

# LAND COVERAGE ANALYSIS OF PAKISTAN USING SATELLITE IMAGERY

A. Sabir<sup>1,\*</sup>, M. Jameela<sup>2</sup>, A. W. Malik<sup>1</sup>

<sup>1</sup> NUST, SEecs, Department of Computing, H-12 Islamabad, Pakistan - (asabir.mscs19seecs, asad.malik)@seecs.edu.pk

<sup>2</sup> York University, Department of Earth and Space Science and Engineering, Canada - (maryumja)@yorku.ca

**KEY WORDS:** Climate Change, Satellite Imagery, Semantic Segmentation, Land Coverage, Deep Learning, Computer Vision

## ABSTRACT:

Pakistan has a unique landscape geographically due to its strategic geo-political importance. It has played a vital role in global climate and politics. There are various semantic segmentation studies performed on remote sensing high-resolution imagery of various urban and rural areas into major classes of buildings, vegetation, water, and roads. These analyses have supported the land coverage study, which can facilitate urban infrastructure management, forestry, disaster management, and climate challenges. Recent climate reports have confirmed the importance of these studies, especially for Pakistan. It's a critical location for the global south to observe the climate catastrophe. This research will focus on three major cities of Islamabad, Karachi, and Quetta and semantically segment the satellite imagery to study the land coverage. Our research contributes the dataset from major cities of Pakistan and compare the performance of state-of-the-art semantic segmentation networks to evaluate the dataset. Benchmark can help in selecting a highly effective deep learning network and generalizing those networks on our prepared dataset. Dataset can be downloaded from here: <https://github.com/Abdullah-Sabir/Pakistan-Land-Coverage-Analysis-Dataset>

## 1. INTRODUCTION

Traditionally, land coverage analysis has been used to study a country's terrain, providing a detailed view for urban infrastructure management, forestry, disaster management, and climate challenges. While technology has made data acquisition faster through remote sensing methods such as airborne LiDAR and satellite imagery, the post-data acquisition analysis remains slow and labor-intensive. However, recent advancements in deep learning and computer vision have created opportunities to automate such analysis. State-of-the-art (SOTA) 2D semantic segmentation networks, such as UNet (Ronneberger et al., 2015a), UNet++ (Zhou et al., 2018b), FPN (Lin et al., 2016), and DeepLab (Chen et al., 2017b), have shown significant improvements in efficiency and speed, but they require large, diverse, and annotated datasets. Our study is motivated by the 2022 flooding and climate change effects in the global south, where there is a lack of data analysis, let alone semantic segmentation datasets for developing countries like Pakistan. Pakistan has a unique geographic landscape and plays a crucial role in global climate and politics, with greenhouse gas emissions that are less than 1% of the world total. Despite lower emissions, it is a hotspot for global warming consequences, and monitoring the situation in Pakistan is essential for recovery, rehabilitation, and safety of the region. We developed a dataset that captures the unique infrastructure of major cities in Pakistan, including Islamabad, Karachi, and Quetta. Our research is focused on developing Pak-dataset, comparing the performance of current semantic segmentation networks, and determining the most effective deep learning network for our created dataset. Pakistan has a distinctive geographical makeup due to its geo-political significance, which has been impacted by global politics and climate. Studies of semantic segmentation have been conducted on remote sensing imagery of urban and rural areas to classify buildings, vegetation, water, and roads. These studies support land coverage analysis, which can enhance urban infrastructure management, forestry, disaster management, and climate preparedness. Given Pakistan's critical role in observing the effects of climate change, this research focuses on semantic segmentation of satellite imagery in three major cities. This research is globally sig-

nificant as the effects of climate change in the global south will eventually impact the global north. The goal is to contribute to

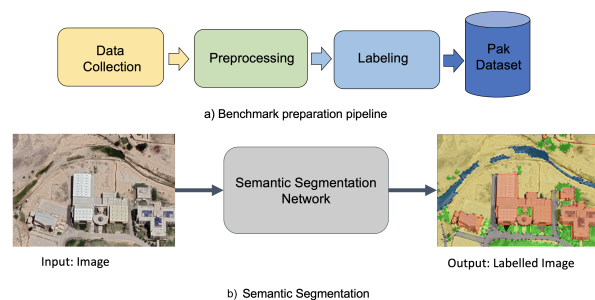


Figure 1: Overview of the contribution; a) Benchmark pipeline, b) semantic segmentation

the dataset from major cities in Pakistan, benchmark the performance of current semantic segmentation networks, and determine the most effective deep learning network for the dataset.

