A REPORT
ON

'Scene Change Detection from a series of photographs'

BY

Name of the Student	ID Number
Adithya Shankar	*2017A3PS0205P*
Siddhant Khandelwal	*2017A7PS0127P*

AT
AI/ML (MMI) Laboratory

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

July, 2016

Acknowledgement

We would like to express our heartfelt gratitude towards Dr Kamlesh Tiwari, for giving us an opportunity to be part of the group working at AI/ML laboratory.

We would also like to thank MapMyIndia for providing us with exciting problems to work on, along with the infrastructure and the tools to build innovative solutions.

Without the help and contribution of the above-mentioned people, this project would not have been a success.

TABLE OF CONTENTS

1 Introduction

MapmyIndia has a large repository of Geo-tagged photographs taken at different time intervals. The objective of the project is to analyse, identify and present scene changes between photographs or a series of photographs taken at different timeframes.

1.1 Aim

The aim of this project is to "***Develop a CNN-based model that detects scene change between two different sets of images.***".

1.2 Literature

Scene change detection (SCD) is a fundamental task in the field of computer vision. The core idea of SCD is detecting changes in multiple images {I1, I2, IM} of the same scene taken at different times. Recently, the convolutional neural network framework has achieved outstanding performance in computer vision tasks.

The paper: Learning to Measure Changes: Fully Convolutional Siamese Metric Networks for Scene Change Detection fits the requirements of the project.

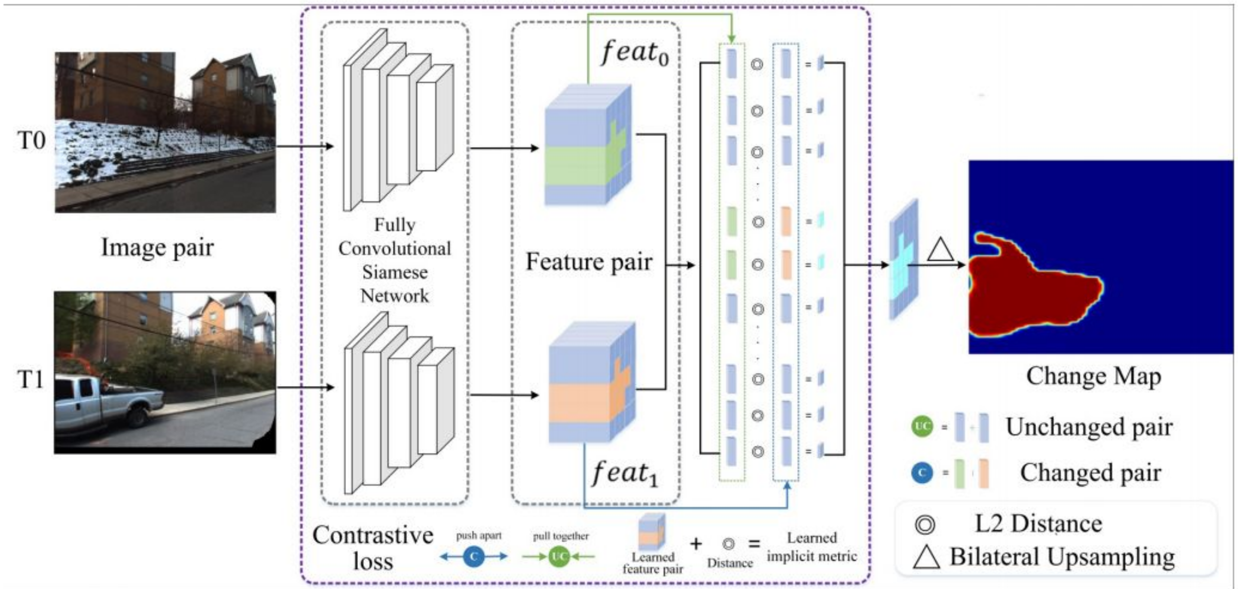
We have coded implementation of both the early-fusion architecture and the late-fusion architecture of the FCN models mentioned in the paper. We have trained this on the CD2014 Scene Change Detection dataset.

Given a pair of images, change detection aims to identify semantic changes at different times[2]. However, the critical challenge in this task is noisy changes generated by challenging factors such as varying illumination, shadows and camera viewpoint differences that are difficult to distinguish from semantic changes, making changes difficult to define and measure owing to the noisy changes and semantic changes that are entangled.

In our work, we implement a novel change detection framework, named the fully Convolutional siamese metric Network(CosimNet). Instead of simple classification, we leverage the intuitive idea of directly comparing a pair of images by customizing a discriminative implicit metric. It contains two parts: the deep features extracted from the Fully Convolutional Siamese Network (FCSN) and the predefined distance metric.

2 Model architecture

The model architecture is Convolutional siamese metric Network(CosimNet). Given a pair of images as input, we forward propagate the input through a full convolutional siamese network to generate feature pairs. Then, we utilize a simple predefined distance metric (l2 or cos) to measure the dissimilarity of the feature pairs. We named the unified processing including deep features and predefined distance metric as an implicit metric. To obtain a better implicit metric, we use the contrastive loss to bring together unchanged pairs and separate changed pairs.



