A multimodal semantic segmentation for airport runway delineation in panchromatic remote sensing images

Rajeshreddy Datla [*,†], Chalavadi Vishnu [†], C Krishna Mohan [†]
[*]Advanced Data Processing Research Institute (ADRIN), Secunderabad, India
[†]Department of Computer Science and Engineering, Indian Institute of Technology Hyderabad, India

ABSTRACT

Monitoring airport runways in panchromatic remote sensing images is helpful for both civil and strategic communities in effective utilization of the large-area acquisitions. This paper proposes a novel multimodal semantic segmentation approach for effective delineation of the runways in panchromatic remote sensing images. The proposed approach aims to learn complementary information from two modalities, namely, panchromatic image and digital elevation model (DEM) to obtain discriminative features of the runway. The fusion of image features and the corresponding terrain information is performed by stacking the image and DEM by leveraging the merits of both Transformers and U-Net architecture. We perform the experiments on Cartosat-1 panchromatic satellite images with the corresponding Cartosat-1 DEM scenes. The experimental results demonstrate a significant contribution of terrain information to the segmentation process in achieving the contours of airport runways effectively.

Keywords: Airport runway, remote sensing images, multimodal segmentation, digital elevation model, Cartosat-1.

1. INTRODUCTION

Remote sensing technology becomes the prominent platform for Earth observation focusing on various aspects, ranging from scene comprehension, the spatial arrangement of different objects in a region, characterization of the terrain, etc. Typically, an acquisition from such platforms can merely observe a few of the aforementioned characteristics. The physical quantity measured in panchromatic image (PAN) is the brightness of the apparent target and therefore spectral information of the targets is lost. However, with the prevailing advantages of higher spatial resolution and large-area coverages, the use of PAN images has been extensively increased in earth observation domain. Monitoring the runways of airports is drawn significant attention of both civilian and strategic communities in order to keep track their changes regularly. These changes may include new runway or extension to an existing runway. Especially, early findings of newly constructed airports or military airbases are of much important to strategic community, because they cannot be obtained easily from public platforms (e.g., maps). Typically, the length and width of an airport runway varies from 245m to 5.5km and 8m to 80m, respectively. The digital elevation model (DEM) to an extent of airport runway is obtained by performing seamless mosaic of individual Cartosat-1 DEM scenes [1]. Similarly, the corresponding PAN image is created by a suitable mosaic operation [18]. In general, the runways are categorized into four types, namely, single, parallel, intersecting, and open-V runways. Understanding the changes in these kinds of runways from PAN images becomes tedious and challenging due to lack of spectral information and the presence of visually similar infrastructures. Further, the appearance of runways in high resolution panchromatic remote sensing images varies across the airports, as the materials used in the construction may influence the intensity values of the runway. Also, the intensity values within a runway vary due to renovation of some parts or extension of the runway. Thus the intensity based threshold methods are not suitable for effective identification of runways.

The airport runways are generally not flat and no two airport runways are same, though they may look visually similar. The terrain information is one of the key characteristics which will be considered in designing the airport runways. A gradual increase or decrease of slope from one end of runway to other helps to drain off the water during rain. The slope or gradient of a runway is the measure of change in runway height over the full length of the runway, usually expressed in terms of percentage. Also, to counteract the tailwind influence on landing, the runway with 3% up-slope is recommended. A 3% slope represents a 3 feet height change in a runway of length 100 feet. These characteristics are usually embedded in the terrain information which would be of significant importance in the runway recognition process. Hence, an effective understanding of airport runway in PAN images necessitates the incorporation of multiple modalities that serve as complementary information over the same scene.