Our Contributions are as follows:

- Pak-dataset of three major cities of Pakistan.
- Benchmark the performance of SOTA networks to evaluate the dataset.

## 2. DATASET APPLICATIONS

In this section, we will briefly discuss the application for our dataset. Satellite imagery datasets can play an important role in solving various problems in planning infrastructure, disaster management, and mapping.

### 2.1 Planning infrastructure:

By providing high-resolution images of cities and rural areas, satellite image datasets can support the development of urban infrastructure, such as roads, buildings, and other facilities. This information can be used to make data-driven decisions about where

\*Corresponding author.

to build new infrastructure, how to upgrade existing infrastructure, and how to maintain and manage the infrastructure over time.

## 2.2 Disaster management:

In the event of a natural disaster, satellite images can be used to assess the extent of damage and to support disaster response and recovery efforts. For example, images can be used to identify disaster affected regions, to locate critical infrastructure that may have been damaged, and to support the deployment of response and recovery resources.

## 2.3 Mapping:

Satellite images can be used to create detailed maps of cities, rural areas, and other landscapes. These maps images can be utilized for numerous purposes, such as guiding navigation, supporting land use planning, and improving emergency response efforts. In summary, satellite image datasets provide valuable information that can be used to solve various problems. The data is essential for decision-makers, planners, and emergency response and can support efforts to make cities and communities more resilient, sustainable, and safe. Satellite images can be used to create detailed maps of cities, rural areas, and other landscapes. These maps can be used for a variety of purposes, such as guiding navigation, supporting land use planning, and improving emergency response efforts.

## 3. OBJECTIVES & RELATED WORK

Our research has two main objectives. Firstly, we explore various benchmark datasets for studying land coverage analysis, aiming to determine whether they can be effectively generalize for use with Pakistan's satellite imagery. Among the datasets we investigated LoveDA (Wang et al., 2021a), OpenEarthMap (Xia et al., 2022), LandCover.ai (Boguszewski et al., 2020), and DeepGlobe (Demir et al., 2018a), as well as several other open datasets. However, a significant limitation of these datasets was their inability to generalize to Pakistan's unique terrain and infrastructure. Secondly, benchmark this dataset for semantic segmentation of 2D images.

### 3.1 Land-Cover Dataset for Semantic Segmentation

The topic of land-cover semantic segmentation has been extensively studied for many years. Early research used low to medium-resolution datasets like MCD12Q1 (Sulla-Menashe and Friedl, 2018), NLCD (Jin et al., 2019), GlobeLand30 (Jun et al., 2014), and LandCoverNet (Alemohammad and Booth, 2020). Nonetheless, these research works concentrated on extensive mapping tasks with a macroscopic viewpoint. But, the advent of modern remote sensing technologies has made it possible to obtain high-resolution imagery from both airborne and spaceborne platforms on a daily basis. These high-resolution (HSR) land-cover datasets provide a micro-level perspective, containing distinct geometric structures and intricate textures. Datasets such as ISPRS Vaihingen, ISPRS Potsdam, Zeebruges (Marcos et al., 2018) and Zurich Summer (Volpi and Ferrari, 2015), were created for urban parsing but have limited coverage and annotated images. LandCover.ai (Boguszewski et al., 2021) and DeepGlobe (Demir et al., 2018b) focus on countryside with coverage that is enormous but with handful of synthetic structures. The chinese dataset GID (Tong et al., 2020) contains both villages and metropolitan areas, but the topographical placement and city identifications are private. HSR land-cover datasets that already exist primarily prioritize accuracy improvement, ignoring transferability. Only the

vital objects that are semantic segmentation in nature are dealt with in iSAID dataset (Waqas Zamir et al., 2019) bringing variable challenges for remote sensing tasks. FCN (Long et al., 2015) variants have been evaluated using these datasets (Chen et al., 2019, Dong et al., 2020, Duan et al., 2020, Wang et al., 2020b), and public datasets have been combined to produce recent UDA algorithm (Yan et al., 2019). However, insufficient common categories might result from combined datasets or inconsistent annotation granularity. Urban and Rural are the two domains that are the composition of The LoveDA dataset (Wang et al., 2021b) encompasses two domains (urban and rural), presenting for land-cover mapping a never before used UDA technique. Nevertheless, our dataset is one of a kind in targeting cities of Pakistan especially in the case of Quetta, which have not received any recognition prior to this particular paper in the field of semantic segmentation.

