Implementation of Deep Learning architectures for Land Cover Change Analysis from High Resolution Satellite Images

Report of Summer Internship Project
From May 2021 to July 2021

Submitted by
B. Bharath Simha (181EC110)
V. Sai Krishna (181EC153)
B.S Trinesh Reddy (181EC108)

Under the guidance of
Dr. Shyam Lal
Dept of E&C, NITK, Surathkal

Dr. Sumam David
Dept of E&C, NITK, Surathkal

For the partial fulfillment for the award of the Degree of
Bachelor of Technology

DEPARTMENT OF ELECTRONICS &COMMUNICATION ENGINEERING NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA, SURATHKAL, MANGALORE - 575025
JULY 2021

Table of Contents

DECLARATION CERTIFICATE

We hereby ***declare*** that the Summer Internship project work entitled **"Implementation of Deep Learning architectures for Land Cover Change Analysis from High Resolution Satellite Images"** which are being submitted to the ***National Institute of Technology, Karnataka, Surathkal***, in partial fulfillment of the requirements for the award of the Degree of **Bachelor of Technology** in **Electronics & Communication Engineering** in the department of **ELECTRONICS AND COMMUNICATION ENGINEERING** is a ***bonafide*** report of the research work carried out by us. The material contained in this project has not been submitted to any other University or Institution for the award of any degree.

B. Bharath Simha Reddy (181EC110)

V. Sai Krishna (181EC153)

B.S Trinesh Reddy (181EC108)

Department of ECE

Place: NITK, SURATHKAL

Date: July 24, 2021

1. Abstract

Land-cover change detection has been a major driver of developments in the analysis of remotely sensed data or geo-spatial data and it is the task of classifying pixels or objects whose spectral characteristics are similar and allocating them to the designated classification classes, such as forests, grasslands, wetlands, barren lands, cultivated lands, and built-up areas. Various techniques have been applied to land cover classification, including traditional statistical algorithms and recent machine learning approaches, such as Neural network models and support vector machines etc.

Considering the Objective as understanding and comparison of different approaches and models for the prediction and performance of change detection in the satellite images, In this Paper we try to implement unsupervised and supervised approaches. Unsupervised approach like Principal Component Analysis, supervised architectures like UNet, UNet++ and Siamese Network models have been implemented. The Dataset used for the purpose is Levir-CD dataset and the predictions of unsupervised approach was better compared to deep learning techniques.

1.1 Introduction

Change detection process is identifying the differences in the state of an object or a natural phenomenon by observing at different times. It has been widely applied in numerous fields such as land cover and land use mapping, natural resource investigation, urban expansion monitoring, environmental assessment and rapid response to disaster events.

Traditional Change Detection (CD) methods can be divided into two categories.

a. Pixel - Based Change Detection (PBCD)
b. Object - Based Change Detection (OBCD)

In PBCD, a difference image (DI) is generated by directly comparing pixel spectral or textual values from which the change map (CM) is obtained by threshold segmentation or clustering.

OBCD has evolved from the concept of object-based image analysis (OBIA) which combines segmentation and spatial, spectral and geographic information along with analyst experience with image-objects in order to model geographic entities. The images are first segmented into disjoint and homogeneous objects, followed by comparison and analysis of bi-temporal objects. It has the ability to improve the identification of changes for the geographic entities found over a given landscape.

In the below implementations Pixel-based change detection is done.

1.2 Literature Survey

Detection of land cover changes has always been a problem of interest to scientists in recent times mainly because it helps identify regions of rapid change. These observations are of equal importance for scientists who battle climate change, or for others who constantly monitor habitat destruction.

Early on, unsupervised methods like PCA and KMeans clustering [2] were used to perform the task of change detection. With the advent of deep learning and availability of more data [5], slowly the transition to supervised methods began. Feature extraction with Convolution networks made possible development of deep learning models which were computationally expensive during the training process but quick and accurate during testing, unlike the unsupervised methods.

In some cases, more discriminative features can be learned by using designed deep learning models with available datasets, which is more beneficial for change detection Zhang et al. [6] utilized the Deep Belief Network(DBN) to learn abstract and invariant features directly from raw images, and then two-dimensional (2-D) polar domain and clustering methods were adopted to generate a CD map. Nevertheless, DBN, unlike CNN, has weak feature learning abilities.

Thus, Siamese CNN architectures with weighted contrastive loss [7] and improved triplet loss [8] were exploited to learn discriminative deep features between changed and unchanged pixels, then difference images were generated based on the Euclidean distances of deep features, and finally the a change map could be obtained by a simple threshold.

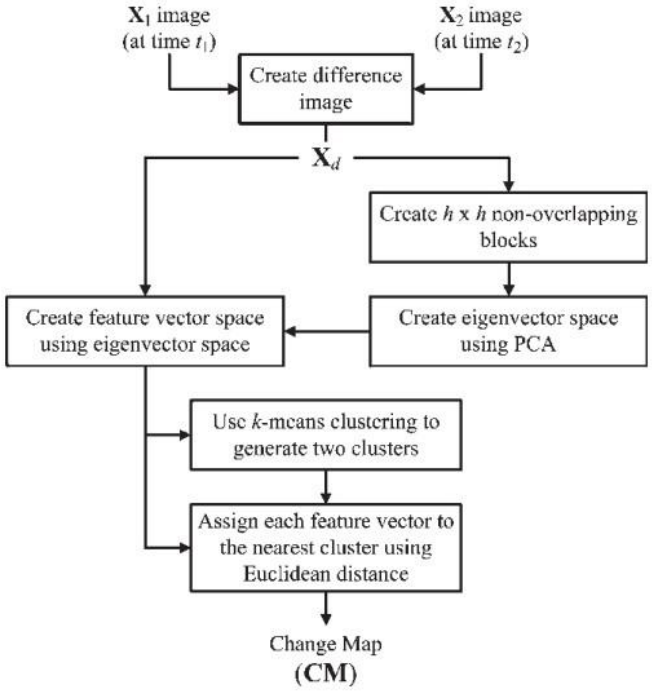
UNET architecture with slight modifications [1] could extract more features and give better results. The proposed architecture is a modified version of [6] on the dataset obtained from [5]. Instead of clustering and thresholding the features layers.They are up-sampled to predict the change map. Processing separately the bi-temporal images and concatenating the outputs to a single layer instead of concatenating the raw images is the underlying difference between the proposed architecture and the implemented networks.

1.3 Unsupervised Implementation

Unsupervised Learning is the implementation of algorithms, where the data points in the dataset are not classified/ labelled and the task here is to find the patterns or relations in the unlabelled data.