In literature, modalities from different sensors such as RGB-D, 3D LiDAR, and thermal data which provide complementary information are explored in understanding the complex scenes [2]. Also, the use of multiple modalities in comparison to single modality has shown significant improvement in the performance of model learning [3, 4, 5]. This motivates us to investigate the performance of airport runway segmentation by employing the terrain information as another modality to the corresponding PAN image. Our proposed method employs both U-Net and Transformers to exercise the innate self-attention capability of Transformers and also by utilizing the CNN features captured through U-Net. The main contributions of this paper are:

- Fusion of terrain information and panchromatic satellite image to capture the semantics of airport runway.
- We model the long-range spatial dependency of DEM and panchromatic image in segmenting the runway by leveraging Transformers and U-Net architecture.
- Efficacy of the proposed multimodal semantic segmentation is evaluated on Cartosat-1 PAN images with the terrain information from Cartosat-1 DEM

The rest of the paper is organized as follows. Sections 2 presents a brief review on airport runway detection and segmentation methods. We describe the proposed approach in Section 3. The experimental results and their analysis are provided in Section 4. Finally, Section 5 concludes this paper.

2. RELATED WORK

This section summarizes the existing works on airport runway detection and segmentation. A texture-based method employed Adaboost algorithm to select best discriminative features from 137 texture features to obtain a coarser representation of airport runway [6]. A semi-automatic approach [7] for airport runway extraction from Google earth images integrates a long straight line finder with a region-based level set evolution (LSE). Initially, long straight boundaries are detected from the images using long straight line finder. Then, initial level curves for LSE are generated semi-automatically using the detected boundaries. These initial level curves follow LSE to obtain the boundaries that represent airport runway. Another runway detection method [8] detects centre-line markings to roughly locate runways in the image using Canny edge map and Hough transforms.

A two-stage approach for airport runway detection [9] is accomplished firstly by classifying the aerial image based on the existence of runway and later by performing localization using both conventional and deep learning methods for line detection. Recently, a framework based on DeepLabv3 semantic segmentation [10] is used to extract the contour of airport runway from multi-spectral remote sensing images. In [11], markings of chevron and runway edge are combined along with the runway length and width to extract airport runway. The existing works explored various approaches based on traditional handcrafted features and deep learning models with more emphasis on airport runway detection on multispectral remote sensing images rather than runway segmentation. In this work, we propose a novel multimodal semantic segmentation approach by combining terrain information and panchromatic remote sensing images.

3. PROPOSED METHOD

In this section, we present an end-to-end semantic segmentation learning model for airport runway delineation which combines terrain information along with panchromatic satellite image by leveraging both Transformers [12] and U-Net architecture [13]. Fig.1 shows the block diagram of the proposed method which is described in the following subsections.

3.1 Encoding with Transformers

For a given image, $\boldsymbol{x} \in \mathbb{R}^{H \times W \times C}$, we obtain a sequence of $\left(N = \frac{H \times W}{P^2}\right)$ flattened 2D patches denoted by $\boldsymbol{x}_p^i \in \mathbb{R}^{P^2 \cdot C}$ for $i = 1, \cdots,$ N, where H & W are height and width of image, C & P represent the number of channels and image patch size, respectively. The vectorized image patches $\boldsymbol{x}_p$ are mapped into K-dimensional embeddings with the help of linear projection. These patch embeddings in conjunction with the learnt position-specific embeddings encode the patch-wise spatial information by retaining the positional information using

$$\boldsymbol{z}_0 = [\boldsymbol{x}_p^1 \psi; \boldsymbol{x}_p^2 \psi; \cdots; \boldsymbol{x}_p^N \psi] + \xi \tag{1}$$

Where $\psi \in \mathbb{R}^{(p^2 \cdot C) \times K}$ and $\xi \in \mathbb{R}^{N \times K}$ represent the patch and position embedding projections, respectively. The output of m^{th} layer in M layers of multi-head self-attention (MSA) and multi-layer perceptron (MLP) blocks, as shown in Fig. 1, is given by

$$z'_m = MSA\big(\alpha(z_{m-1})\big) + z_{m-1}\,, \tag{2}$$

$$z_{m_1} = MSA\big(\alpha(z'_m)\big) + z'_m\,, \tag{3}$$

where $\alpha(\cdot)$ denotes layer-normalization operator. And the image representation is encoded as z_M.

