STATS 402 - Interdisciplinary Data Analysis <Building Change Detection using CNN Model on Aerial Images: A Comparative Analysis of 2012 and 2016 Datasets> Milestone Report: Stage 2

1st Jingheng Huan
Duke Kunshan University
Beijing, China
jh730@duke.edu

2nd Luyao Wang
Duke Kunshan University
Suzhou, China
lw337@duke.edu

3rd Fanbin Xu
Duke Kunshan University
Chengdu, China
fx31@duke.edu

***Abstract*—Following the completion of Milestone Report 1, our group has made significant progress in our satellite image analysis project. In this report, we will discuss our current status, the techniques we have adopted, any adjustments to our technical route, our initial data preprocessing results, and our plans for the next two weeks.**

I. Current Status and Techniques Adopted

A. Dataset and Data Preprocessing

The dataset [1] we will use for this project covers an area where a 6.3-magnitude earthquake occurred in February 2011 and rebuilt in the following years. This dataset consists of aerial images obtained in April 2012 that contain 12796 buildings in 20.5 km^2 (16077 buildings in the same area in 2016 dataset).

The dataset we downloaded contains both the whole and cropped images cropped from the whole image of 2012 and 2016. The labelled images of changed buildings based on the whole images are provided (Fig. 1) while labelled images of changed buildings of each cropped image are not provided. However, the labeled images of changed buildings are necessary for training the model. Therefore, we manually crop the whole images to create corresponding labeled images of changed buildings.

We crop the images by 500×500 pixels and got 1200 images for 2012, 2016 and labelled images separately. Each image is named with three parameters, for example, "2016_0_1243.jpg" indicates the image is cropped from the whole image of 2016 and cropped from x coordinate 0 and y coordinate 1243 (top left corner of the cropped image).

B. Deep Learning Model

We first tried a semantic segmentation model for the 2012 images and labels. The result successfully identifies the region where buildings were built (Fig. 2).

This image segmentation model builds upon a ResNet50 FCN (fully convolutional network) model. We use PyTorch built-in torchvision.models.segmentation.fcn_resnet50 model and change the number of the classifier from the default 21 classes to 1 class since we are only doing a binary classification (Fig. 3).

For the change detection model, we applied the same ResNet model with shared weights for both image of year 2012 and year 2016 as a feature extractor of two images from two eras. We add a simple FCN after the output of the siamese ResNet feature extractor of to predict the difference image. The structure in Fig. 4 is a combination of two FCNs in Fig. 3 with shared weights and a sequential model of FCN is added after the output of the siamese FCN feature extractor.

II. Plan for Next Two Weeks

A. Dataset and Data Preprocessing

Further refine our image preprocessing techniques: We will explore additional image processing techniques and improve upon our existing methods to obtain more accurate and informative results from the satellite images.

We notice that the tones of the photographs from the two eras differ. We must preprocess the photos to reduce the impact of tone discrepancies in order to solve this problem. The first strategy we considered was to equalize the pixel values in the two sets of photos. In order to make the photos more comparable, the brightness and contrast of the images must be changed. To do this, we can employ strategies like histogram equalization or adaptive histogram equalization.

Another tactic is to use image registration techniques to align the two sets of images such that they are in the same coordinate system. This technique can help with the tonal discrepancies as well as the issue that the building is not always in the same location in the two photographs because we can observe a slight shift between the images that were shot at the two different times. Image registration can be done using a variety of methods, such as feature-based methods, intensity-based methods, and hybrid methods. Using

