Indian Institute of Technology, Guwahati

CS590 (Deep Learning)

Department of Computer Science and Engineering

Enabling Country-Scale Land Cover Mapping with Meter-Resolution Satellite Imagery

AI Geeks

1.Theodore Suanlalfaka (200121060)
2.Heibormi Suchiang (200121020)
3.Piyush S Shende (200121038)
4.Somesh Raj (200121052)
5.Surwade Bhushant Anant (200121055)

September 29, 2023

Abstract

In the realm of land cover classification, high-resolution satellite images offer valuable and intricate spatial insights, particularly within complex urban landscapes. However, leveraging these images for detailed land cover mapping on a large scale has been hindered by challenges such as intricate land cover patterns, expensive collection of training samples, and significant distribution shifts in satellite imagery due to geographical and acquisition variations. Addressing this gap, we introduce the Five-Billion-Pixels dataset, a comprehensive land cover dataset. This dataset comprises over 5 billion labeled pixels extracted from 150 high-resolution Gaofen-2 (4 m) satellite images, annotated across 24 distinct categories encompassing artificial-constructed, agricultural, and natural classes.

Moreover, we propose a deep learning-driven unsupervised domain adaptation approach designed to transfer classification models from a labeled dataset (referred to as the source domain) to unlabeled data (referred to as the target domain) for extensive land cover mapping. Our approach integrates an end-to-end Siamese network employing dynamic pseudo-label assignment and a class balancing strategy to facilitate adaptive domain joint learning. To ascertain the versatility of our dataset and the proposed approach across various sensors and geographical regions, we conduct land cover mapping in five megacities in China and six cities in five other Asian countries, utilizing satellite images from PlanetScope (3 m), Gaofen-1 (8 m), and Sentinel-2 (10 m). Despite entirely unlabeled input images and spanning a cumulative study area of 60,000 km2, the experiments demonstrate promising results. The proposed approach, trained with the Five-Billion-Pixels dataset, facilitates high-quality, detailed land cover mapping at meter-resolution not only across China but also in several other Asian countries.

Table of Contents

Chapter 1

Introduction

Land cover information is critical across various domains like environmental science, climate monitoring, food security, urban planning, disaster management, and ecosystem protection. Human activities are increasingly affecting both urban and natural environments, necessitating timely and accurate large-scale land cover data to guide construction efforts and mitigate adverse environmental changes.

Over the years, extensive research has focused on large-scale land cover mapping using low-/medium-spatial resolution remote sensing images, such as MODIS, Landsat TM, and ETM+ satellite images, yielding significant achievements. However, these images lack the spatial detail needed to distinguish specific land cover categories, especially those prevalent in the built environment. Recent efforts utilizing higher spatial resolution satellite imagery like Sentinel have improved land cover mapping, but challenges remain due to complex land structures and feature distribution shifts caused by varied imaging conditions.

Advancements in satellite technology offer higher spatial resolution, providing richer ground object information. However, this detailed information complicates land cover classification, particularly in highly heterogeneous areas like densely populated megacities. Additionally, feature distribution shifts due to diverse imaging conditions pose challenges in using high-resolution satellite images for large-scale land cover mapping.

To address these challenges, this work presents a comprehensive land cover dataset, Five-Billion-Pixels, containing over 5 billion labeled pixels from high-resolution Gaofen-2 satellite images. It covers a 24-category system encompassing artificial-constructed, agricultural, and natural classes. Furthermore, a deep-learning-based unsupervised domain adaptation approach is proposed to facilitate large-scale land cover mapping, even with entirely unlabeled input images. This approach demonstrates promising results across various sensors and geographical regions, showcasing its potential for meter-resolution land cover mapping in China and other Asian countries.

Chapter 2

Problem Statement and Aim

2.1 Problem statement

The main objective is to develop UDA model for semantic segmentation using remote sensing data.

2.2 Aim:

The primary aim of this study is to enhance large-scale land cover mapping by leveraging high-resolution satellite imagery, overcoming challenges related to complex land patterns, costly training data collection, and distribution shifts in satellite imagery. The objective is to develop a robust approach that effectively utilizes high-resolution imagery to provide detailed and accurate land cover information across diverse categories at a larger scale.

Chapter 3

Objectives:

1. Dataset Creation:
- Develop the Five-Billion-Pixels dataset, comprising over 5 billion labeled pixels extracted from high-resolution Gaofen-2 satellite images.
- Annotate the dataset with a comprehensive 24-category system, encompassing artificial-constructed, agricultural, and natural classes.

2. Algorithm Development:
- Design a deep-learning-based unsupervised domain adaptation approach to transfer classification models trained on labeled datasets (source domain) to unlabeled data (target domain).
- Implement an end-to-end Siamese network employing dynamic pseudo-label assignment and a class balancing strategy to enable adaptive domain joint learning.

3. Validation and Generalization:
- Validate the generalizability of the proposed approach across different sensors and geographical regions by conducting land cover mapping in various cities using satellite images from different sources (PlanetScope, Gaofen-1, Sentinel-2).
- Demonstrate the efficacy of the approach by achieving high-quality and detailed land cover mapping across the entire country of China and select Asian countries, even with entirely unlabeled input images.

4. Comparison and Evaluation:
- Compare the performance of the proposed approach with existing methods for land cover mapping using high-resolution imagery.
- Evaluate the accuracy, efficiency, and applicability of the approach in comparison to other approaches, especially in challenging urban environments with complex land cover patterns.

5. Integration and Application:
- Integrate the developed approach into practical applications related to environmental science, urban planning, disaster management, and ecosystem protection.
- Facilitate the use of high-resolution satellite imagery for guiding human settlement construction and mitigating negative environmental changes through informed land cover mapping.