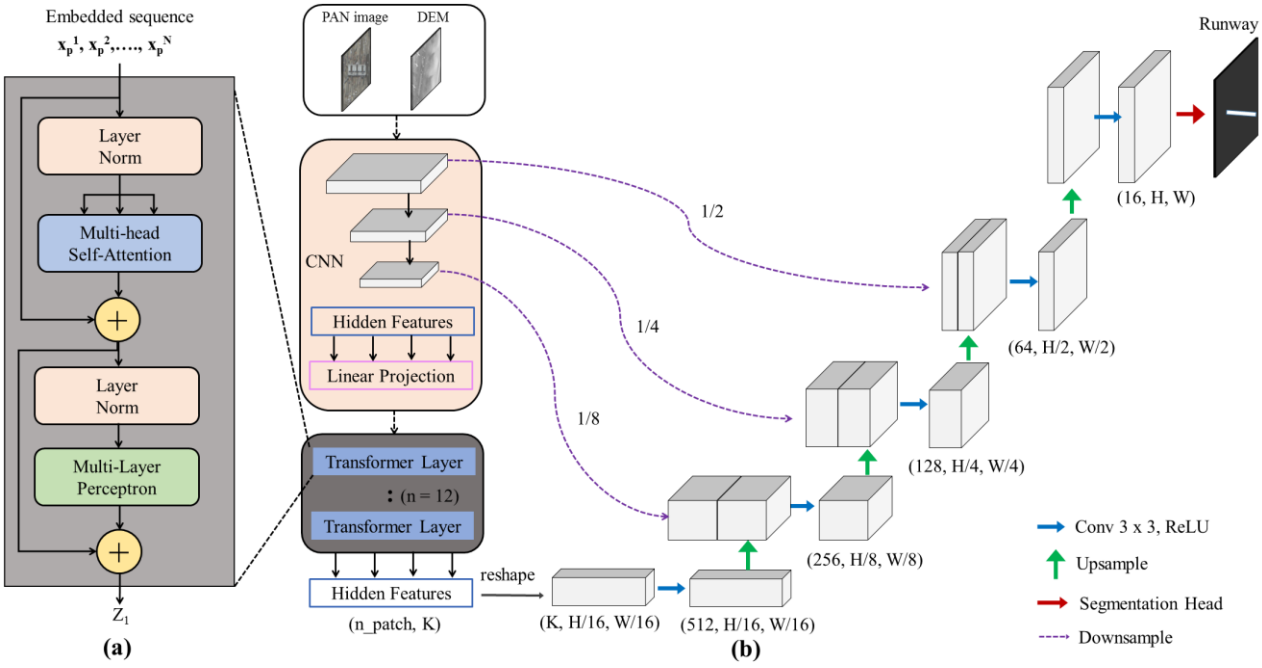


Figure 1. Block diagram of the proposed approach for airport runway segmentation from panchromatic remote sensing. (a) Illustration of Transformer layer; (b) Leveraging U-Net architecture for multimodal segmentation

3.2 Transformers with U-Net

In order to obtain the dense predicted segmentation output, the encoded representation $z_M = \mathbb{R}^{\frac{H}{P^2}\frac{W}{P^2} \times K}$ is upsampled to original resolution of the image. The spatial order is recovered by resizing the encoded features from $\frac{H}{P^2}\frac{W}{P^2}$ to $\frac{H}{P} \times \frac{W}{P}$. Further, the number of channel of the resized encoded features is reduced to the number of classes using 1×1 convolution. The segmentation output at full resolution is finally obtained by employing bilinear upsampling to the feature map. Though Transformers with upsampling produces a reasonable segmentation, this would not compensate the loss of details at low-level such as contour of the airport runway due to $\frac{H}{P} \times \frac{W}{P} \ll H \times W$. So, our proposed method involves a hybrid CNN-Transformer that acts as both encoder and cascaded upsampler to achieve precise localization.