### 3.2 Semantic Segmentation

Semantic segmentation is the process of categorizing every pixel in an image into different categories or classes. This is a challenging and tedious process in computer vision, and numerous approaches have been proposed to solve it. Recently, deep learning techniques have shown favorable results for semantic segmentation. Here, we discuss four popular deep learning-based architectures for semantic segmentation: UNet (Ronneberger et al., 2015b), UNet++ (Zhou et al., 2018a), FPN (Lin et al., 2017), and DeepLab (Chen et al., 2017a). UNet (Ronneberger et al., 2015b) is a popular deep learning model for semantic segmentation. The architecture of UNet compose of a decoder and an encoder blocks. The encoder network downsamples the input image and extracts high-level features, while the decoder network upsamples the features to generate the final segmentation map. The skip connections among the encoder and decoder blocks help in preserving the spatial information of the input image. UNet++ (Zhou et al., 2018a) is an extension of UNet. The architecture of UNet++ is similar to UNet, but it has more skip connections between the decoder and encoder network, which helps in preserving spatial information even better. UNet++ also introduces a nested and dense skip connection architecture, which further boost the performance of the model. FPN (Lin et al., 2017) is a deep learning-based model for object detection. However, FPN has also been widely used for semantic segmentation. FPN compose of a backbone network and a feature pyramid network. The backbone network take out features from the inserted image, while the FPN generates a feature pyramid that captures features at different scales. The feature pyramid is then used to produce the resultant segmentation map. DeepLab (Lin et al., 2017) is a deep learning-based model for semantic segmentation. The architecture of DeepLab consists of a backbone network and an Atrous Spatial Pyramid Pooling (ASPP) module. The ASPP module uses atrous convolutions to capture multi-scale contextual data, which helps in improving the segmentation performance. DeepLab also uses skip connections among the backbone network and ASPP module to reattain the spatial information of the input image. All of these architectures are designed to improve the performance of semantic segmentation by preserving spatial information and capturing multi-scale contextual information. They have shown promising results on various datasets and have been commonly utilize in numerous implementations, such as autonomous driving, medical image analysis, and object detection. Secondly, a comparative study of different methods for land coverage analysis using satellite imagery. Traditional image segmentation methods include thresholding, region-based approaches, edge-based approaches, watershed methods, and clustering methods. While these methods have shown reasonable performance, they require domain knowledge for feature crafting, thresholding, and clus-

tering. In contrast, deep learning semantic segmentation neural networks have shown significant improvements in performance and have been widely adopted by academia and industry.

## 4. PAK-DATASET

### 4.1 Dataset Overview

In our research, we have created a unique satellite imagery dataset, specifically for Pakistan. This dataset, named Pak-Dataset, consists of 600 images (initially 150 images of 1024x768, during the experiments, each image is split into 4, making it 600 images of 512x512 resolution, further details mentioned in section 6.1) from three different cities, which were obtained through the use of the Google Earth API. The dataset provides coverage of a wide range of terrain, including Quetta, a valley and the least developed major city in Pakistan; Karachi, the country's largest seaport and most organically developed metropolitan city; and Islamabad, the capital city, which was established through urban planning in the 1960s, nearly 20 years after the creation of Pakistan. Our dataset includes five major semantic segmentation classes: background, buildings, vegetation, water, and roads. All images are of dimension 512x512 and pre-processed to eliminate noise, adjust brightness and contrast, and correct for geometric distortions. The data covers an area of approximately  $7km^2$  as shown in figure 2.

### 4.2 Preparation of Dataset benchmark

To develop a dataset of satellite images of Pakistan using Google Earth Pro, these steps are followed:

- Google Earth Pro: We configured Google Earth Pro on the computer. It's an open source software that allows anyone to explore the world map, including satellite imagery of cities, landscapes, and more.
- Determine the area of interest: We identified the cities or regions in Pakistan which were selected to reflect in your dataset. We selected Quetta, Karachi and Islamabad for their unique infrastructure, diverse terrain and drastically different environmental characteristics. We have used Google Earth Pro to navigate to these areas and zoom in on the satellite imagery.
- Export the imagery: Once, we have navigated to the desired cities (i.e. Quetta, Karachi and Islamabad), We exported satellite imagery as a high-resolution image of  $1024 \times 768$ . These images later then be processed for different categories.
- Pre-processing of images: We pre-processed to remove noise, adjust the brightness and contrast, and correct for geometric distortions.
- Annotation: The next step is to annotate the images to label different objects and areas within the images, such as background, buildings, vegetation, water, and roads. This process is known as labeling segments and regions with enclosing class.

### 4.3 Data Labeling

As discussed in above section we labeled data into five classes.

- Building: This class represents the man-made built areas such as houses, hospitals, malls and all other commercial buildings.

