Augmented Convolutional Neural Networks for Remote Sensing Change Detection

Sarvesh Patil
sarveshpatil@vt.edu

Pranjal Ranjan
pranjalranjan@vt.edu

Badhrinarayan Malolan
badhrinarayan@vt.edu

Ankit Parekh
ankitparekh@vt.edu

Abstract

With the advent of advanced remote sensing techniques, a large volume of remotely sensed data is being produced, potentially transforming multiple applications and fields. Deep learning methods have recently gained popularity in utilizing this stream of data, with continuous improvements taking place in both the architectures used for training and preprocessing techniques. This work aims to put forward some new ideas in the intersection of these two areas concerning training deep learning models for change detection. The emphasis is on the usage of signal processing ideas such as edge detection and multiresolution analysis, as well as the popular deep learning paradigm of transfer learning for augmenting classical change detection models. The newly proposed training methods are evaluated on the LEVIR-CD+ dataset, and additionally, challenges such as dataset shift, few shot learning and noise injection are also explored to gauge their efficiency. The results indicate significant improvement in Intersection over Union (IoU) when augmented methods are compared with the unaugmented model on the original dataset, as well as the challenges considered for testing.

1 Introduction

The analysis of massive landscapes or areas of the earth is often very difficult to achieve on the ground without appropriate wide-scale image data obtained from a considerable height from the earth's surface. Remote sensing satellites have revolutionized this field of study by enabling the easy acquisition of this type of information. This data is used for various public interest applications such as weather forecasting, land planning, urban expansion, disaster management, etc. Deep learning-based methods such as CNNs have found success in computer vision applications and thus have become popular in the analysis of such images.

Our work focuses on the task of change detection in remote sensing images, where we compare images acquired by remote sensing techniques at two different points of time. We have chosen the LEVIR-CD+ dataset to carry out our experiments and test our model. Apart from baselining the dataset on classical change detection models, we also attempt to improve them by designing augmentations to these models in the form of additional input information for further processing, these augmentations include Canny Edge Maps, Haar MRA components and Imagenet pretrained Encoders.

From the perspective of a research study, we have utilized this task as an opportunity to tackle some of the prevalent problems in the domain of deep learning. Dataset shift is one of the most common issues that occur in most deep learning problems, where the distributions of train and test datasets differ resulting in performance degradation. So we have tested our LEVIR-CD+ trained model on the WHU dataset to observe its performance. Another shortcoming of large supervised learning models is that they require huge amounts of data for training. Hence we limited the training data of our model to various extents to test its Few-Shot Learning capability. Finally, noise added to the data is an adversary of deep learning models that is quite troublesome both at training and testing. We have tested our augmented networks against these problems to see how robust they are in comparison to ordinary deep-learning networks.

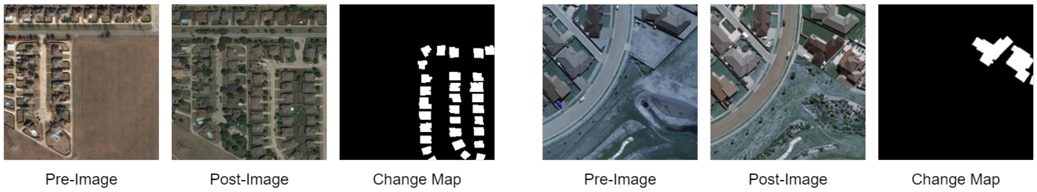
2 Related Work

M. Hussain et al.[1] have documented various change detection techniques and categorized them into two classes based on the unit of image analysis namely pixel-based and object-based. They review the most commonly used techniques in remote sensing, their applications, and related issues. Pixel-based approaches to change detection make a comparison between two images on a pixel by pixel basis, while object-based change detection approaches attempt to map the spatial environment in which the image exists by first identifying various classes of objects underlying a given image.