This hybrid CNN-Transformer model first extracts features for the input and generates a feature map. Then, patch embedding is perfromed over a set of 1×1 patches, which are extracted from the generated feature map instead of raw images. This methodology improves the decoding path due to the incorporation of CNN feature maps with high resolution. We also employ cascaded upsampler that facilitates multiple upsamplings in order to produce a final segmented contour by decoding the hidden features. After obtaining the reshaped hidden features $\frac{H}{P} \times \frac{W}{P} \times K$ from $z_M = \mathbb{R}^{\frac{H}{P^2}\frac{W}{P^2} \times K}$, multiple blocks of upsampling are cascaded to achieve full resolution $H \times W$ from $\frac{H}{P} \times \frac{W}{P}$. Each block

comprises of a $2\times$upsampling, a 3×3 convolutional layer with ReLU activation function. It can be seen from Fig. 1 that the U-Net with Transformer enables to aggregate the features at multiple resolutions through skip-connections.

3.3 Post-processing

The airport runways are often confused with road highways due to their look-alike features. Hence, the road network information from open street map (OSM) layers is used to suppress the false positives.

4. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed multimodal segmentation on Cartosat-1 panchromatic remote sensing images and the corresponding Cartosat-1 DEM scenes.

4.1 Datasets and experimental settings

Fig. 2 shows some samples of the PAN image and the corresponding DEM data used in the experiments and the details of these datasets are provided in Table 1. Total 150 scenes of Cartosat-1 PAN images (2.5m) and the corresponding scenes of Cartosat-1 digital elevation model (DEM) (10m) containing large and medium airports are considered in this study. The panchromatic images (2.5m) are downsampled 4 times to the resolution of DEMs using bilinear interpolation. Then, we apply four kinds of data augmentation techniques such as horizontal flipping, 10 random crops, rotation by 90, 180, 270 degrees resulting into 2100 images and 2100 DEMs. We split both these datasets with 80% - 20% train-test ratios resulting into 1680 training samples and 420 test samples.

Table 1. Datasets used in semantic segmentation experiments

Dataset	Scene size	Spatial resolution (m)
Cartosat-1 PAN image [14]	512×512	2.5
Cartosat-1 DEM [14]	512×512	10