3 IMPLEMENTATION

The backbone of CosimNet was based on DeeplabV2, whose last classification layer was removed. We trained all the models using a stochastic gradient descent algorithm with a momentum equal to 0.90 and the weight decay equal to 5e-5. All the code is implemented in the PyTorch platform. The code for our implementation can be found here.

We have referred the implementation here.

3.1 DATASET

The CDnet dataset consists of 31 videos depicting indoor and outdoor scenes with boats, trucks, and pedestrians that have been captured in different scenarios. It contains a range of challenging factors, including dynamic backgrounds, camera jitter, shadow, intern object motion, PTZ, and night video, which aims to solve foreground detection in complex outdoor conditions. In detail, we built a total of 91595 image pairs, which consist of a training set and a validation set with 73276 pairs and 18319 for each. All images were scaled to 512 × 512 during training. The preprocessed dataset is made available by the authors of the referred paper. The dataset can be obtained here.

The dataset that we received from MMI was significantly different from what we had expected. After a rigorous analysis of the dataset, we realised that the dataset required more details than what was provided. We requested MMI to re-issue us the updated dataset. Our analysis is as follows:

- The reason for going ahead with the CD2014 Dataset and not the dataset provided by MMI is mainly due to our lack of understanding of the dataset and how it is structured:
  - Currently, the structure of the dataset is as follows:
    - Image_Change_Karnataka_Part1

                ● Image_Change_Karnataka_04042019
                    ○ F
                    ○ L
                    ○ R
                ● Data

        ■ Image_Change_Karnataka_Part2
            ● Image_Change_Karnataka_16022018
                ○ F
                ○ L
                ○ R
            ● Image_Change_Karnataka_31122018
                ○ F
                ○ L
                ○ R

    ○ If we take a look at the structure of Image_Change_Karnataka_Part2, this suggests that we have a series of photographs of the same set of locations captured in a time-series form and that we are supposed to detect changes in scenes between the two different dates, i.e. 16/02/2018 and 31/12/2018.

    ○ However, we went through the entire directory manually in search of a common point which we can consider as a start and have realised that either the series of photographs do not belong to the same path, or even if they are, there seems to be a lapse in the visited locations.

    ○ One possibility could be that we need to treat the entire Part 2 directory as one single path and Part 1 directory as the same path on a different date. However, the size of the directories suggests otherwise, Part 1 - 4.6 GB, and Part 2 - 6.3 GB. Therefore, this case is ruled out too.

- There seems to be a lapse in our understanding of the structure of the dataset. We will move ahead once we fully understand the intentions behind it.

- We realised that the files in the DATA folder of the Part 1 directory suggest that the dataset is intended to be browsed using software like ArcGIS. We are not sure if we are moving in the right direction and it would be generous of you to suggest ways of browsing the dataset efficiently.

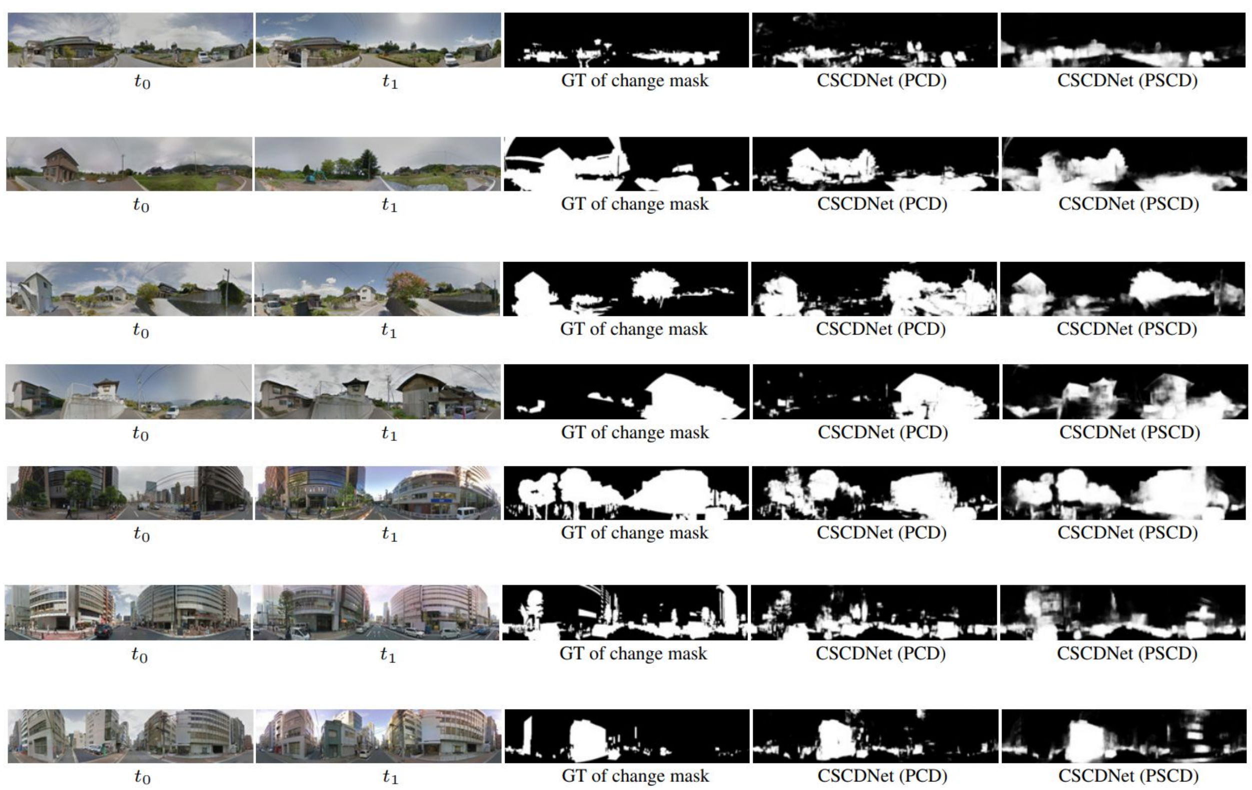
MMI helped us with this dataset problem by issuing us with a new dataset on 29th November 2019. This new dataset consisted of 2 folders of images taken from a moving vehicle. The photos in the two folders were taken at 2 different time stamps, one being in the year 2016 and the other being in the year 2018. The photos taken were of Hyderabad this time and not Karnataka.