Jia Liu et al.[2] have developed an adaptive training strategy for effective learning by splitting the dataset into samples of different difficulty levels based on the amount of change, training the model on easier samples (less-changed) first and gradually feeding the difficult ones (more-changed). D. Peng et al.[3] proposed an effective encoder-decoder architecture for semantic segmentation capable of capturing changes with varying sizes effectively in complex scenes through a novel loss function was designed and an effective deep supervision strategy. H. Chen et al.[4] tackle the issue of illumination variations and misregistration errors between temporal samples through a self-attention mechanism to build useful Spatio-temporal dependencies between any two pixels at different time intervals and use it to generate more discriminative features. R.A. Ansari et al.[5] have used a Multiresolution based approach to extract textural features from SAR images to achieve change detection. S. Fang et al.[6] have used the combination of a Siamese network and a NestedUNet architecture to help preserve localization information in the deep layers of the network.

There has been use of the attention mechanism towards the end of improving results of change detection in remote sensing images. X. Peng et al.[7] use dense attention to enhance the texture extraction capability of the CNN and to model temporal differences. Furthermore, H. Chen et al.[8] and J. Chen et al.[9] have also addressed the problem of imbalanced data samples containing more unchanged samples than changed samples leading to model learning pseudo-changes by punishing attention to unchanged feature pairs and increasing attention to changed feature pairs through a weighted double-margin contrastive loss function which captures long-range dependencies. C. Zhang et al.[10] & Z. Zheng et al.[11] explore supervised approaches for change detection using bi-temporal and single-temporal samples.

3 Datasets



(a) LEVIR-CD+

(b) WHU-CD

Figure 1: Examples from both datasets

For the purposes of our experiments, we shall be using two datasets; a primary dataset, which will be used to train and test various model architectures, and an alternate dataset, which will be only used to test the models trained on the primary dataset.

Each instance comprises of a set of 3 images - Pre-image, Post-image, & Label. The details of the two datasets are as follows -

1. LEVIR-CD+ (Primary)[12] - This dataset consists of 985 VHR (0.5 m/pixel) bi-temporal Google Earth images and labels, with a size of 1024x1024 pixels, from multiple areas located in Texas, USA. We have pre-processed these to a 512x512 pixel resolution and cropped them into 256x256 images resulting into 3940 pairs and labels. We have split these into 2036 sets for training, 512 sets for validation, and 1392 sets for testing.
2. WHU-CD (Alternate)[13] - This dataset consists of around 2416 pairs of aerial images of independent buildings in Christchurch, New Zealand at a spatial resolution of 0.075m. We applied the same pre-processing steps used for the primary dataset and took 15% of these images of build a test dataset of 364 sets.

4 Evaluation Metric & Training Specifications

We have used IoU (Intersection over Union) as our metric for evaluation for all models. IoU is a commonly used evaluation metric for semantic image segmentation and defined as the ratio of true positives (TP) to the sum of true positives (TP), false positives (FP), and false negatives (FN) for an individual class. It quantifies the degree of overlap of the segmentation change map with the ground-truth label.

$$IoU = TP/(TP + FP + FN) \tag{1}$$

In order to ensure fairness in the training process we have kept our training specifications consistent across all our networks and experiments. We have used the categorical cross-entropy loss function over DICE and MSE to better facilitate convergence to the global minima. The loss function was used in conjunction with Adam adaptive learning rate optimizer which works well even with little tuning of hyper-parameters. The initial learning rate was set to 0.01 and IOU for validation set was monitored for learning rate reduction with a patience of 5 epochs. We kept a batch-size of 16 and used the Keras Machine Learning Library on a Tensorflow-backend to carry out all our experiments, training the models for 50 epochs with an early stopping patience of 15 epochs.

5 Experiment

5.1 Phase I

In this phase, we experiment with three fully convolutional network (FCN) architectures[14] to find the optimal architecture for our task that can be used as a base model for our proposed augmentations.

