FloodNet: A High Resolution Aerial Imagery Dataset for Post Flood Scene Understanding

21BCE1266 - Shayari Bagchi
Link: https://github.com/Shayari08/DIP
21BCE5475-Chavi Arora
Link: https://github.com/Chavi02/DIP-project

Introduction

Visual scene understanding has the potential to help us significantly advance in decision support systems, encompassing the categorization of scenes and the comprehension of interrelationships among object classes at both instance and pixel levels.

Although publicly available datasets like ImageNet, Microsoft COCO, PASCAL Visual Object Classes, and Cityscapes have accelerated the development of deep learning algorithms, there's a scarcity of aerial imagery datasets due to the challenges of obtaining annotations. Aerial scene understanding datasets are vital for applications like urban management, city planning, and infrastructure maintenance, particularly in post-disaster scenarios. Existing datasets mostly focus on the classification or semantic segmentation of individual classes, lacking the specificity needed for disaster damage assessment.

The pre-existing post-disaster management datasets either use satellite images or social media images, satellite images are too costly while social media pictures are noisy and cannot be scaled for a suitable deep learning model. In the aftermath of natural disasters like hurricanes, wildfires, and severe flooding, rapid response and recovery on a large scale are imperative. Ensuring the response team has swift access to high-resolution aerial images is crucial for effective disaster management. To address this gap, we can use FloodNet, a high-resolution aerial imagery dataset for three computer vision tasks: classification, semantic segmentation, and visual question answering (VQA). FloodNet offers clarity to scenes by providing images from low altitudes, reducing obstructions from clouds and smoke compared to satellite

images. This dataset includes pixel-level annotations, setting it apart from existing natural disaster datasets that primarily focus on classification and object detection. FloodNet aims to assist rescue teams in efficiently managing operations during emergencies. Our contribution includes introducing this high-resolution UAV imagery dataset and comparing the performance of various classification, semantic segmentation, and VQA methods on FloodNet. To our knowledge, this is the first semantic segmentation and VQA work focused on UAV imagery for disaster damage assessment.

Related Works

In the field of computer vision, in recent years there has been significant progress in the field of semantic segmentation in satellite images, particularly in object detection and semantic labelling.

In one study, with a focus on deep learning advances, the paper provided an extensive assessment of the literature on processing methods for field boundary extraction from satellite imagery. The task formulation is described as semantic segmentation, in which every image pixel is classified into classifications, such as boundaries and fields. Using deep convolutional neural networks (CNNs) and taking advantage of its ability to extract features and grasp spatial context is a noteworthy development. Because of the multitasking nature of the model design, it is possible to estimate field, border, and distance probabilities for each pixel at the same time. Furthermore, conditional inference processes are incorporated to improve predictions by taking certain contextual data into account. In-depth discussions are held regarding post-processing techniques for improving the retrieved boundaries, including morphological procedures, thresholding, and clustering. Alternative approaches are also covered in the paper, such as instance segmentation strategies like Mask R-CNN and Faster R-CNN. Moreover, transfer learning's potential for model adaptation to new geographic areas is considered, highlighting the need for flexible processing methods for precise and effective field boundary extraction from satellite data.

Semantic segmentation has alsp been extensively used to understand man-made features like roads, buildings, land use, and land cover types.

Large-scale datasets have played a crucial role in advancing deep learning-based approaches for these tasks. However, existing

datasets have limitations in capturing the complexity and diversity of real-world urban scenes.

To address this gap, the Cityscapes dataset was introduced as a benchmark suite and large-scale dataset for pixel-level and instance-level semantic labeling in urban scenes Cityscapes consists of a diverse set of stereo video sequences recorded in streets from 50 different cities, providing a comprehensive representation of urban environments. The dataset includes 5,000 images with high-quality pixel-level annotations and an additional 20,000 images with coarse annotations to enable methods that leverage weakly-labeled data. The Cityscapes dataset stands out in terms of its dataset size, annotation richness, scene variability, and complexity, surpassing previous attempts in these aspects. It has been designed to spark progress in semantic urban scene understanding by creating a large and diverse dataset, developing a sound evaluation methodology, and providing an in-depth analysis of dataset characteristics.

The significance of the Cityscapes dataset is evident in several observations. Firstly, the relative performance order of state-of-the-art methods on Cityscapes differs from more generic datasets, indicating the need for specialized datasets for urban scene understanding.

Secondly, applying an off-the-shelf fully-convolutional network trained on Cityscapes outperforms current state-of-the-art methods on other urban scene datasets, highlighting the unique benefits of Cityscapes Lastly, the challenges posed by Cityscapes, with current best-performing methods achieving lower IoU scores compared to other datasets, emphasize the need for further advancements in urban scene understanding.

The introduction of Cityscapes has paved the way for future research and progress in semantic urban scene understanding. Its large-scale and diverse nature, along with its comprehensive annotations, make it a valuable resource for developing and evaluating new approaches in this field. The dataset is expected to drive advancements in scene understanding and contribute to the development of more efficient and accurate convolutional neural network architectures for urban environments.