- Vegetation: This class is made of grass, trees and bushes.
- Water: All water bodies.
- Roads: It represents the roads, side walks, highways.
- Background: Any region or object that doesn't fall in above mentioned classes.

The final annotation is carried out using the Matlab Image Labeler app, and we refined these annotations further using a Matlab script to merge regions and polygons. The annotated images are then organized into a dataset, with the labels and image data stored in a format that can be used for deep learning. Generating a dataset utilizing Google Earth Pro may entail significant time and effort, especially for large datasets and extensive coverage areas. However, the resulting dataset holds immense value for various applications, including but not limited to urban infrastructure management, forestry, disaster management, and climate preparedness.

## 5. 2D SEMANTIC SEGMENTATION

The dataset is developed so it can be used to benchmark various deep learning networks for semantic segmentation especially for climate hotspot such as Pakistan with unique terrain and infrastructure. Comparing the performance of SOTA semantic segmentation networks is a crucial step in evaluating the quality and usefulness of a dataset, as it helps to ensure that the data is suitable for the intended use case and that the best possible results can be obtained from the data. A common evaluation approach involves employing different deep learning networks for semantic segmentation on the same dataset and thoroughly analyzing their results. The objective is to identify the network that performs best on the specific dataset and uncover any areas that may require improvement in terms of network performance. In our study, we employ transfer learning using ResNet-50 as the backbone for semantic segmentation over all five classes. We split the dataset into training and testing sets, based on cities and cross-city dataset split. To train the model, we use weighted cross entropy with an Adam optimizer, learning rate of  $1e-4$ , and batch size of 4. Data augmentation techniques is used to address class imbalance scenarios, improving the overall model performance. Our experimental results compared the performance of LoveDA and Pak-Dataset on UNet, UNet++, FPN, and DeepLabv3 networks, achieving an mIOU of 43% for UNet, 45% for UNet++, 36% for FPN, and 33% for DeepLabv3, which is comparable to LoveDA. The dataset also provided insights into the land coverage in the major cities of Quetta, Karachi, and Islamabad as shown in figure 3. However, the size of our dataset is significantly smaller. This indicates that in future we can achieve better results by expanding our dataset and using techniques of domain adaptations.

## 6. EXPERIMENT AND RESULTS

### 6.1 Dataset Description

The Pakistan Climate Change dataset comprises 150 satellite images of Pakistan, captured at an altitude of approximately 300m from the ground. The dataset has been divided into three equal parts, comprising a total of 150 images. Specifically, the first 50 images represent Quetta, the second 50 images depict Karachi, and the last 50 images showcase Islamabad. These images encompass objects from five distinct classes: water, road, building,

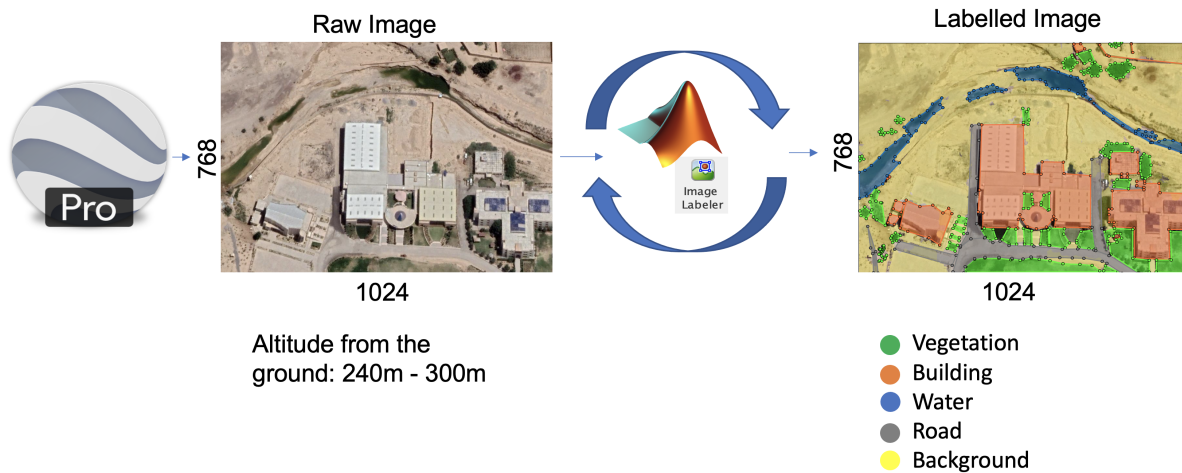


Figure 2: Data preparation and annotation pipeline

vegetation, and background. Any objects that do not fall within the first four classes are considered part of the background class. Each image has a spatial area of  $41800m^2$ . The whole dataset covers an area of approximately  $7km^2$ . The images are geometrically registered and pre-processed. The Pakistan Climate Change dataset is annotated using a comprehensive annotation pipeline. The dataset is intended for evaluation tasks such as semantic segmentation. The dataset is designed to provide a diverse and comprehensive collection of images for researchers in the field of remote sensing and computer vision. The images are annotated into five classes, namely water, road, building, vegetation, and background, enabling deep learning algorithms to identify and classify different features in the images accurately.