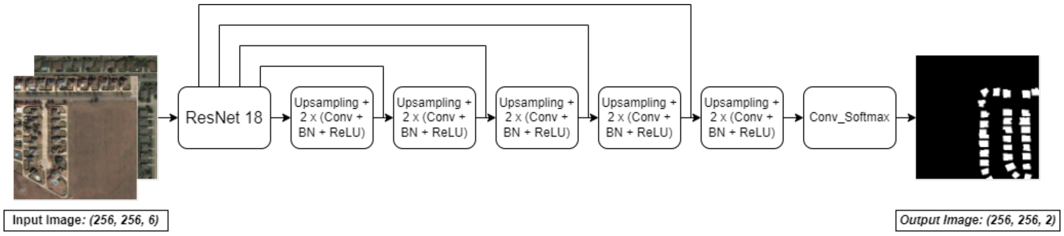

ResNet 18
Upsampling + 2 x (Conv + BN + ReLU)
Upsampling + 2 x (Conv + BN + ReLU)
Upsampling + 2 x (Conv + BN + ReLU)
Upsampling + 2 x (Conv + BN + ReLU)
Upsampling + 2 x (Conv + BN + ReLU)
Conv_Softmax
Input Image: (256, 256, 6)
Output Image: (256, 256, 2)

Figure 2: Early Fusion Model

The first FCN architecture (Figure 2) is the early fusion model, an adaptation of the U-Net model. In this network, the input is the channel-wise concatenation of the pre- and post- image while the output is the change map. The input is passed to a ResNet 18 encoder, which downsamples the image spatially while increasing its temporal range. The output of the encoder is then fed to a decoder network which upscales the image back to its original spatial dimensions. There are several skip connections from the encoder to various stages of the decoder similar to those present in the U-Net.