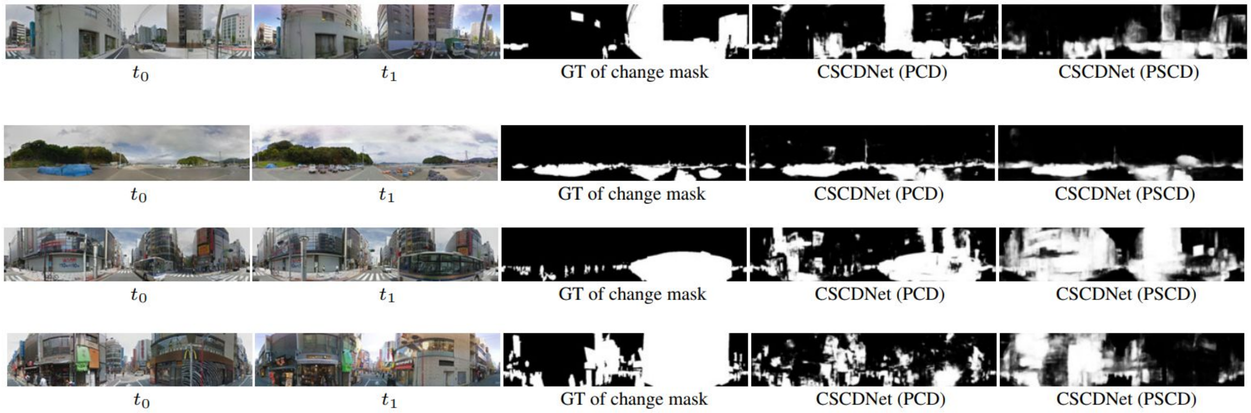
After a thorough analysis of the new datasets we encountered a few more problems that could not be solved given the limited time we had left to complete the project. Some of these problems were-

- Unlike before there were no separate folders having images taken from different views of the car i.e F,L,R view. The images were taken randomly and mixed in the folder. This prevented us from using techniques like image stitching to create a larger panoramic image which consists of photos from all 3 views taken at a given instant stitched together. This image could then be fed into the network to better help in the scene change detection task.
- There were a few instances where there was an abrupt discontinuity between the photos i.e. there was a location change and view change between successive photos.
- Like before there was no clear demarcation between test and train sets and we were unable to understand what changes between two objects would constitute a valid scene change detection.
- After manually going through both the datasets we came to the conclusion that the two folders consisted of images of two different locations i.e the path followed by the car in the two folders were different and hence we could not apply scene change detection because we did not have 2 photos of the exact same location taken at different time stamps.

4 RESULTS

In the figure given below, we show (1) change map and (2) prediction produced by the implementation of CoSimNet. These images are from the mentioned CD2014 dataset mentioned above.





5 References

Learning to Measure Changes: Fully Convolutional Siamese Metric Networks for Scene Change Detection

https://arxiv.org/pdf/1810.09111.pdf

Pytorch implementation of scene change detection - Fully Convolutional Siamese Network for Scene Change Detection

https://github.com/gmayday1997/SceneChangeDet

CDnet2014 Dataset

https://drive.google.com/drive/folders/1bUcUZcx8eRFZMsDuzVSo8ZkpLNhkEwNu?usp=sharing

Backbone model: deeplabv2

https://drive.google.com/file/d/1vma3tTX_ecKvInd91CWMEivbxhT5Xjfa/view?usp=sharing

The new dataset provided by MMI. (Year- 2016)

http://180.179.208.101/getFile.php?uname=BITS&pwd=gagidrothi

The new dataset provided by MMI. (Year- 2018)

http://180.179.208.101/getFile.php?uname=BITS&pwd=sleverirac