## 6.2 Experimental Configuration

To test the effectiveness of various architectures in semantic segmentation, especially in remote sensing, several common architectures and their variants were evaluated using the Pakistan Climate Change dataset. The networks that were selected for testing were UNet (Ronneberger et al., 2015b), UNet++ (Zhou et al., 2018a), DeepLabV3+ (Chen et al., 2017a), FPN (Lin et al., 2017). The F1-Score and intersection over union (IoU) was utilized to measure the accuracy of the semantic segmentation, following the standard practice (Long et al., 2015, Wang et al., 2020a). The F1-Score, which represents the average of the F1-Score across all categories, was used to report the F1-Score for each class.

These metrics offer valuable insights into the overall performance of each network, as well as their performance on specific classes or objects within the dataset. After comparing the performance of the different networks, the most effective network can then be selected for use on the dataset. This information can also be used to guide future improvements to the networks, such as fine-tuning the network architecture or incorporating additional data into the training process.

## 6.3 Results and Discussion

The results show performance comparison between SOTA networks for 2D semantic segmentation in figure 4. In table 6.3, results reflected the higher performance by UNet++. We also compare the performance of SOTA over precision and recall of each category namely background, water, vegetation, building and road. The selected SOTA are UNet, UNet++, FPN and DeepLab shown in table 6.3.

In the table 6.3, we present f1 score of each category according to different resolution, the resolution we have chosen are  $128 \times 128$  and  $512 \times 512$ . The interesting insight we observed that the higher resolution images had much better performance when training models especially with UNet and UNet++. The categories like water and road that have minor share in the dataset are not very well recognized by the models especially in feature pyramid network and DeepLab because those architectures are not very well suited for semantic segmentation. The U-Net and U-Net++ models are specifically designed for semantic segmentation tasks and have been shown to outperform other segmentation models such as FPN and DeepLab in many cases. U-Net and U-Net++ have skip connections that allow for better feature reuse and combination. In U-Net, the result of each encoder block is concatenated with the corresponding decoder block, allowing for more precise localization of features. U-Net++ takes this a step further by including multiple paths for information flow, allowing for more diverse feature combinations.

The U-Net architecture has a symmetrical decoder-encoder structure that allows for the collection of both high-level and low-level features. This can be particularly useful in tasks such as biomedical image segmentation, where fine-grained details are important. U-Net and U-Net++ use convolutional layers that are deeper and wider than those in FPN and DeepLab. This deeper and wider architecture allows for the extraction of more complex features that can better capture the variability and intricacies of the input data. U-Net and U-Net++ are end-to-end trained using back-propagation, allowing the network to learn more complex and nuanced features in an efficient manner.

## 7. CONCLUSION

To summarize up, satellite image datasets are a crucial resource that can aid in solving diverse problems in infrastructure planning, disaster management, and mapping. Such datasets are vital for decision-makers, planners, and emergency response and can help create more resilient, sustainable, and secure communities. Land coverage analysis through semantic segmentation can produce detailed maps of cities, rural areas, and other landscapes, useful for navigation, land use planning, and emergency response. Our future work intends to improve the enrichment of the datasets and expanding on these datasets to cover more cities.

Methods	F1 Score per class					mIoU	Overall Accuracy
	Background	Water	Vegetation	Building	Road		
UNet	0.49	0.02	0.78	<b>0.63</b>	<b>0.41</b>	0.43	0.60
UNet++	<b>0.53</b>	<b>0.21</b>	<b>0.83</b>	0.61	0.38	<b>0.45</b>	<b>0.63</b>
FPN	0.50	0.19	0.63	0.52	0.41	0.36	0.54
DeepLab	0.38	0.02	0.62	0.63	0.43	0.33	0.50

Table 1: Semantic segmentation comparative study for climate change Pak-dataset.