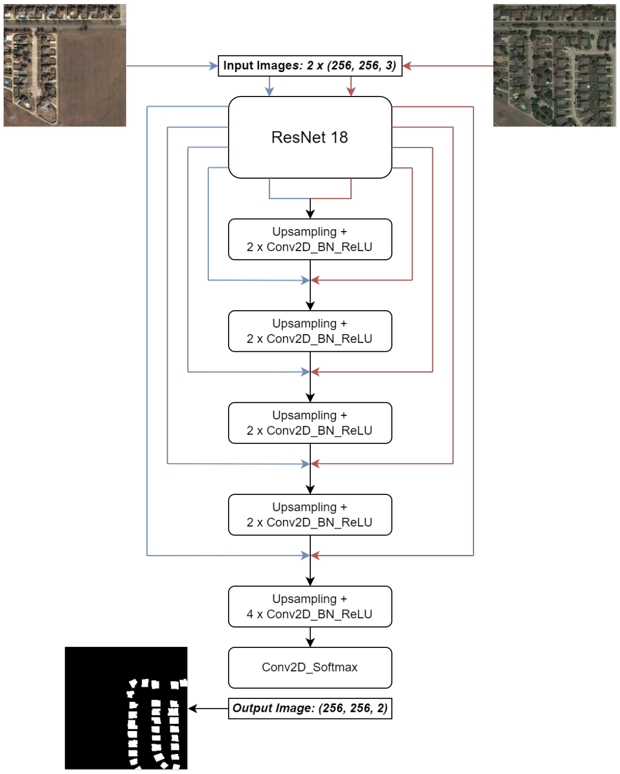


Figure 3: Siamese Conc. Model

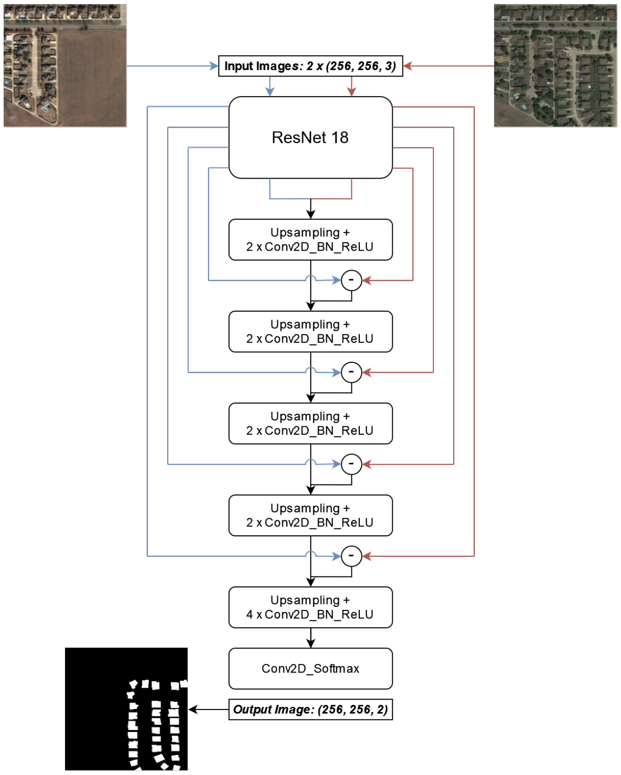


Figure 4: Siamese Diff. Model

The other two FCN architectures are variations of the Siamese network model. As opposed to in the early fusion model, where the pre- and post-images are concatenated before being fed to the model, in the Siamese model the images are passed seperately and the outputs of the ResNet 18 encoder are concatenated and passed to the decoder. The two variations differ in how the skip connections are fed to the decoder - in one variation, the corresponding skip connections are concatenated before being connected with the decoder. This is known as the Siamese-concatenation model (Figure 3). In another variation, the difference of the skip connections is fed to the decoder instead, and the model is called the Siamese-difference model (Figure 4).

For the experiment, we train and test all three models on the primary dataset. The model that performs the best is considered as the base model for further phases. The results can be found in 6.1.

5.2 Phase II

In this phase, we will apply various augmentation techniques to our chosen base model and test their effect on the change detection performance.

5.2.1 Edge Augmentation

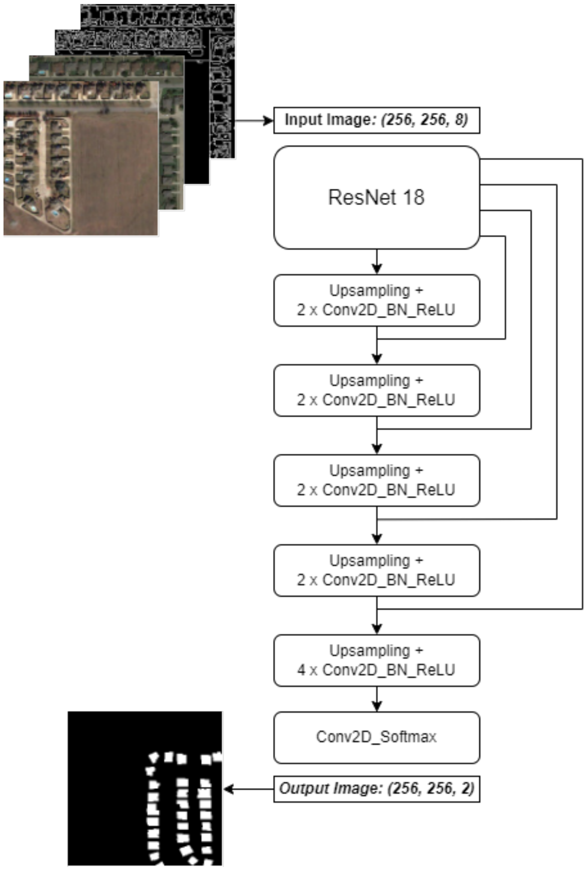

Input Image: (256, 256, 8)
ResNet 18
Upsampling + 2 x Conv2D_BN_ReLU
Upsampling + 2 x Conv2D_BN_ReLU
Upsampling + 2 x Conv2D_BN_ReLU
Upsampling + 2 x Conv2D_BN_ReLU
Upsampling + 4 x Conv2D_BN_ReLU
Conv2D_Softmax
Output Image: (256, 256, 2)