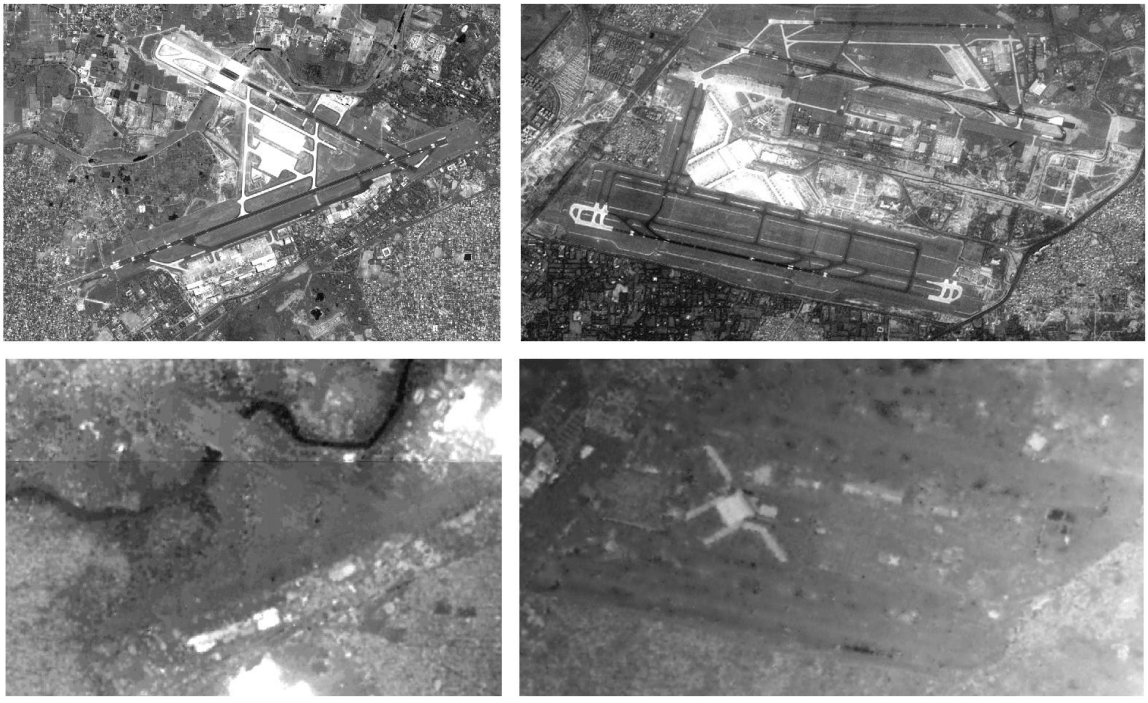


Figure 2. Samples of PAN image (Row 1) and the corresponding DEM scenes (Row 2) of Cartosat-1 satellite.

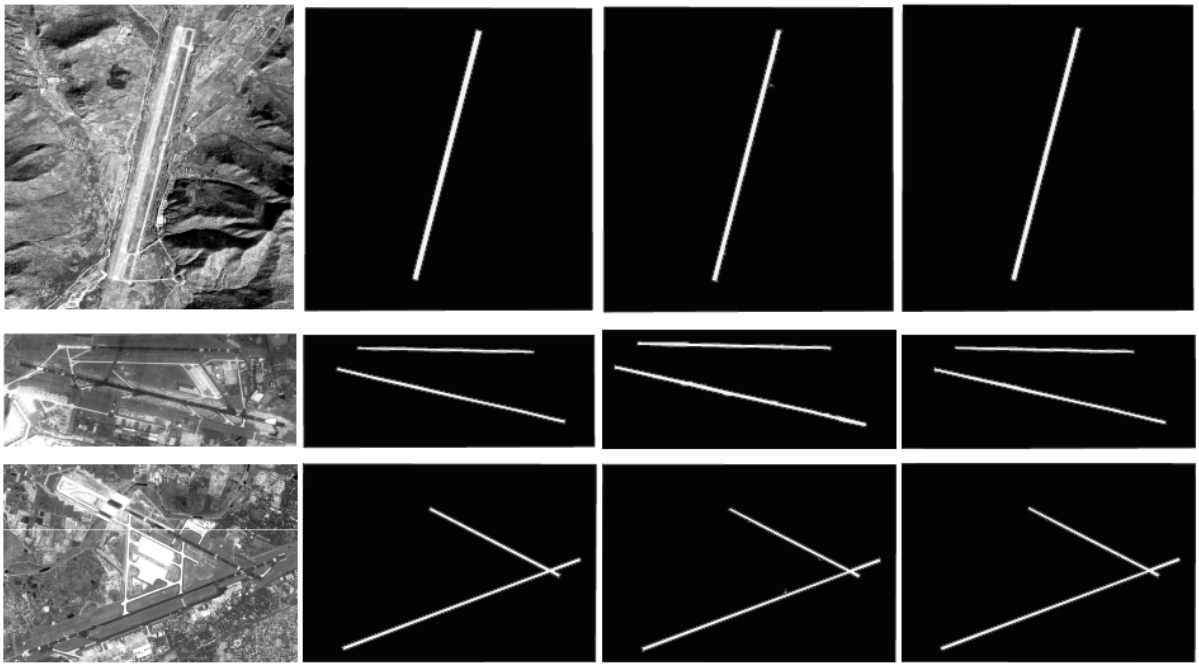
We use ViT [15] as encoder containing 12 Transformers. The hybrid encoder is designed by combining ResNet-50 and ViT. Our network provides promising results by training for 75 epochs comprising of 210 iterations with batch size of 8 samples using standard stochastic gradient descent optimization, by setting learning rate to 0.01, momentum to 0.9 and weight decay to $1e^{-4}$.