Clustering is a technique of dividing a dataset into groups/clusters based on their similarities or differences. Data points in the same group/cluster are as similar as possible and different from the data points present in different clusters/groups. K- means Clustering is a process of clustering the data points into k clusters/groups.

Principal Component Analysis is a statistical method used for feature extraction. It is a process of dimensional reduction without loss in the key information of the dataset points and is done by computing the principal components(uncorrelated variables), where the uncorrelated components are computed with the help of a set of correlated variables in the dataset.



Algorithm for Change detection using PCA

Algorithm:

The PCA algorithm works well for highly correlated data making it most suitable for images.

The first step is to find the image difference of bi-temporal images, such that the associated pixels with the land changes will be different compared to the pixels associated with unchanged areas. And then the image is standardized so that all the variables contribute equally to the changes in the model.

Creation of non - overlapping blocks is done by Partitioning the difference image into h×h non-overlapping blocks (where h≥2) and flatten them into row vectors. The image can be resized to make both dimensions a multiple of h.

Eigenvectors are computed for the covariance matrix and the corresponding eigenvalues are calculated. Sorting eigenvalues and corresponding eigenvectors in the decreasing order. Choose the first k vectors, where k is the number of dimensions you wish to have in the new dataset. In the last step, the data is transformed by re-orienting from the actual axes to the axes represented by the computed principal components.

As we know the objective of K-means clustering, with the newly transformed data computed by principal components performing the K- means clustering(where k = 2 in this case, as we need to check for changed and unchanged pixels). The pixel data points are clustered into two groups and then a binary change map can be created assuming 1 as change in the pixels/ land change and 0 as no change in pixels. It can be done by unsupervised ThresholdingAlgorithm for change detection using PCA.

1.4 Supervised Implementation

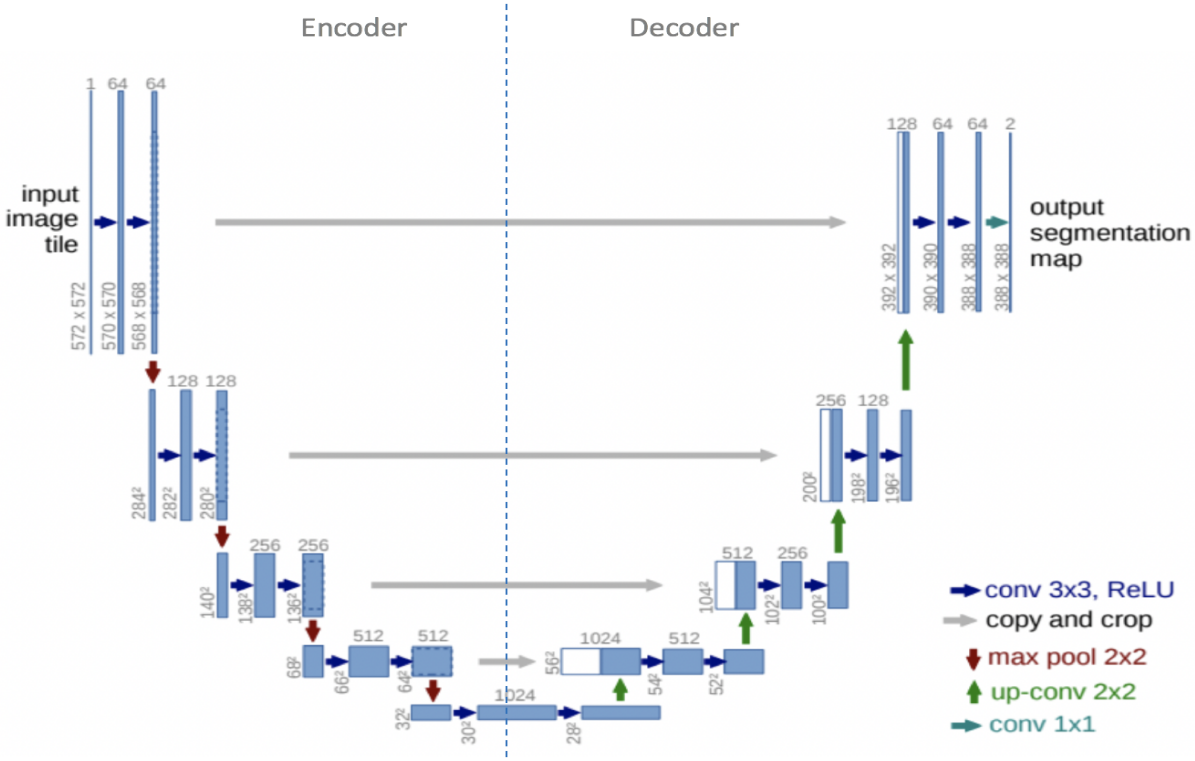
In the supervised implementation, the training samples are taken which are the areas on the ground for which there is Ground truth is known. The spectral signatures of the training areas are used to search for similar signatures in the remaining pixels of the image and are classified accordingly. Since the selection of the training samples and a biased selection can badly affect the accuracy of classification expert knowledge is required.

Popular techniques include the Maximum likelihood principle and Convolutional neural network. The Maximum likelihood principle calculates the probability of a pixel belonging to a class (i.e. feature) and allots the pixel to its most probable class. Newer Convolutional neural networks based methods account for both spatial proximity and entire spectra to determine the most likely class.

2 Benchmark Models

Implementation of UNet, UNet++ and a simpler version of the Siamese network model has been done.

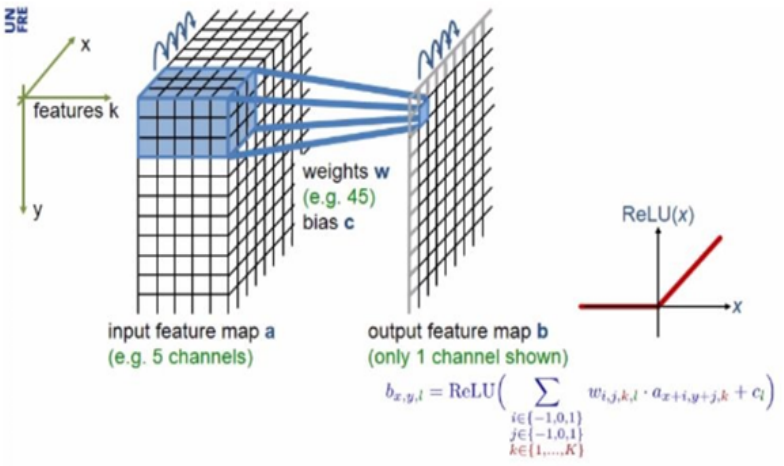
2.1 U-Net Architecture



U-net architecture (example for 572x572 pixels in the lowest resolution)

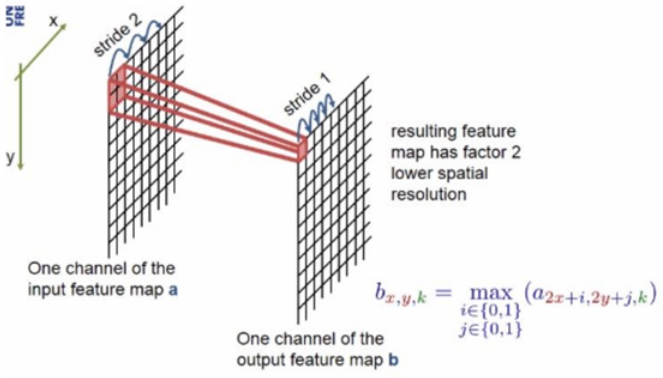
U-Net is a fully convolutional neural network architecture. The network consists of two paths, one is the encoding path and other is the decoding path.