Figure 5: Edge augmentation

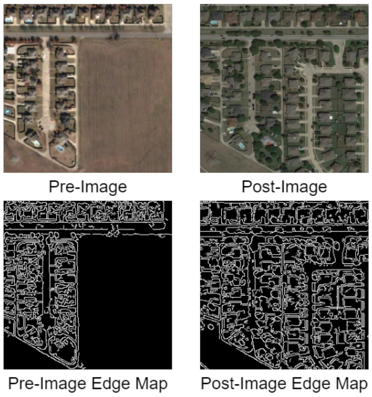


Figure 6: Edge maps of two input images

We use Canny Edge Detection to compute edge maps of the input pre- and post- images and feed them along with the input images. Given that our task is to detect change in building footprint, edge maps can help detect and localise building footprints in both pre- and post- images. This should help provide information to the model directly which it might have had to otherwise learn to extract.

We use a lower threshold of 100 and an upper threshold of 200 for the canny edge detector.

5.2.2 MRA Augmentation

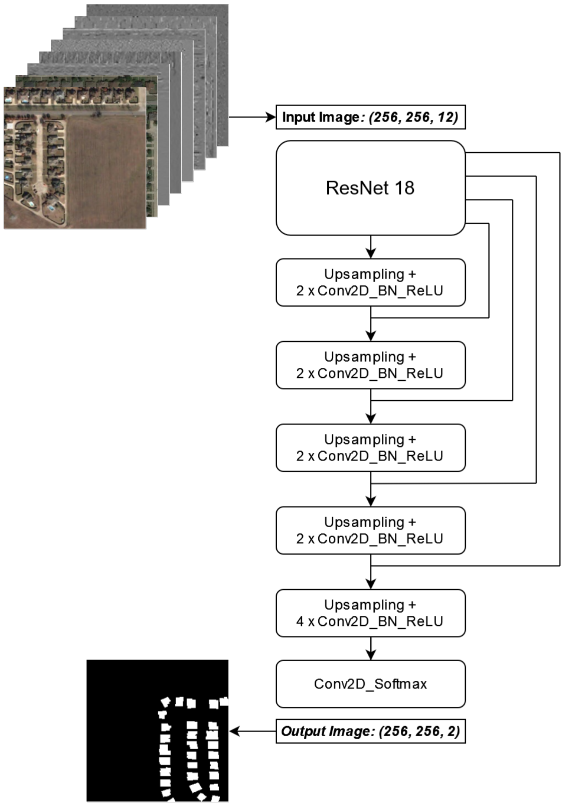


Figure 7: MRA augmentation

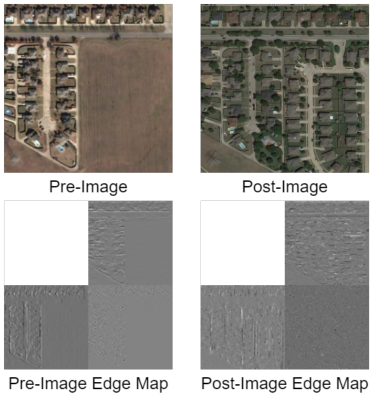


Figure 8: MRA decompositions of two input images

We use Multiresolution Analysis (MRA) to compute first-level decompositions of the input images to be fed along with them to the model. We use the 2-D Discrete Wavelet Transform (DWT) to perform this decomposition, the result of which is an approximation of the image and three directional edge maps (horizontal, vertical and diagonal), all at half the resolution of the original image. For our purposes, we shall use only the directional edge maps.

We use the "haar" wavelet to perform the 2-D DWT.