Methods	Background		Water		Vegetation		Building		Road	
	P	R	P	R	P	R	P	R	P	R
UNet	0.52	0.47	0.01	0.04	0.83	0.73	0.64	0.63	0.31	<b>0.57</b>
UNet++	0.55	0.51	<b>0.12</b>	<b>0.69</b>	0.80	<b>0.86</b>	<b>0.67</b>	0.55	0.33	0.44
FPN	0.45	<b>0.56</b>	<b>0.12</b>	0.44	0.68	0.60	0.56	0.49	<b>0.47</b>	0.36
DeepLab	<b>0.58</b>	0.28	0.01	0.55	<b>0.88</b>	0.48	0.51	<b>0.80</b>	0.39	0.48

Table 2: Semantic segmentation comparative study for climate change Pak-dataset over precision and recall of each category.

Methods	Resolution	F1 Score per class					Overall Accuracy
		Background	Water	Vegetation	Building	Road	
UNet	128x128	0.54	0.00	0.49	0.00	0.00	0.42
UNet++		<b>0.56</b>	0.00	0.59	0.00	0.00	0.45
FPN		0.55	<b>0.35</b>	<b>0.65</b>	0.00	0.01	<b>0.47</b>
DeepLab		0.55	0.01	0.56	0.00	0.00	0.44
UNet	512x512	0.49	0.02	0.78	<b>0.63</b>	0.41	0.60
UNet++		<b>0.53</b>	<b>0.21</b>	<b>0.83</b>	0.61	0.38	<b>0.63</b>
FPN		0.50	0.19	0.63	0.52	0.41	0.54
DeepLab		0.38	0.02	0.62	<b>0.63</b>	<b>0.43</b>	0.50

Table 3: Semantic segmentation image resolution based ablation study for climate change Pak-dataset.