Encoding path consists of repeated application of convolution layers with filter of size 3X3 without padding followed by 2X2 max pooling layer with stride =2 for downsampling and activation function is “RELU”.During the encoding process, the spatial information is reduced while feature information is increased.



U-Net Convolutional Path

Decoding path combines feature and spatial information through a sequence of upconvolutions up-convolutions and concatenations with high-resolution features from the contracting path. Every step in the decoding path consists of an upsampling of the feature map followed by a 2x2 convolution (up-convolution) that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU.



Max Pooling

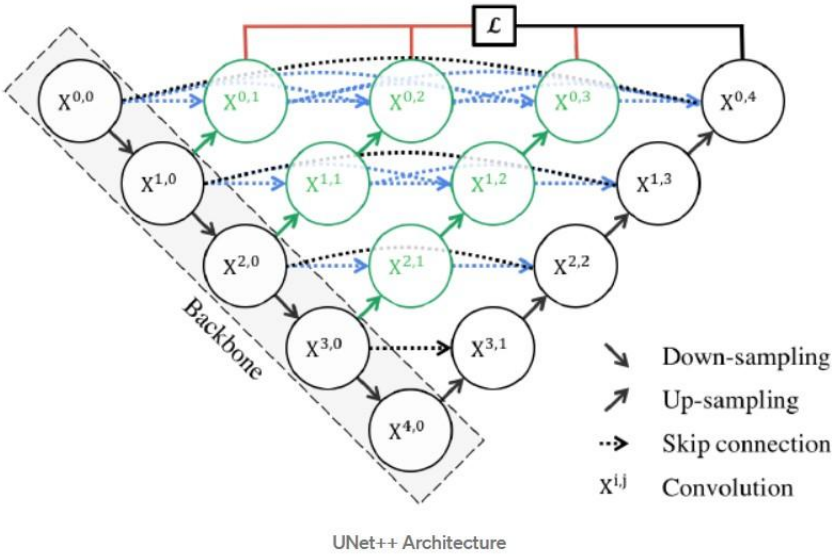
At the last layer , a 1x1 convolution is used to map each 64 component feature vector to the desired number of classes. In total the network has 23 convolutional layers.

2.2 U-Net ++ Architecture

Addition to the original U-Net a Dense block and convolution layers between the encoder and decoder in this architecture. The improvements this architecture has than the previous are better skip pathways, dense skip connections and deep supervision.

Having convolutional layers on skip pathways leads to decrease in the semantic gap between encoder and decoder feature maps.
Having dense skip connections on skip pathways improves gradient flow.

Having deep supervision enables model pruning. It improves or in the worst case achieves comparable performance to using only one loss layer.



UNet++ Architecture

Fig -1

1. Redesigned skip pathways (shown in green)
2. Dense skip connections (shown in blue)
3. Deep supervision (shown in red), as in the below figure.

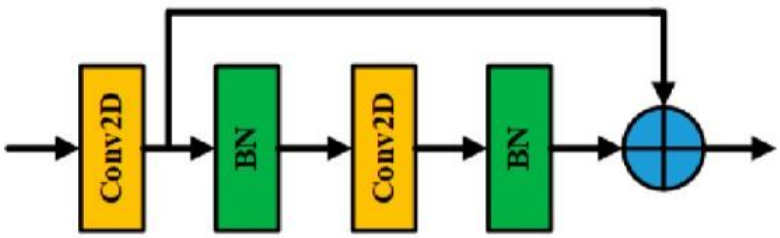


Fig 2

UNET++ was first designed and applied in 2019. The nested dense skip pathways have great benefits for extracting multi-scale feature maps from multi-level convolution pathways, which is similar to the UNET architecture.

The model architecture is shown in Fig.1. Each convolution block has a set of two Convolution2D and Batch Normalization (BN) operations as shown in Fig.2. The semantic levels of the encoder feature maps are closer to those in the corresponding decoder part, which facilitates the optimization of the optimizer.
Assume x(i,j) represents the output of node X(i,j), where i denotes the i'th down-sampling layer along the encoder way, j denotes the j'th convolution layer along the skip pathway. The accumulation of feature maps by x(i,j) can be expressed as:

$$x_{(i,j)} = \begin{cases} H(_{(i-1,j)}, & \text{if } j = 0 \\ H([[x^{(i,k)}]_{(k=0)}^{(j-1)}, \bigcup(x^{(i+1,j-1)})]), & \text{otherwise} \end{cases}$$

where H(.) represents a convolution operation followed by an activation function, S(.) is an up-sampling layer, and [.] denotes the concatenation operation.

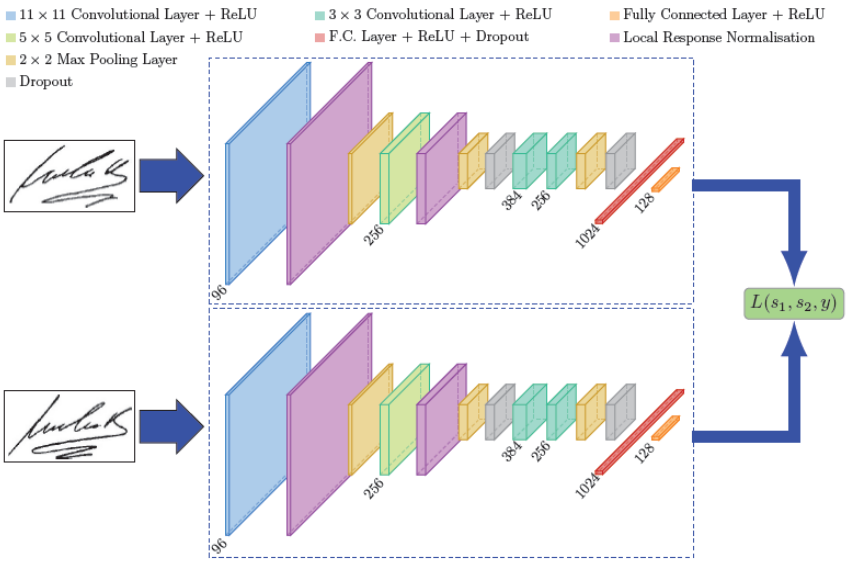
In general, nodes at level j=0 receive only one input from a previous downsampling layer, while nodes at level j>0 receive j +1 inputs from both the skip pathways and the up-sampling layer. It is noteworthy that residual modules are adopted in the convolution unit, which facilitates better convergence abilities for the deep networks.