4.2 Results and analysis of the proposed method

Table 2 provides the performance comparison of the proposed method with state-of-the-art segmentation methods. It can be observed that our proposed model achieves better performance in segmenting the airport runways. Also, it can be observed from Table 2 that the proposed method achieves significant improvement in the mean intersection over-union (mIoU) and F1-score. This signifies the capability of Transformers as effective encoders with U-Net for segmentation task in remote sensing images. Fig. 3 shows the segmentation results on different types of runways. A typical panchromatic satellite image is considered to be a large-scale image as it covers a region with minimum of 10 km swath

Table 2. Performance comparison with different semantic segmentation methods

Methods	PAN		PAN + DEM	
	mIoU (%)	F1-score	mIoU (%)	F1-score
PSPNet [16]	73.54	0.854	76.49	0.861
U-Net [13]	76.29	0.892	78.93	0.903
DeepLabV3 [17]	79.65	0.923	81.14	0.934
Proposed method	83.62	0.952	90.72	0.982



(a) Cartosat-1 PAN image (b) Ground truth (c) Ours (PAN only) (d) Ours (PAN+DEM)

Figure 3. Segmentation outputs from Cartosat-1 data. Row 1: single runway; Row 2: open-v runway; Row 3: intersecting runway.

and 30km length on the ground. An image with ground sampling distance (GSD) of 0.6m for an area 300 sq.km is depicted by 800 mega pixels, which occupies approximately 1.5 GB space. Therefore, we partition the PAN image into

scenes with resolution 512 × 512 with patch size P as 16, as inferencing is not viable from large-scale images directly due to computational limitations. Our proposed method also generalizes well on the partial airport runways which are resulted from partitioning with the inclusion of random cropping in the training data.

4.3 Discussion

The experiments are conducted by downsampling Cartosat-1 panchromatic images (2.5m) to 10m for compliance with digital elevation model (DEM), which is of 10m grid spacing derived from Cartosat-1 stereo-pairs. Thus the features at the original resolution of image are not utilized effectively. However, the performance of runway segmentation can be further improved by employing the finer terrain information to the level of details present in panchromatic remote sensing images.

5. CONCLUSION

In this work, we presented a multimodal semantic segmentation for airport runway delineation from panchromatic remote sensing images. The proposed method employs Transformers and U-Net for effective segmentation by considering two inputs, i.e., PAN image and terrain data as grey-scale images. The joint information from these two-channels helps to learn the spatial and terrain patterns effectively in order to delineate the contour of the airport runway. The proposed method is demonstrated on both medium and large airports from Cartosat-1 PAN images. The significance of Transformers as effective encoders with U-Net is demonstrated for segmentation task in remote sensing images. Our experiments have also shown that the additional terrain information to the PAN image significantly improved the performance of runway segmentation. In future, the proposed method would be extended to small airports with the help of very high resolution DEMs.

REFERENCES

[1] Rajeshreddy Datla and C. Krishna Mohan, "A novel framework for seamless mosaic of Cartosat-1 DEM scenes," Computers & Geosciences, vol. 146, pp.104619, (2021).

[2] Yifei Zhang, Désiré Sidibé, Olivier Morel, Fabrice Mériaudeau, Deep multimodal fusion for semantic image segmentation: A survey, Image and Vision Computing, Volume 105, (2021).