In conclusion, the Cityscapes dataset has emerged as a significant contribution to the field of semantic urban scene understanding. Its size, diversity, and comprehensive annotations make it a valuable resource for training and testing approaches for pixel-level and instance-level semantic labeling in complex urban street scenes. The dataset has already shown its potential by enabling advancements in performance and highlighting the need for specialized datasets in this domain Future adaptations and updates to Cityscapes are expected to further drive progress in semantic urban scene understanding.

However, most of these analyses are still limited to static snapshots of data involving images acquired at a single time instance. This means that the existing approaches primarily focus on analyzing individual images without considering the temporal aspect.

To overcome this limitation and determine the area impacted by a disaster, it is crucial to extend these approaches to time-series data to detect areas of change. By analyzing multiple snapshots of satellite images captured at different time periods, we can gain insights into the dynamic changes that occur in the aftermath of a disaster.

A simple solution to detect change in time-series data is to directly compare the raw RGB values of satellite images. However, due to factors such as different seasons, lighting conditions, and noise, the pixel values across time-series data can vary significantly even in areas with no disaster impact. Therefore, researchers have explored various techniques to improve disaster mapping from satellite images.

In one study, the authors highlight the use of MODIS (Moderate Resolution Imaging Spectroradiometer) data to develop models for disaster detection. They emphasize the importance of leveraging different satellite data sources and efforts for disaster response. More recent approaches have focused on using Convolutional Neural Networks (CNNs) for disaster detection from satellite images. These CNN-based models can effectively detect damaged buildings by

considering damaged and non-damaged buildings as two distinct classes.

However, a limitation of these approaches is the requirement for large training datasets specifically for damaged areas, which can be expensive and not easily scalable. To address this challenge, a proposed approach aims to locate areas of maximal disaster damage by using man-made features as a reference. By training models based on Fully-Convolutional Neural Networks to detect roads and buildings from satellite imagery, prediction masks can be generated in regions with disaster impact.

The key idea is to compute the relative change between multiple snapshots of data captured before and after a disaster. By analyzing the change in high-level man-made features, such as roads and buildings, it becomes possible to identify areas of maximal damage and prioritize disaster response efforts. Importantly, this approach is invariant to seasonal variations, lighting conditions, and noise differences in time-series data.

Compared to previous approaches that require training CNNs specifically for detecting damaged features, the proposed approach utilizes models trained on general road and building datasets, which are relatively inexpensive and scalable to other similar natural disasters. Evaluation on both human-annotated datasets and state-provided datasets of actual disaster-impacted areas demonstrates a strong positive correlation between predicted disaster areas and actual disaster-impacted areas.

Proposed Methods

The study offers a number of techniques for examining the FloodNet dataset, including as semantic segmentation, visual question answering (VQA), and image categorization-

Image Classification:- This task involves categorizing images into flooded or non-flooded categories. Three cutting-edge convolutional neural network (CNN) architectures are used in the paper: InceptionNetv3, ResNet50, and Xception. The FloodNet dataset is used to fine-tune these networks once they have been pretrained on ImageNet. For 30 epochs, the CNN models are trained with binary cross-entropy loss and customized hyperparameters to achieve maximum efficiency.

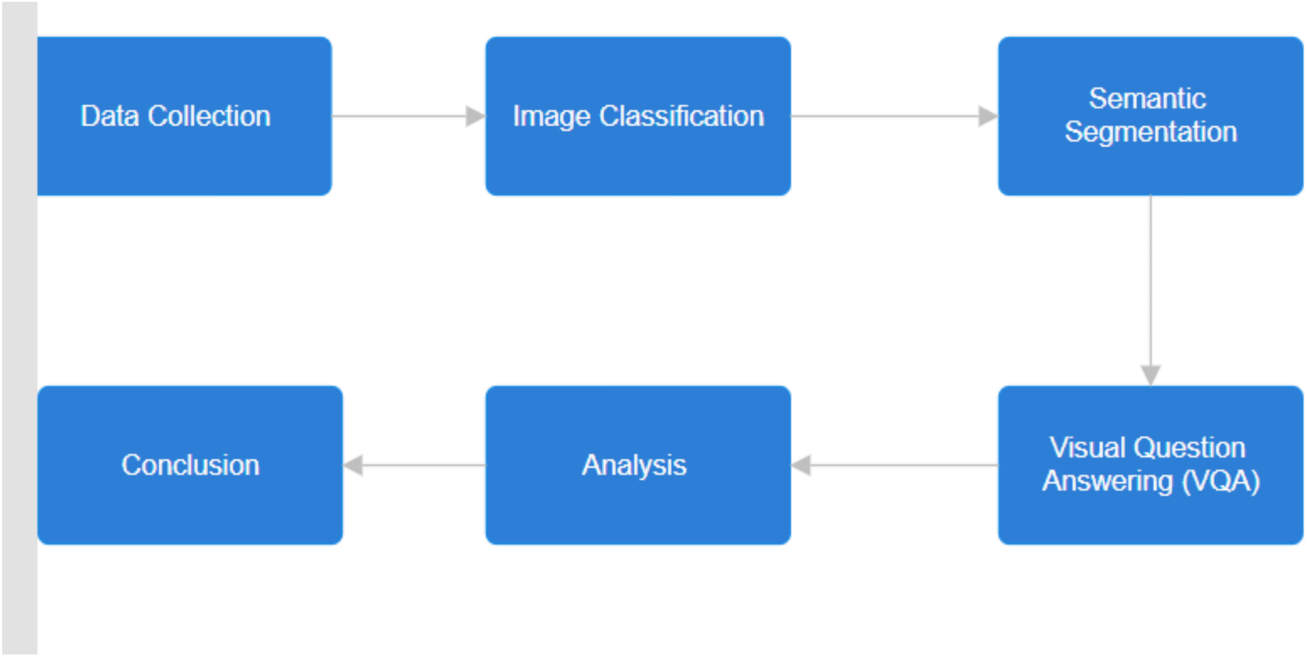
Semantic Segmentation:- This technique seeks to differentiate between various items and areas in an image by applying semantic labels to every pixel. Three segmentation models—PSPNet, ENet, and DeepLabv3+—are proposed in this study. ResNet101 serves as the backbone for PSPNet, while ENet and DeepLabv3+ have different topologies. Specific learning rate schedules, weight decay, and other hyperparameters are used during the training process for each model. To avoid overfitting, image augmentation methods including flipping, scaling, shuffle, and random rotation are used.