Chapter 4

Literature review

The paper https://x-ytong.github.io/project/Five-Billion-Pixels.html. is about creating UDA models using U net built on top of a Siamese network for satellite image segmentation. Due to problems such as distribution shifts, geographical differences , information from High-resolution satellite images are not good enough for land cover classification. To solve the problem it presents a large scale and cover dataset, Five-Billion-Pixels. It contains more than 5 billion pixel of 150 high resolution Gaofen -2 satellite images. Due to cloud obstruction, it is important to use images captured at different times and positions by the sensor in the satellite. The main goal of this paper is to use UDA model(Unsupervised Domain Adaption Model) i.e to provided classification to unlabeled data(Target Domain) from labelled data(Source Domain). Since we are using Deep Convolution Neural Network learning models the performance will rely heavily on the quality and quantity of the data. The two domains may have very diverse conditions that could lead to widely dispersed feature space and class imbalance which could hamper our model ,So to resolve these problems we use pseudo labeling techniques. Siamese neural network is a branch network consisting of two branch use mainly in similarity measurement.

The main contributions of this paper is
-1)presenting land-cover classification of Five billion pixels,
-2) the implementation of UDA approach for the land cover mapping
-3)land cover mapping on five megacities in China and six cities in other five Asian countries.
The main concern for our project is that we propose a scheme that could improve the UDA model for semantic segmentation used for this land cover classification.

Chapter 5

Study Data

All images are from Gaofen-2(GF-2) satellite. The Multi spectral images used for Five Billion Pixels possess a spectral range of blue, green red and near infrared with an image resolution of 6800x7200 pixels. The five billion pixels are annotated manually using process such as coarse labeling, fine labeling, fine checking and spot checking. The category system of five billion pixels includes areas such as industrial area, urban residential, etc. Unclear areas are unlabeled. Study areas include five Chinese cities Beijing, Chengdu, Guangzhou, Shanghai and Wuhan. Additional Asian cities Bangkok, Thailand; Delhi, India; Naypyidaw, Myanmar; Seoul, South Korea; Tokyo, Japan; and Yangon, Myanmar. Due to large study area, two annotation strategies are used sparse labeling and dense labeling. Sparse labeling annotates particular objects while dense labeling annotates each and every pixel.

Chapter 6

Methodology

We consider two domains Ds(source domain) and Dt(target domain) we pre-train a semantic segmentation model and use it as backbone to construct a Siamese network. We will use a U-Net which is a encoding decoding architecture as the backbone for the land cover classification and domain adaption. Since there are 4 channels in the Five million pixel dataset we use a kernel of 3x3x4 is used. Due to lack of annotated data we are inspired to use pseudo labeling. Siamese network is used to collect pseudo labels of which the number is dynamically increased with training iterations. Shannon Entropy is used to measure the confidence of each pixel in target domain. Lower entropy represents higher classification confidence.

$$E^{(h,w)}xT = -\frac{1}{\log(K)} \sum_{k=1}^{K} x^T \log(F(h; w; k)x^T)$$

Pixels with high confidence is extracted and softmax function is used to obtain the category probability vector:

$$P_{xT}^{(h,w)} = \frac{\exp(F_{xT}^{(h,w)})}{\sum_{k=1}^{K} \exp(F_{xT}^{(h,w,k)})}$$

The pseudo-labeling will change dynamically with each iteration.
Class balancing is a common strategy for the training of semantic segmentation models, but it is rarely used in UDA approaches because the category information in the target domain is unknown. Since we assign pseudo-labels to D_T , it is possible to reduce the distribution bias caused by unbalanced categories through this strategy.
For D_S , we count the ratio of the number of pixels in each category to the number of all labeled pixels. Supposing that the ratio of the class k is µk, its weight is

$$W_k = \frac{1}{\log(1 + \mu_k)}$$

Then, the loss function of the DT branch is calculated as

$$Loss_{DT} = \sum_{n=1}^{N} W_{ln} F_{CE}(l_n, P_n^{xT})$$

where $F_{CE}()$ is the Cross Entropy loss function, l_n and P_n^{xT} denote the pseudo-label and the category probability vector of the nth pixel selected from x_T , respectively. If there are errors in pseudo-labels, a small number of mistakes may eventually lead to a relatively large bias during the iterative training. When gradually learning the distribution of D_T , to maintain the discrimination of the network for the true labels, we adopt joint learning of both the D_S branch and the D_T branch. The overall loss function of the Siamese network is

$$Loss = Loss_{D_S} + Loss_{D_T}$$

where $Loss_{D_S}$ is calculated by all pixels of x_S and is also applied with class-balanced weighting. When the training of the Siamese network is completed, forward propagation is performed on only one of the branches during the inference phase.

Chapter 7

Conclusion

7.1 Discrepancies:

On areas like overpass, railway station , airport, garden land, park, and pond, these categories represent small percentages in the data set and they inherently easier to confuse because they cover a much smaller area in the cities compared to other categories . Errors occur in the forested mountains and at the mosaic borders. Small areas of mountains around Beijing are misclassified as water bodies, and some areas of mountains around Chengdu are misclassified as irrigated field. So basically overfitting is happening here. This is also due to spectral shifts.

7.2 Proposed scheme to possibly enhance the model:

1)Adjusting the depth of the U net model to reduce overfitting and possibly lead to a better classification.

2)Using Diluted Convolution in some layers to ger larger receptive field and to reduce down-sampling.

3)Recently, based on Sentinel satellite imagery, European Space Agency (ESA) and Google have released global 10 m land cover mapping projects, we can use this labeled data to train our model for locations like cities, railway stations, airports, etc,