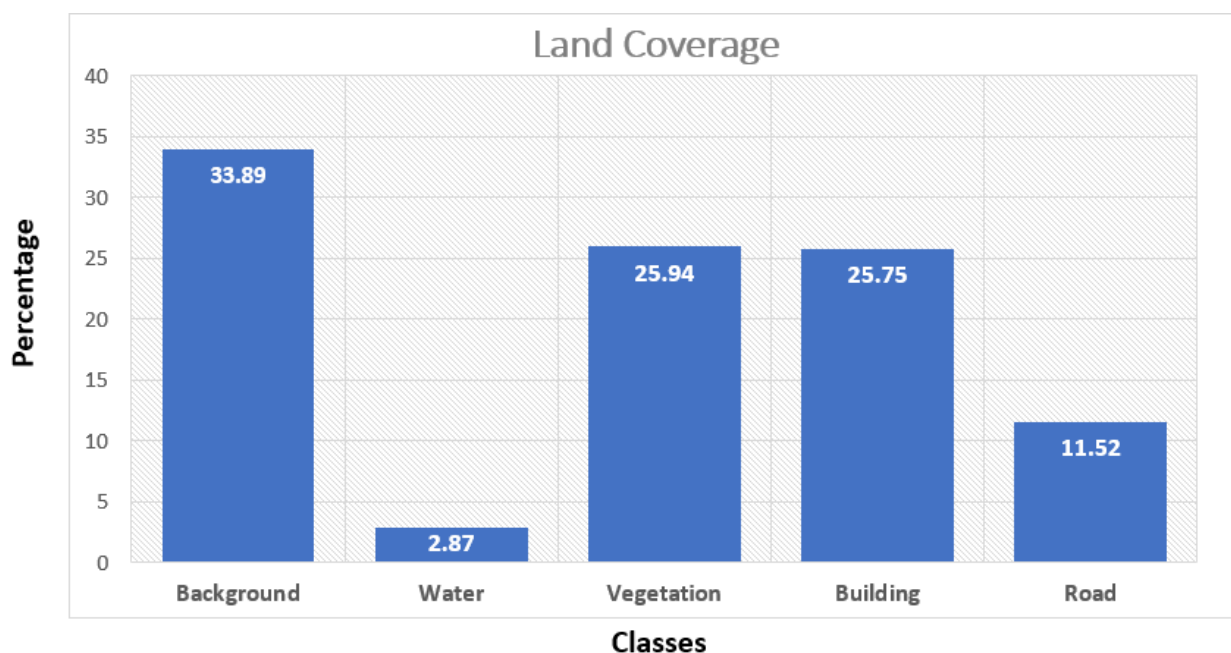


Figure 3: Land Coverage of Pak-Dataset

## REFERENCES

- Alemohammad, H. and Booth, K., 2020. Landcovernet: A global benchmark land cover classification training dataset. arXiv preprint arXiv:2012.03111.
- Boguszewski, A., Batorski, D., Ziemia-Jankowska, N., Dziedzic, T. and Zambrzycka, A., 2021. Landcover.ai: Dataset for automatic mapping of buildings, woodlands, water and roads from aerial imagery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1102–1110.
- Boguszewski, A., Batorski, D., Ziemia-Jankowska, N., Zambrzycka, A. and Dziedzic, T., 2020. Landcover.ai: Dataset for automatic mapping of buildings, woodlands and water from aerial imagery. CoRR.
- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A. L., 2017a. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence 40(4), pp. 834–848.
- Chen, L., Papandreou, G., Schroff, F. and Adam, H., 2017b. Rethinking atrous convolution for semantic image segmentation. CoRR.
- Chen, W., Jiang, Z., Wang, Z., Cui, K. and Qian, X., 2019. Collaborative global-local networks for memory-efficient segmentation of ultra-high resolution images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 8924–8933.
- Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D. and Raskar, R., 2018a. Deepglobe 2018: A challenge to parse the earth through satellite images. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) pp. 172–17209.
- Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D. and Raskar, R., 2018b. Deepglobe 2018: A challenge to parse the earth through satellite images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 172–181.
- Dong, Y., Liang, T., Zhang, Y. and Du, B., 2020. Spectral-spatial weighted kernel manifold embedded distribution alignment for remote sensing image classification. IEEE Transactions on Cybernetics 51(6), pp. 3185–3197.
- Duan, Y., Huang, H., Li, Z. and Tang, Y., 2020. Local manifold-based sparse discriminant learning for feature extraction of hyperspectral image. IEEE transactions on cybernetics 51(8), pp. 4021–4034.
- Jin, S., Homer, C., Dewitz, J., Danielson, P. and Howard, D., 2019. National land cover database (nlcd) 2016 science research products. In: AGU Fall Meeting Abstracts, Vol. 2019, pp. B111–2301.
- Jun, C., Ban, Y. and Li, S., 2014. Open access to earth land-cover map. Nature 514(7523), pp. 434–434.
- Lin, T., Dollár, P., Girshick, R. B., He, K., Hariharan, B. and Belongie, S. J., 2016. Feature pyramid networks for object detection. CoRR.
- Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B. and Belongie, S., 2017. Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117–2125.
- Long, J., Shelhamer, E. and Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440.
- Marcos, D., Volpi, M., Kellenberger, B. and Tuia, D., 2018. Land cover mapping at very high resolution with rotation equivariant cnns: Towards small yet accurate models. ISPRS journal of photogrammetry and remote sensing 145, pp. 96–107.
- Ronneberger, O., Fischer, P. and Brox, T., 2015a. U-net: Convolutional networks for biomedical image segmentation. CoRR.
- Ronneberger, O., Fischer, P. and Brox, T., 2015b. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18, Springer, pp. 234–241.

