For testing, each class contains 75 images.

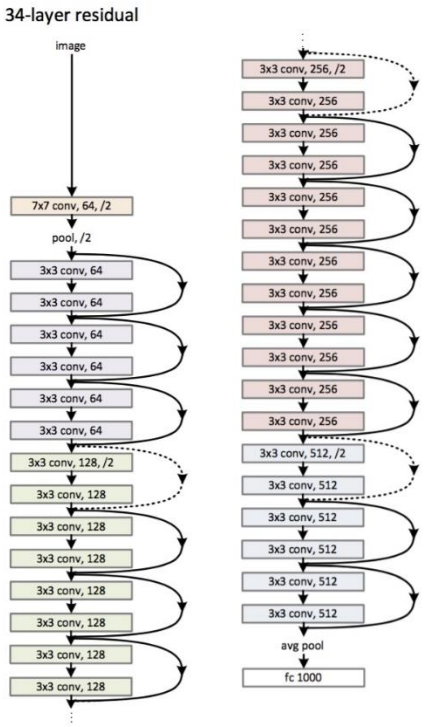
Dataset link: https://www.kaggle.com/datasets/mahmoudreda55/satellite-image-classification

Here is the overview of our dataset:

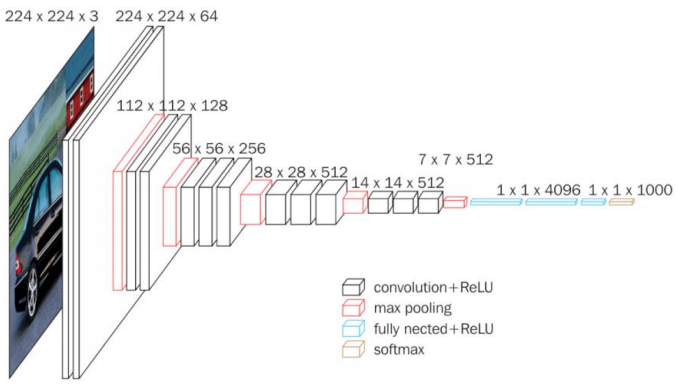


CNN Models

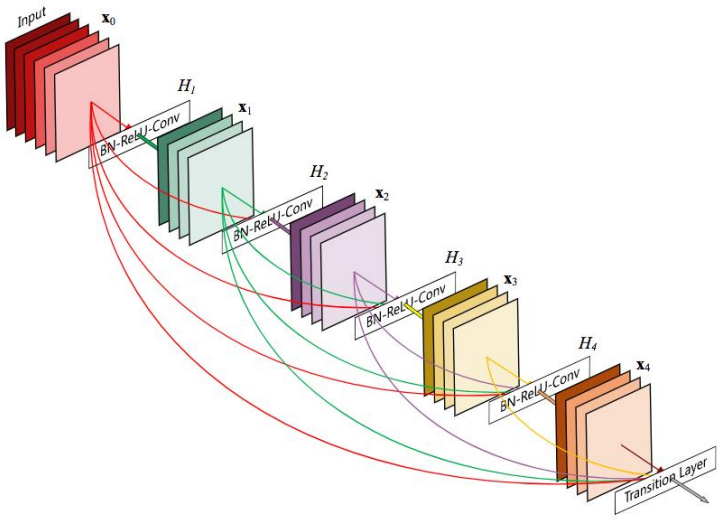
- **Resnet-34:**
  Resnet 34 is a 34-layer convolutional neural network that can be utilized as a state-of-the-art image classification model.



- **VGG-16:**
  VGG stands for Visual Geometry Group. It is a standard deep Convolutional Neural Network (CNN) architecture with multiple layers. The “deep” refers to the number of layers. VGG-16 consists of 16 convolutional layers.



- **DenseNet-161:**
  DenseNet is a new CNN architecture. DenseNet is composed of Dense blocks. In those blocks, the layers are densely connected together: Each layer receives in input all previous layers' output feature maps.



- **Swin Transformer:**
  Swin Transformer is a transformer-based deep learning model with state-of-the-art performance in vision tasks. It builds hierarchical feature maps by merging image patches (shown in gray) in deeper layers and has linear computation complexity to input image size due to computation of self-attention only within each local window (shown in red).

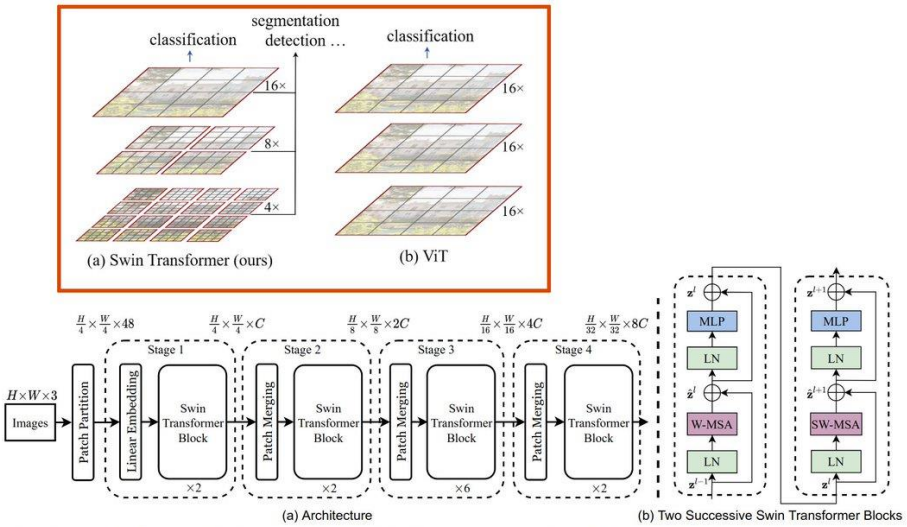


Figure 3. (a) The architecture of a Swin Transformer (Swin-T); (b) two successive Swin Transformer Blocks (notation presented with Eq. (3)). W-MSA and SW-MSA are multi-head self attention modules with regular and shifted windowing configurations, respectively.

Comparison of Performance Scores:

Name of Architecture	Precision	Recall	F1 Score
Resnet-34	0.92	0.92	0.92
VGG-16	0.93	0.93	0.93
DenseNet-161	0.91	0.90	0.90
Swin Transformer	0.88	0.86	0.85

Decision:

From the comparison of performance scores, it is seen that the Precision, Recall, and F1 Score is highest in VGG-16 architecture and the second highest in Resnet-34 architecture. In comparison to the other architecture, it can be said that VGG-16 has shown better precision, recall & f1 score. Swin Transformer got the lowest accuracy of all the architecture we have used.