As seen in Fig 2, a 2-D convolution layer (Conv2D) is implemented first, which is followed by a batch normalization layer. Then, a further Conv2D and batch normalization layer is applied. Finally, the output is generated by adding the outputs from the second BN layer and the first Conv2D layer. It should be noted that scaled exponential linear units (SeLUs) is adopted as the activation function instead of ReLU.

Another major difference is the multilevel full-resolution feature maps-generating strategy. Only a single-level feature map is generated in the UNET architecture through the pathway. Thus, the strengths of the four full-resolution feature maps are combined to generate a one-hot encoded Change Map.

2.3 Siamese Architecture

In the modern Deep learning era, Neural networks are almost good at every task, but neural networks rely on more data to perform well. But, for certain problems like face recognition and signature verification, and land cover change detection we can't always rely on getting more data.



Siamese network used in signet

To solve this kind of task we have a new type of neural network architecture called Siamese Networks.

It uses only a few numbers of images to get better predictions. The ability to learn from very little data made Siamese networks more popular in recent years.

A Siamese Neural Network is a class of neural network architectures that contain two or more identical subnetworks. 'identical' here means, they have the same configuration with the same parameters and weights. Parameter updating is mirrored across both sub-networks. It is used to find the similarity of the inputs by comparing its feature vectors, so these networks are used in many applications.

Below is the simple implementation of Siamese Architecture, where base layers are common layers for two input images, which consists of 5 convolutional layers, and the feature maps obtained from the base layers are then used to find the difference between them and a dense layer is used as a final layer with softmax activation.

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 512, 512, 3)	0	
input_2 (InputLayer)	[(None, 512, 512, 3)	0	
conv2d (Conv2D)	(None, 512, 512, 64)	1792	input_1[0][0]
conv2d_5 (Conv2D)	(None, 512, 512, 64)	1792	input_2[0][0]
conv2d_1 (Conv2D)	(None, 512, 512, 64)	36928	conv2d[0][0]
conv2d_6 (Conv2D)	(None, 512, 512, 64)	36928	conv2d_5[0][0]
conv2d_2 (Conv2D)	(None, 512, 512, 64)	102464	conv2d_1[0][0]
conv2d_7 (Conv2D)	(None, 512, 512, 64)	102464	conv2d_6[0][0]
conv2d_3 (Conv2D)	(None, 512, 512, 32)	51232	conv2d_2[0][0]
conv2d_8 (Conv2D)	(None, 512, 512, 32)	51232	conv2d_7[0][0]
conv2d_4 (Conv2D)	(None, 512, 512, 16)	528	conv2d_3[0][0]
conv2d_9 (Conv2D)	(None, 512, 512, 16)	528	conv2d_8[0][0]
lambda (Lambda)	(None, 512, 512, 16)	0	conv2d_4[0][0] conv2d_9[0][0]
dense (Dense)	(None, 512, 512, 2)	34	lambda[0][0]

Summary of Implemented Siamese Network Model

3. Training and Implementation Details

3.1 Dataset

The data used is the Levir-CD dataset. 445 training and 128 validation bi-temporal images resized to (512,512,3) are used for training and validation set respectively.

3.2 Pre-processing

As preprocessing step images are normalized to (0-1 range) and the corresponding ground truth change map is one-hot encoded, i.e. the corresponding mask image is one hot encoded to two classes i.e changed and not changed.

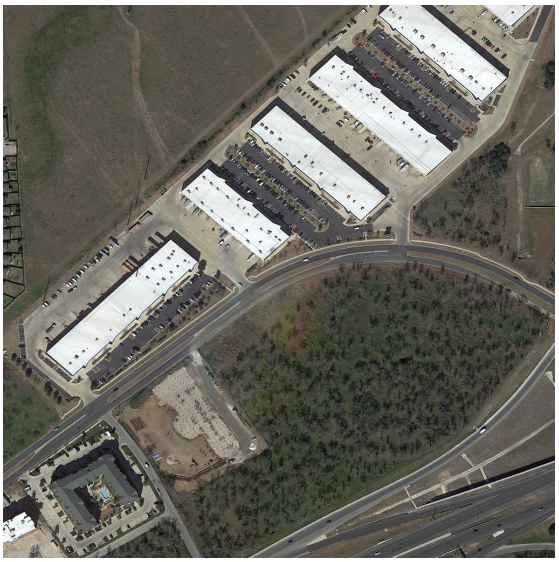
1. For UNet, UNet++ models the shape of input is (512,512,6), where the bi-temporal images are concatenated and sent to the model.
2. For the Siamese model, the shape of input is (512,512,3).

4.1 Simulation Results

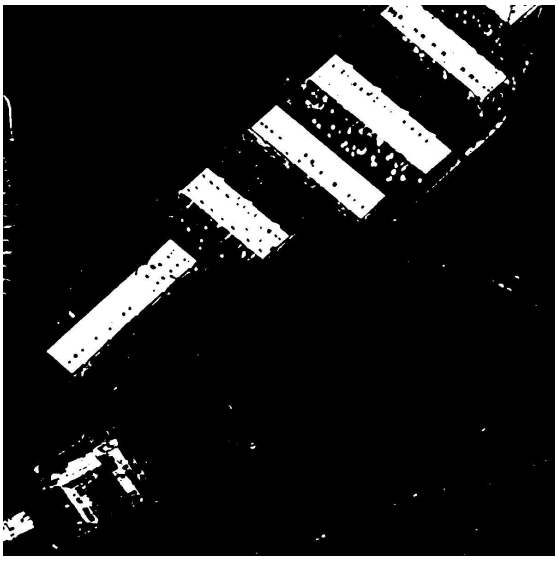
4.1.1 Unsupervised:

1.1 :