5.2.3 Feature Augmentation

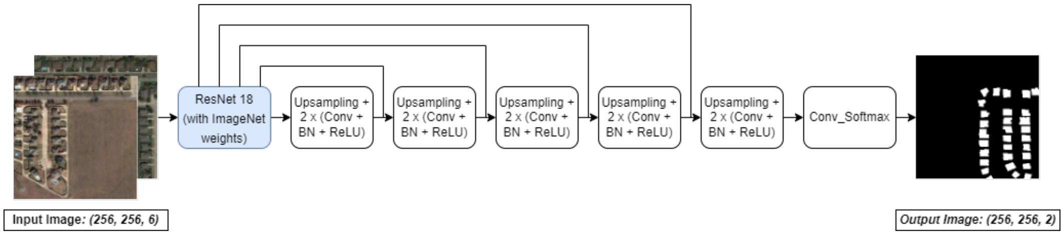

ResNet 18 (with ImageNet weights)
Upsampling + 2 x (Conv + BN + ReLU)
Upsampling + 2 x (Conv + BN + ReLU)
Upsampling + 2 x (Conv + BN + ReLU)
Upsampling + 2 x (Conv + BN + ReLU)
Upsampling + 2 x (Conv + BN + ReLU)
Conv_Softmax
Input Image: (256, 256, 6)
Output Image: (256, 256, 2)

Figure 9: Feature Augmentation

In this augmentation technique, instead of initializing the ResNet 18 encoder with random weights, we load it with ImageNet weights. The idea behind this approach is to implement transfer learning, where we pre-load the network with generic feature-extracting filters learnt by training on a particularly large corpus of data and use it as a starting point to fine-tune those filters for our specific task.

Since the ImageNet weights are taken from a model trained on 3-channel images, we modify the weights of the first channel so as to facilitate loading them into our 6-channel input model. For the first layer, we stack two copies of the convolutional filters and the batchnorm values to match the shape of the input layer. The rest of the layers are loaded unchanged.

5.3 Phase III

In this phase, we evaluate the three different augmented architectures (edge maps, MRA decompositions, and features) and the unaugmented architecture on three specifically designed challenges that further test the performance of these models. The intuition behind this phase is to specifically find which of the augmentations offers improved robustness in certain difficult situations as opposed to an unaugmented model.

5.3.1 Dataset Shift

In this challenge, we explore the dataset shift problem, wherein a model trained on a particular dataset is tested on a different dataset. Essentially this gauges the overall ability of the models to generalize even beyond the distribution from which the training set is sourced. Usually, deep learning models fall short of the expected generalizability owing to the overparameterized nature of such architectures. The aim is to see if the augmentations proposed in this work offer better generalizability across more than just the training dataset's distribution. This challenge is especially relevant in the field remote sensing, since systems are trained on datasets that capture only a few locations and conditions, but their application would be

To test the models through this particular challenge, we use an alternate dataset and create a test set to evaluate the already trained models on the original dataset. Here the original dataset is LEVIR-CD+, while the alternate dataset used is the WHU-CD dataset.

It is worth noting that the spatial resolutions of the two datasets are different, which would create a good barrier to generalization. Additionally, the site of data capture and the sensor used differ across the two datasets. This variation in parameters should be sufficient to consider the two datasets as being sourced from different distributions.

5.3.2 Few-Shot Learning

In this challenge, we explore the challenge of few-shot learning. Here, a model is trained on a smaller subset of the original dataset to see its capability to learn the target task with lesser data. Deep learning models often struggle with the problem of overfitting, and this only gets worse with smaller datasets. Therefore a better few-shot learning capability of models is a valuable quality.

For the specific application of remote sensing-based change detection, this is relevant because all datasets are not created by equal sampling rates. Therefore this would result in variation in datasets in terms of their size. A deep learning model would be required to learn distinguishing features for the particular with a minimal amount of data without significant degradation in performance.

For designing an experiment to test our models through this challenge, we create smaller subsets of the original dataset by random sub-sampling - taking 50%, and 25% of the original size, and train all variants of models on these to get their respective few-shot test performances.