Visual Question Answering (VQA):- VQA involves generating responses to image-related queries. Lastly, the study describes two baseline techniques for VQA: Multimodal Factorized Bilinear (MFB) with co-attention and simple concatenation or element-wise product of image and text features. These techniques are combined with feature extraction networks, such as a Two-Layer LSTM for questions and VGGNet for images. The dataset is divided into testing, validation, and training sets. Stochastic gradient descent (SGD) with cross-entropy loss is then used to optimize the models.

Dataset

The data collected by DJI Mavic Pro quadcopters and other small unmanned aerial vehicle (UAV) platforms following Hurricane Harvey, a Category 4 hurricane that made landfall in the vicinity of Texas and Louisiana in August 2017, is included in the FloodNet Dataset. This dataset consists of images and videos taken during several flights that were undertaken between August 30, 2017, and September 4, 2017, mostly in Ford Bend County, Texas, and adjacent regions that were immediately impacted. It is notable for two primary reasons. First of all, it provides a high degree of authenticity since it includes footage taken by tiny unmanned aerial vehicles (UAVs) during the emergency response phase, giving a precise depiction of the procedures and surroundings during a crisis. Second, it stands apart from other datasets that might contain images from larger, fixed-wing assets operating at altitudes over the 400 feet AGL limit for small unmanned aerial vehicles (sUAVs) since it is the only collection of small UAV photography expressly designed for catastrophe scenarios. With the FloodNet dataset, all flights were notable for being carried out at 200 feet AGL, which produced high-resolution imagery that demonstrated the post-flood devastation to impacted areas. Following Hurricane Harvey, these photos show a variety of items, including roads, buildings, and other structures, along with their respective states —that is, whether they are flooded or not. An important concern in the creation of this dataset for tasks like semantic segmentation and visual question answering was the annotation of properties associated with these objects.

The FloodNet dataset's training and testing sets were created using a methodical process described in the paper. The dataset was specifically divided into training, validation, and testing sets, with 30% of the data going toward testing and the remaining 70% going toward training. About 70% of the entire dataset was used as the training set, guaranteeing that there was enough information to train the models for a variety of tasks, including picture classification, semantic segmentation, and visual question answering. After Hurricane Harvey, a variety of scenarios and settings were recorded by UAV platforms; this piece of the dataset was carefully selected to reflect these variables. With a similar distribution of samples between the two sets, the testing and validation sets each made up around 15% of the overall dataset. The purpose of these subgroups was to assess how well the trained models performed and how well they could generalize. Testing set was left unaltered until the last assessment step, which measured models' efficacy in real-world scenarios. The validation set was used to adjust hyperparameters and track the model's performance during the training phase.



References:

- M. Rahnemoonfar, T. Chowdhury, A. Sarkar, D. Varshney, M. Yari and R. R. Murphy, "FloodNet: A High Resolution Aerial Imagery Dataset for Post Flood Scene Understanding," in IEEE Access, vol. 9, pp. 89644-89654, 2021, doi: 10.1109/ACCESS.2021.3090981.

- F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," Google, Inc.

- M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding,”
- Dailmer AG R&D, TU Darmstadt, MPI Informatics, TU Dresden, www.cityscapes-dataset.net

- L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam

- Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," Google Inc.

- W. Chen, Z. Jiang, Z. Wang, K. Cui, and X. Qian, "Collaborative Global-Local Networks for Memory-Efficient Segmentation of Ultra-High Resolution Images," Department of Computer Science and Engineering, Texas A&M University, Department of Electrical and Computer Engineering, Texas A&M University

- J. Doshi, S. Basu, and G. Pang, "From Satellite Imagery to Disaster Insights," CrowdAI, Facebook

- F. Waldner et al., "Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network," Remote Sensing of Environment (2020)