A



B

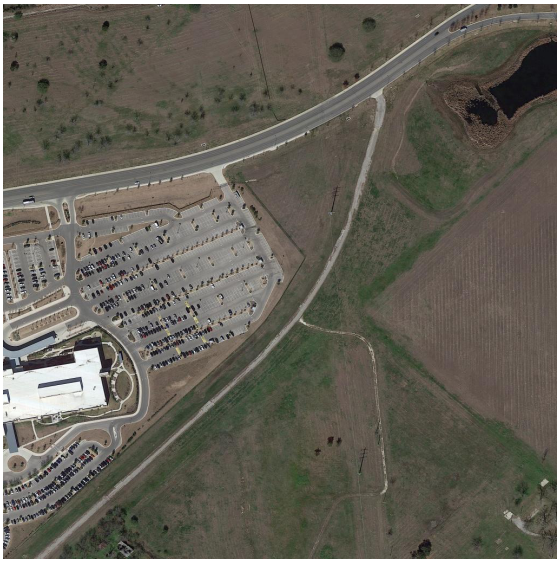


Predicted

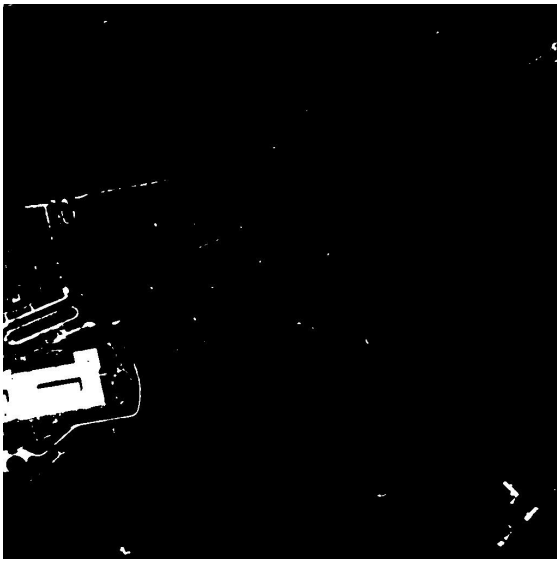
Ground Truth

1.2 :

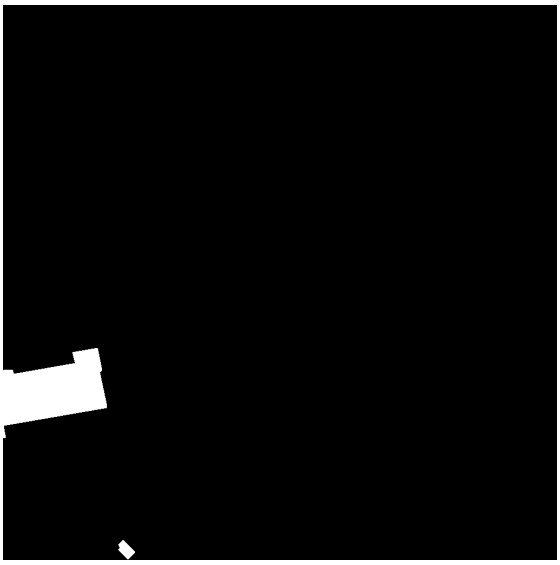
A



B



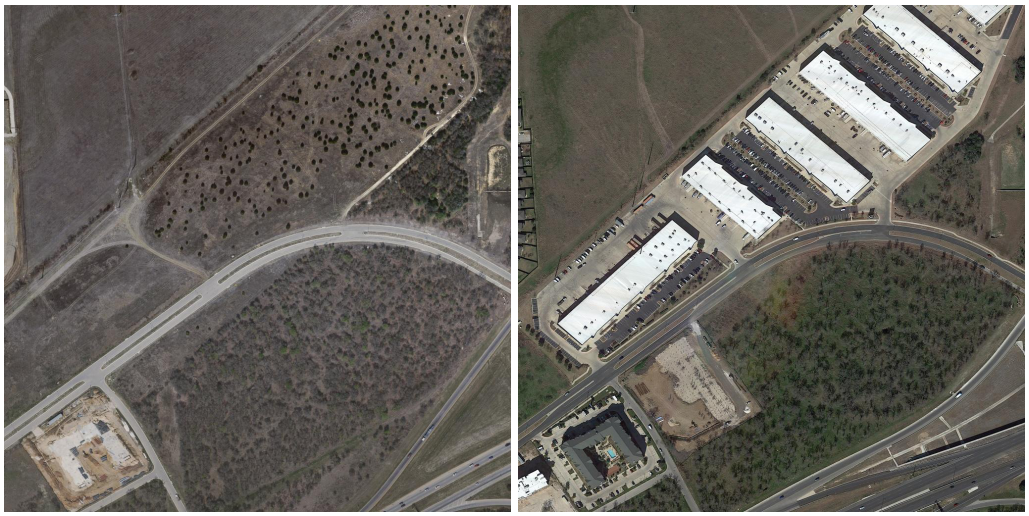
Predicted



Ground Truth

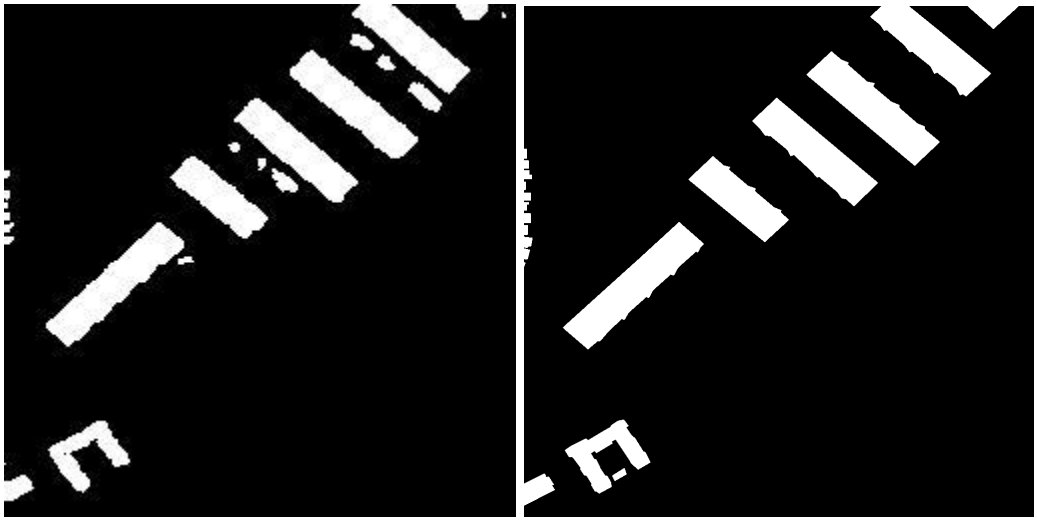
4.1.2. Supervised Implementation

2.1 :