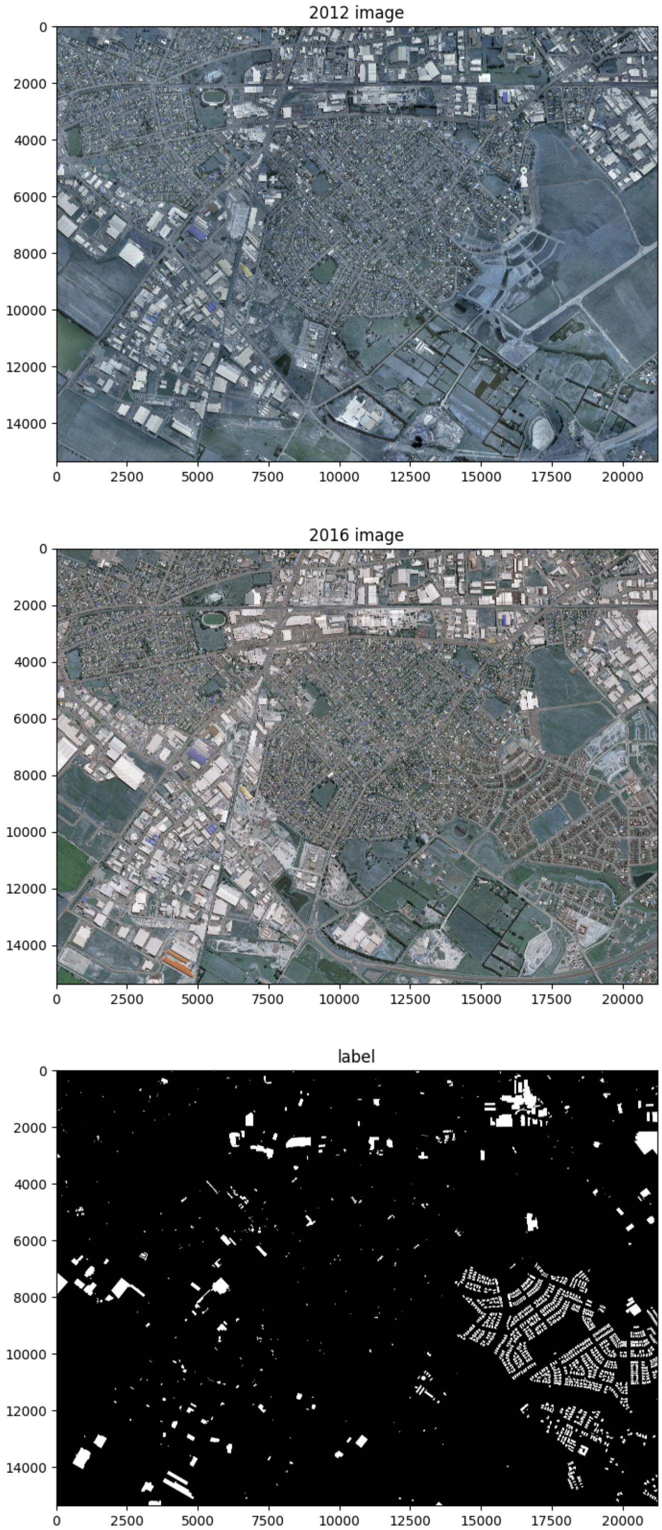

Fig. 1. The whole images of 2012, 2016 and change label [1].

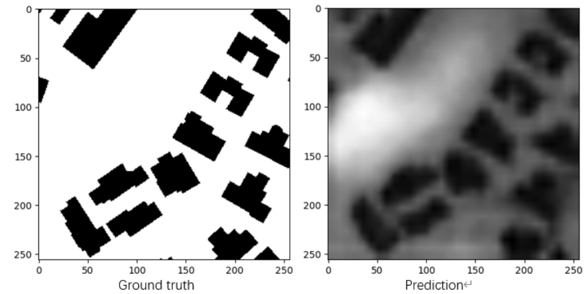

Fig. 2. Ground truth (left) and prediction (right) by the segmentation model.

```
FCN
├─IntermediateLayerGetter: 1-
│    └─Conv2d: 2-1
│    └─BatchNorm2d: 2-2
│    └─ReLU: 2-3
│    └─MaxPool2d: 2-4
│    └─Sequential: 2-5
│    │    └─Bottleneck: 3-1
│    │    └─Bottleneck: 3-2
│    │    └─Bottleneck: 3-3
│    └─Sequential: 2-6
│    │    └─Bottleneck: 3-4
│    │    └─Bottleneck: 3-5
│    │    └─Bottleneck: 3-6
│    │    └─Bottleneck: 3-7
│    └─Sequential: 2-7
│    │    └─Bottleneck: 3-8
│    │    └─Bottleneck: 3-9
│    │    └─Bottleneck: 3-10
│    │    └─Bottleneck: 3-11
│    │    └─Bottleneck: 3-12
│    │    └─Bottleneck: 3-13
│    └─Sequential: 2-8
│    │    └─Bottleneck: 3-14
│    │    └─Bottleneck: 3-15
│    │    └─Bottleneck: 3-16
├─FCNHead: 1-2
│    └─Conv2d: 2-9
│    └─BatchNorm2d: 2-10
│    └─ReLU: 2-11
│    └─Dropout: 2-12
│    └─Conv2d: 2-13
├─FCNHead: 1-3
│    └─Conv2d: 2-14
│    └─BatchNorm2d: 2-15
│    └─ReLU: 2-16
│    └─Dropout: 2-17
│    └─Conv2d: 2-18
=============================
Total params: 35,306,818
Trainable params: 35,306,818
Non-trainable params: 0
Total mult-adds (G): 37.00
```