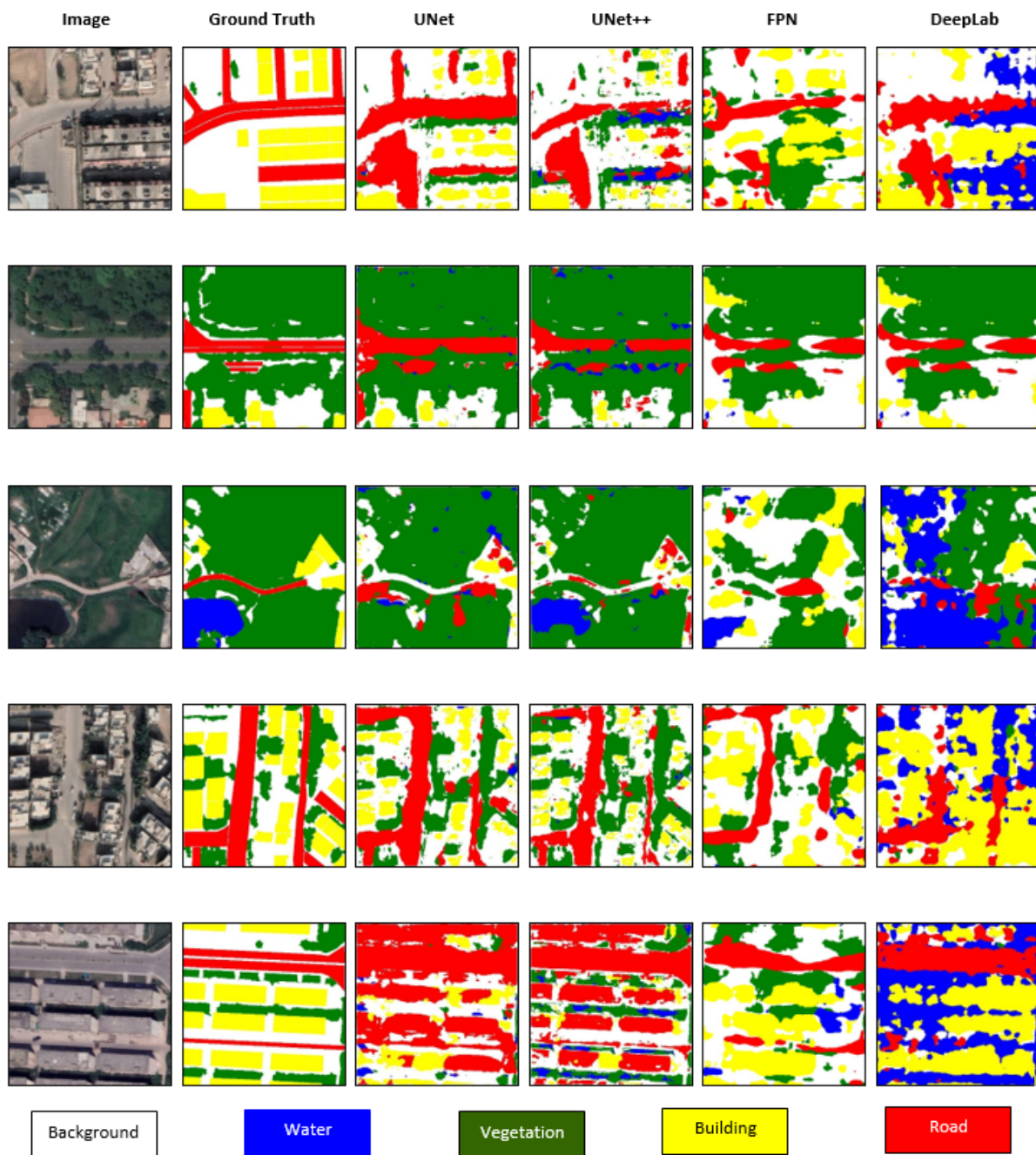


Figure 4: Semantic segmentation output on images from the Pakistan climate change test set. Some minor-scale objects such as water and vegetation are difficult to recognize in Quetta. Water is easy to recognize in Karachi and same is true for vegetation in Islamabad.

Sulla-Menasha, D. and Friedl, M. A., 2018. User guide to collection 6 modis land cover (mcd12q1 and mcd12c1) product. Usgs: Reston, Va, Usa 1, pp. 18.

Tong, X.-Y., Xia, G.-S., Lu, Q., Shen, H., Li, S., You, S. and Zhang, L., 2020. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sensing of Environment* 237, pp. 111322.

Volpi, M. and Ferrari, V., 2015. Semantic segmentation of urban scenes by learning local class interactions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1–9.

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu,

D., Mu, Y., Tan, M., Wang, X. et al., 2020a. Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence* 43(10), pp. 3349–3364.

Wang, J., Zheng, Z., Ma, A., Lu, X. and Zhong, Y., 2021a. Loveda: A remote sensing land-cover dataset for domain adaptive semantic segmentation. *CoRR*.

Wang, J., Zheng, Z., Ma, A., Lu, X. and Zhong, Y., 2021b. Loveda: A remote sensing land-cover dataset for domain adaptive semantic segmentation. *arXiv preprint arXiv:2110.08733*.

Wang, J., Zhong, Y., Zheng, Z., Ma, A. and Zhang, L., 2020b. Rsnet: The search for remote sensing deep neural networks in

recognition tasks. *IEEE Transactions on Geoscience and Remote Sensing* 59(3), pp. 2520–2534.

Waqas Zamir, S., Arora, A., Gupta, A., Khan, S., Sun, G., Shahbaz Khan, F., Zhu, F., Shao, L., Xia, G.-S. and Bai, X., 2019. isaid: A large-scale dataset for instance segmentation in aerial images. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 28–37.

Xia, J., Yokoya, N., Adriano, B. and Broni-Bediako, C., 2022. Openearthmap: A benchmark dataset for global high-resolution land cover mapping.

Yan, L., Fan, B., Liu, H., Huo, C., Xiang, S. and Pan, C., 2019. Triplet adversarial domain adaptation for pixel-level classification of vhr remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 58(5), pp. 3558–3573.

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N. and Liang, J., 2018a. Unet++: A nested u-net architecture for medical image segmentation. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, Springer, pp. 3–11.

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N. and Liang, J., 2018b. Unet++: A nested u-net architecture for medical image segmentation. *CoRR*.