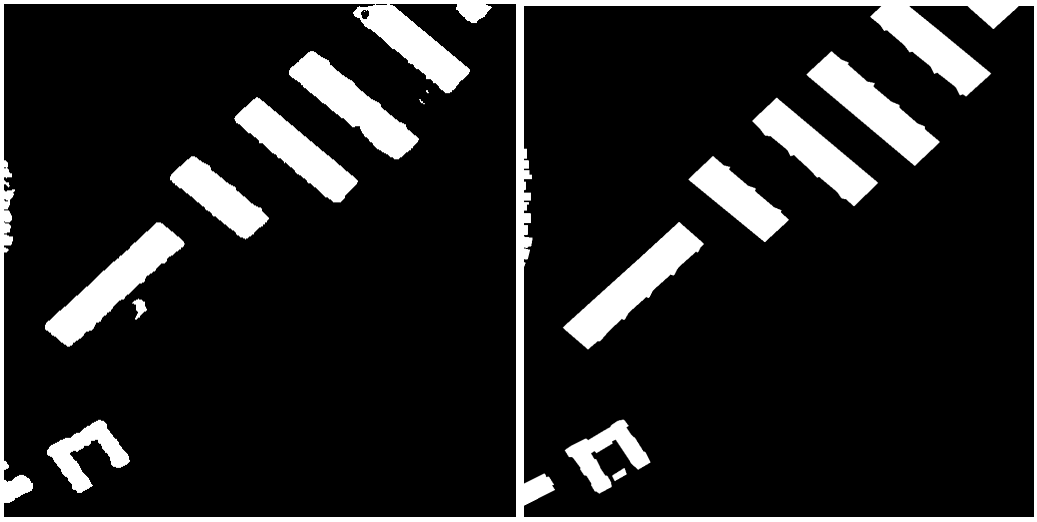
A **B**

a. UNet



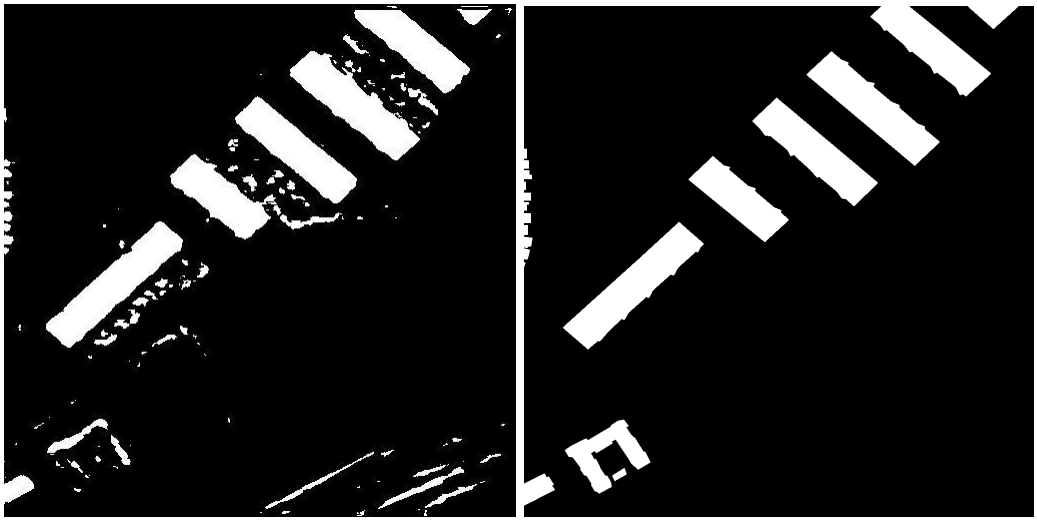
Predicted **Ground Truth**

b. UNet++



Predicted **Ground Truth**

c. Siamese (Simpler implementation of network)



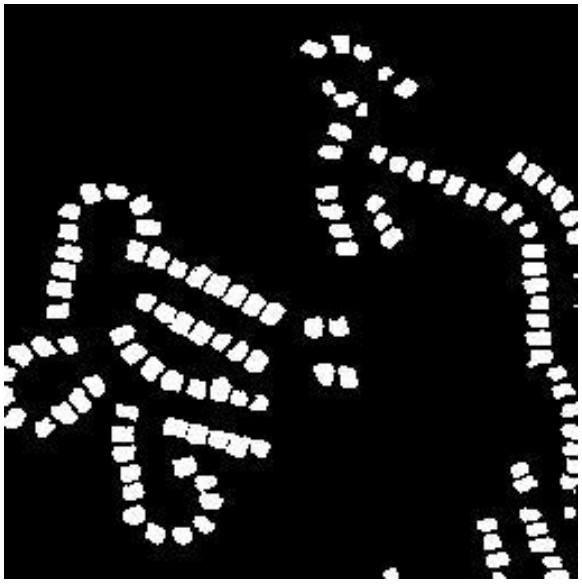
Predicted **Ground Truth**

2.2 :