Fig. 3. Structure of the FCN model for semantic segmentation

```
SiameseNetwork
├─FCN: 1-1
│    └─IntermediateLayerGetter: 2-1
│    │    └─Conv2d: 3-1
│    │    └─BatchNorm2d: 3-2
│    │    └─ReLU: 3-3
│    │    └─MaxPool2d: 3-4
│    │    └─Sequential: 3-5
│    │    └─Sequential: 3-6
│    │    └─Sequential: 3-7
│    │    └─Sequential: 3-8
│    └─FCNHead: 2-2
│    │    └─Conv2d: 3-9
│    │    └─BatchNorm2d: 3-10
│    │    └─ReLU: 3-11
│    │    └─Dropout: 3-12
│    │    └─Conv2d: 3-13
│    └─FCNHead: 2-3
│    │    └─Conv2d: 3-14
│    │    └─BatchNorm2d: 3-15
│    │    └─ReLU: 3-16
│    │    └─Dropout: 3-17
│    │    └─Conv2d: 3-18
├─FCN: 1-2
│    └─IntermediateLayerGetter: 2-4
│    │    └─Conv2d: 3-19
│    │    └─BatchNorm2d: 3-20
│    │    └─ReLU: 3-21
│    │    └─MaxPool2d: 3-22
│    │    └─Sequential: 3-23
│    │    └─Sequential: 3-24
│    │    └─Sequential: 3-25
│    │    └─Sequential: 3-26
│    └─FCNHead: 2-5
│    │    └─Conv2d: 3-27
│    │    └─BatchNorm2d: 3-28
│    │    └─ReLU: 3-29
│    │    └─Dropout: 3-30
│    │    └─Conv2d: 3-31
│    └─FCNHead: 2-6
│    │    └─Conv2d: 3-32
│    │    └─BatchNorm2d: 3-33
│    │    └─ReLU: 3-34
│    │    └─Dropout: 3-35
│    │    └─Conv2d: 3-36
├─Sequential: 1-3
│    └─Conv2d: 2-7
│    └─BatchNorm2d: 2-8
│    └─ReLU: 2-9
│    └─Conv2d: 2-10
│    └─BatchNorm2d: 2-11
│    └─ReLU: 2-12
│    └─Conv2d: 2-13
│    └─BatchNorm2d: 2-14
│    └─ReLU: 2-15
├─Sigmoid: 1-4
======================================
Total params: 35,309,392
Trainable params: 35,309,392
Non-trainable params: 0
Total mult-adds (G): 74.06
```

Fig. 4. Structure of the siamese model for change detection

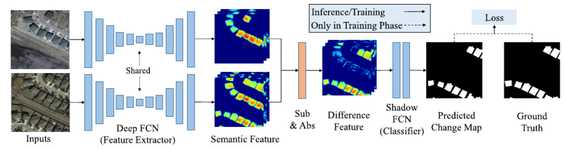

Fig. 5. Illustration of Deep Learning Network with Siamese network for Change Detection [3].

feature-based approaches, significant aspects in the images, like corners or edges, are identified and compared between the two images. Intensity-based methods use methods like mutual information or normalized cross-correlation to reduce the difference between the pixel intensities of the two images.

Aside from the two methods mentioned above, we will also consider cleaning the dataset we produced. Since we discover that using all photographs for training does not yield enough results, accuracy improves when we utilize only images that appear to show changes in buildings. As a result, a dataset that includes pictures of apparent changes to structures could be helpful for change detection.

B. Deep Learning Model

Evaluate our models: We have not evaluate our model quantitatively, so in next two weeks we will evaluate the performance of our machine learning and deep learning models using appropriate evaluation metrics, such as accuracy, precision, recall, and F1-score.

Iterate and refine our models: Based on the evaluation results, we will iterate and refine our models by tuning the parameters to improve their performance and ensure their effectiveness in analyzing satellite images.

For the FCN network, we plan to try skip connection [2] to build a more robust net.

We use FCNResNet50 as a feature extractor backbone for two images from two eras. Therefore, we can try using the pretrained model of FCNResNet50 and train only last several layers to make traning faster.

REFERENCES

[1] "-OpenDataLab-," Opendatalab.com, 2023. https://opendatalab.com/Building_change_detection_dataset (accessed Apr. 07, 2023).

[2] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," arXiv.org, 2014. https://arxiv.org/abs/1411.4038 (accessed Apr. 07, 2023).

[3] H. Chen, W. Li and Z. Shi, "Adversarial Instance Augmentation for Building Change Detection in Remote Sensing Images," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-16, 2022, Art no. 5603216, doi: 10.1109/TGRS.2021.3066802.