[3] X. Xu, Y. Li, Gangshan Wu, and Jiebo Luo, "Multimodal deep feature learning for RGB-D object detection," Pattern Recognition, vol. 72, pp. 300–313, (2017).

[4] A. Asvadi, L. Garrote, C. Premebida, P. Peixoto, and U. Nunes, "Multimodal vehicle detection: fusing 3D-LIDAR and color camera data," Pattern Recognition Letters, vol. 115, pp. 20–29, (2018).

[5] ZoltanKoppanyi, DorotaIwaszczuk, BingZha, Can JozefSaul, Charles K.Toth, and AlperYilmaz, "Chapter 3 - multimodal semantic segmentation: Fusion of RGB and depth data in convolutional neural networks," in Multimodal Scene Understanding, Michael Ying Yang, Bodo Rosenhahn, and Vittorio Murino, Eds., pp. 41–64, Academic Press, (2019).

[6] Ö. Aytekin, U. Zöngür and U. Halici, "Texture-Based Airport Runway Detection," in IEEE Geoscience and Remote Sensing Letters, vol. 10, no. 3, pp. 471-475, (2013).

[7] Z. Li, Z. Liu, and W. Shi, "Semiautomatic airport runway extraction using a line-finder-aided level set evolution," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 7, pp. 4738–4749, (2014).

[8] Philip T. G. Jackson, Carl J. Nelson, J. Schiefele, and B. Obara, "Runway detection in high resolution remote sensing data," 2015 9th International Symposium on Image and Signal Processing and Analysis (ISPA), pp. 170–175, (2015).

[9] Javeria Akbar, Muhammad Shahzad, Muhammad Farooq Malik, A. Ul-Hasan, and Fasial Shafait, "Runway detection and localization in aerial images using deep learning," 2019 Digital Image Computing: Techniques and Applications (DICTA), pp. 1–8, (2019).

[10] Z. Men, J. Jiang, Xian Guo, L. Chen, and D. Liu, "Airport runway semantic segmentation based on DCNN in high spatial resolution remote sensing images," ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLII-3/W10, pp. 361–366, (2020).

[11] Wei Ding and JidongWu, "An airport knowledge-based method for accurate change analysis of airport unways in VHR remote sensing images," Remote Sensing, vol. 12, pp. 3163, (2020).

[12] A Vaswani, N Shazeer, N Parmar, J Uszkoreit, L Jones, A.N. Gomez, and Kaiser, "Attention is all you need," In: Advances in neural information processing systems, pp. 5998–6008, (2017).

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," ArXiv, vol. abs/1505.04597, (2015).

[14] R.K. Jaiswal and S. Bhatawdekar, "1.10 - indian earth observation program," in Comprehensive Remote Sensing, Shunlin Liang, Ed., pp. 280–298. Elsevier, Oxford, (2018).

[15] A. Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, M. Dehghani, Matthias Minderer, Georg Heigold, S. Gelly, Jakob Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," ArXiv, vol. abs/2010.11929, (2020).

[16] Hengshuang Zhao, J. Shi, Xiaojuan Qi, XiaogangWang, and J. Jia, "Pyramid scene parsing network," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6230–6239, (2017).

[17] Liang-Chieh Chen, G. Papandreou, Florian Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," ArXiv, vol. abs/1706.05587, (2017).

[18] Debabrata Ghosh, Naima Kaabouch, "A survey on image mosaicing techniques," Journal of Visual Communication and Image Representation, Volume 34, Pages 1-11, (2016).

AUTHORS' BACKGROUND

Your Name	Title*	Research Field	Personal website
Rajeshreddy Datla	Scientist-SE/ External PhD candidate	Remote sensing imagery analysis, computer vision, machine learning	
Chalavadi Vishnu	PhD candidate	Aerial and drone imagery analysis, machine learning	
C Krishna Mohan	Full Professor	Pattern Recognition, machine learning	https://www.iith.ac.in/~ckm/