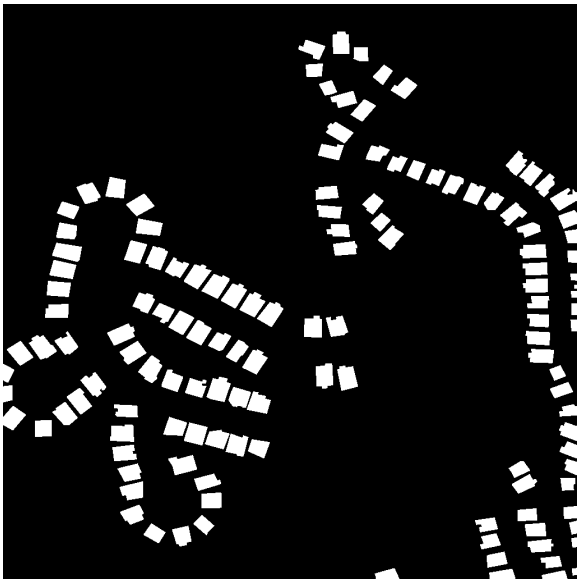
A

B

a) UNet

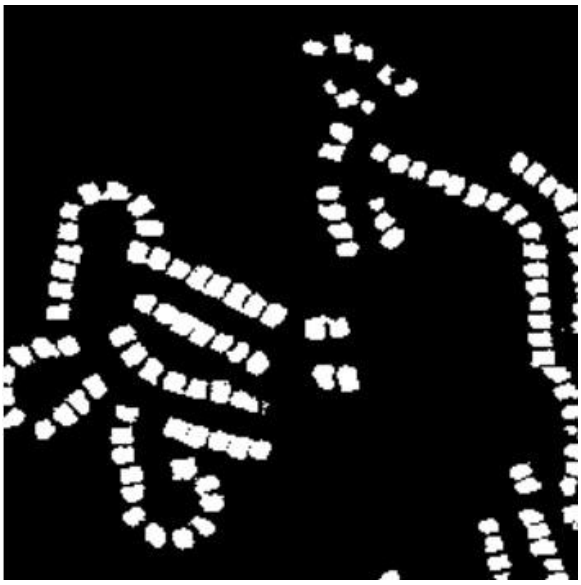


Predicted

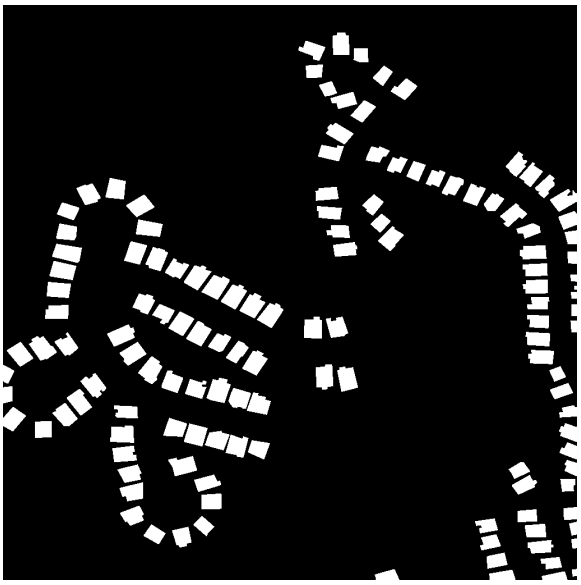


Ground Truth

b) UNet++

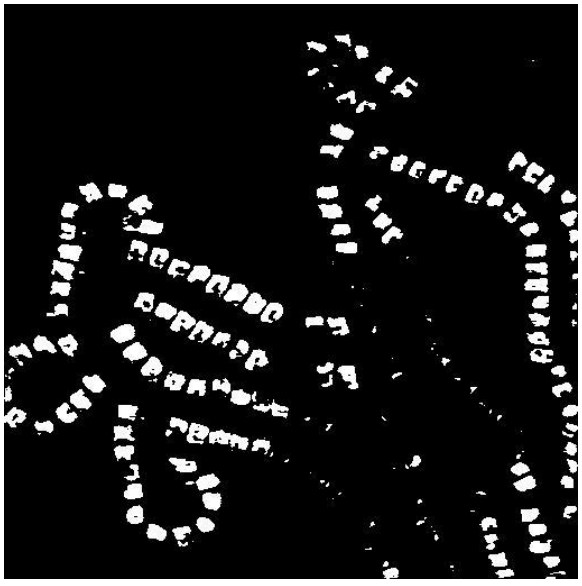


Predicted

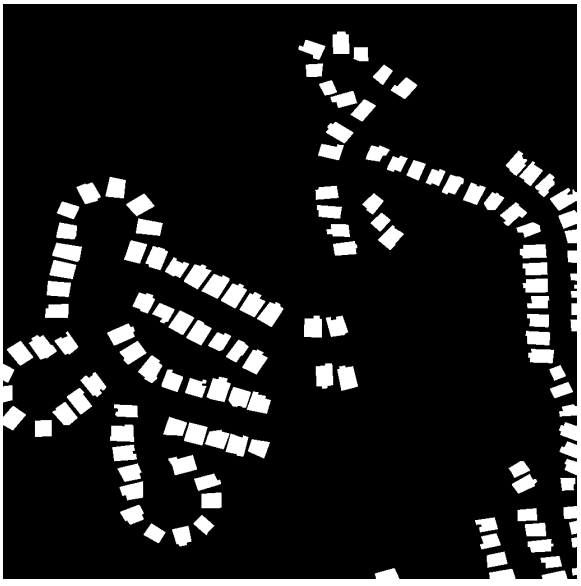


Ground Truth

c) Siamese Network



Predicted

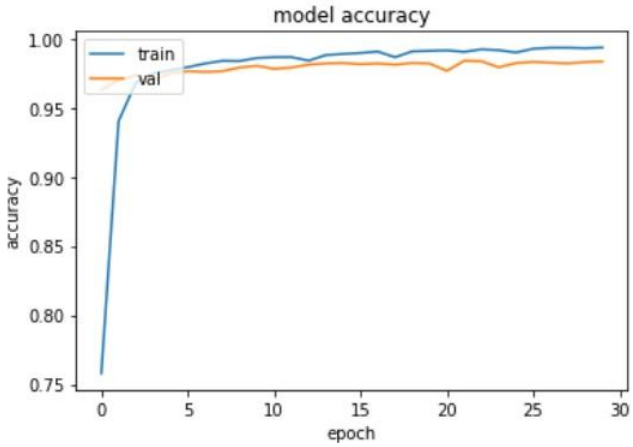


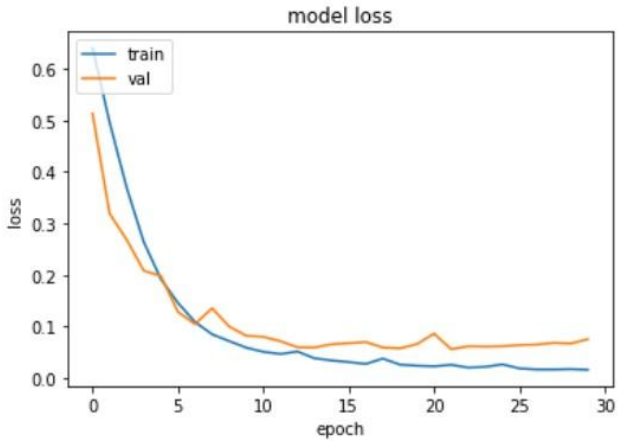
Ground Truth

4.2 Observations and Metrics Comparison

Accuracy and Loss graphs

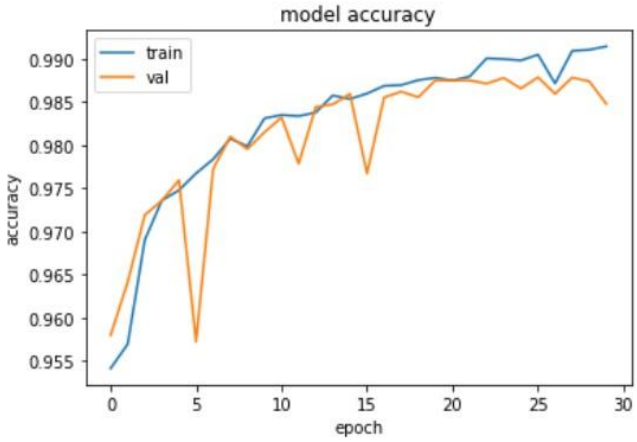
a. UNet

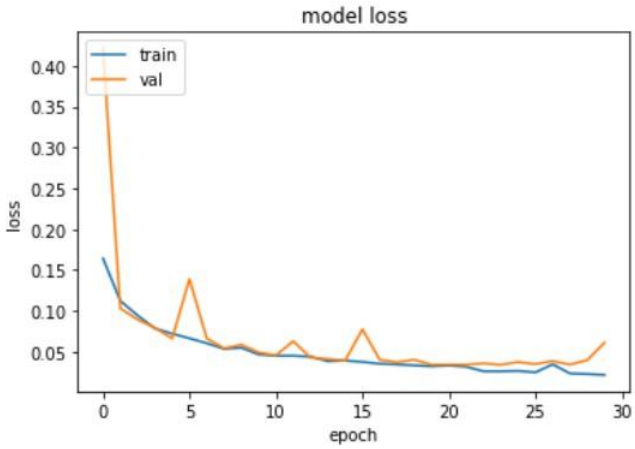




Total params of UNet Model: 31,054,258

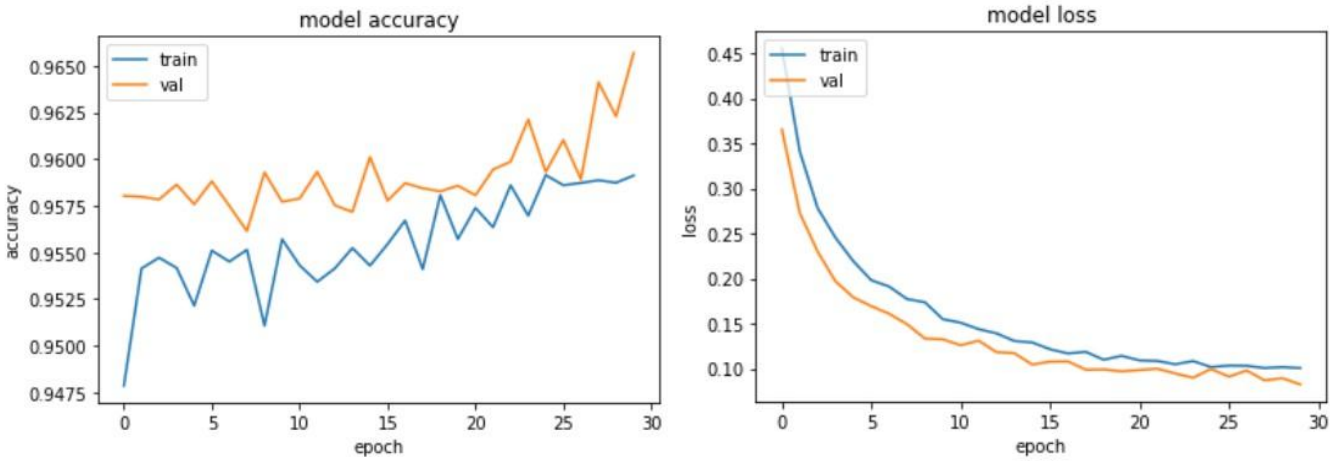
b. UNet++





Total params of UNet++ Model: 4,882,914

c. Siamese



Total params of Siamese Network Model: 385,922

Metrics Table :

	UNet	UNet++	Siamese
Accuracy	0.98388	**0.98901**	0.96046
Loss	0.0572	**0.0342**	0.168
Precision	0.9706	0.9817	0.9549
Recall	0.9706	0.9817	0.9549

5.1 Conclusion :

Implementation of PCA K-means (Unsupervised method) as well as UNet, UNet++, Siamese Network Architectures (Supervised Methods) has been done successfully and the model's accuracy, loss, precision, and recall for various loss functions and different activations for the last layer as been computed and compared as shown in the metrics table. The loss function which we used for the models is "Binary Crossentropy" and activation of the last layer is "Softmax" has provided good predictions compared to other changes.

Overall we were able to effectively generate the change maps for the bi-temporal images using the above models and among all the models UNet++ has shown good accuracy and loss, as well the predictions of the change map are good compared to the other models.

6.1 References

1. End-to-End Change Detection for High Resolution Satellite Images Using Improved UNet++, Daifeng Peng, Yongjun Zhang and Haiyan Guan

2. Unsupervised Change Detection in Satellite Images Using Principal Component Analysis and k-Means Clustering, Turgay Celik.

3. Learning to Measure Changes: Fully Convolutional Siamese Metric Networks for Scene Change Detection, Enqiang Guo, Xinsha Fu, Jiawei Zhu, Min Deng, Yu Liu, Qing Zhu, and Haifeng Li

4. Turgay Celik, "Unsupervised change detection in satellite images using Principal Component Analysis and K-means clustering", IEEE Geoscience and Remote Sensing Letters, Vol. 6, No.4, October 2009.

5. Change Detection inRemote Sensing Images using Conditional Adversal Networks, M.A.Lebedev, Yu.V.Vizilter, O.V.Vygolov, V.A.Knyaz, A.Yu.Rubis.

6. Zhang, H.; Gong, M.; Zhang, P.; Su, L.; Shi, J. Feature-level change detection using deep representation and feature change analysis for multispectral imagery. IEEE Geosci. Remote Sens. Lett. 2016, 13, 1666–1670

7. Yang Zhan, Kun Fu, Menglong Yan, Xian Sun, Hongqi Wang, and Xiaosong Qiu, "Change detection based on deep siamese convolutional network for optical aerial images," IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 10, pp. 1845–1849, 2017

8. Zhang, M.; Xu, G.; Chen, K.; Yan, M.; Sun, X. TripletBased Semantic Relation Learning for Aerial Remote Sensing Image Change Detection. IEEE Geosci. Remote Sens. Lett. 2019, 16, 266–270

9. Zhan, Y.; Fu, K.; Yan, M.; Sun, X.; Wang, H.; Qiu, X. Change detection based on deep siamese convolutional network for optical aerial images. IEEE Geosci. Remote Sens. Lett. 2017, 14, 1845–1849.

10. https://towardsdatascience.com/siamese-networks-line-by-line-explanation-for-beginners-55b8be1d2fc6

11. https://towardsdatascience.com/understanding-semantic-segmentation-with-unet-6be4f42d4b47

12. https://towardsdatascience.com/biomedical-image-segmentation-unet-991d075a3a4b

13. “Fully Convolutional Siamese Networks For Change Detection”, Rodrigo Caye Daudt, Bertrand Le Saux, Alexandre Boulch