5.3.3 Noise Robustness

Remotely sensed images are often impacted by different kinds of noises, which might be because of the particular sensors used for data capture, or just the conditions of capture in general. It might not always be practical to employ complex denoising techniques, especially when one is not even sure about what noise has contaminated the signal. In this scenario, deep learning models which are robust against the presence of noise are preferred.

With this motivation, we test the robustness of augmented models against the effect of noise in images compared to the unaugmented model. We do so by injecting two different kinds of noise on only the test set of the original dataset - LEVIR-CD+, while the training set remains clean. These noises are chosen by identifying some of the common noise occurrences in remotely sensed data, and are finalized to be Gaussian Noise, and Salt & Pepper Noise. All models are thus trained on images without noise, and they are tested on images that have presence of noise in them.

For Gaussian Noise, we choose the mean as 0, standard deviation of 0.5, and amplitude of 0.2, while for Salt & Pepper Noise the threshold used is 0.025 for both noise segments.

6 Results

6.1 Phase I

Table 1: Baseline models comparison

Model	IoU
Early Fusion	55.02 %
Siamese Conc.	53.61 %
Siamese Diff.	50.80 %

From Table 1, we can see that the Early Fusion model has the highest IoU of the three networks. Therefore, we shall continue to use this model as our base for further experiments.

6.2 Phase II

Table 2: Augmented models comparison

Model	IoU
Unaugmented	55.02 %
Edge-Augmented	57.14 %
MRA-Augmented	56.04 %
Feature-Augmented	55.35 %

From Table 2, we can see a clear increase in performance using augmented models as compared to the unaugmented model. The edge-augmented model has the highest IoU, followed by the MRA-augmented model.

6.3 Phase III

Table 3: Dataset Shift - Train: LEVIR-CD+, Test: WHU-CD

Model	IoU
Unaugmented	35.30 %
Edge-Augmented	38.62 %
MRA-Augmented	35.67 %
Feature-Augmented	36.61 %

Table 4: Few-Shot Learning

Model	IoU	
	25% Train	50% Train
Unaugmented	47.77 %	51.07 %
Edge-Augmented	48.75 %	54.88 %
MRA-Augmented	48.6 %	53.98 %
Feature-Augmented	35.63 %	35.06 %

Table 5: Noise Robustness

Model	IoU	
	Gaussian Noise	Salt and Pepper Noise
Unaugmented	43.74 %	46.22 %
Edge-Augmented	45.49 %	47.06 %
MRA-Augmented	50.06 %	51.39 %
Feature-Augmented	46.08 %	44.86 %

From Tables 3, 4, 5, we can see that the augmented models mostly outperform the unaugmented model. Particularly, the edge-augmented and the MRA-augmented models have better performance than the unaugmented model.

7 Conclusion

The results obtained by the three-phased experiment support the hypothesis that augmented convolution neural networks outperform unaugmented networks for the task of change detection. This improvement in performance is seen not just while training and testing on one dataset, but also on specially engineered challenges of dataset shift, few shot learning and noise injection.

In general, the Edge-Augmented variant is the best-performing model. The intuition behind this trend could be that edge detection is essential to the feature extraction conducted by convolutional layers in the networks. Thus, feeding edge maps along with the input helps with more efficient computation by the deep learning architecture. For noise robustification, a multiresolution decomposition of the input helps the most, the reason behind which can be understood by how MRA works - an image is decomposed into a lower resolution and into its constituent approximation and directional details. As the noise is added at a higher resolution, an appropriate wavelet used for decomposing the image can effectively segment out noises in different directions and thus help the deep learning model identify the areas affected by noise. The feature augmented network is not as effective as the edge and MRA augmented models, but still generally outperforms the unaugmented network. The less effectiveness of this augmentation can be attributed to the fact that we are using Imagenet weights for pertaining, which may not be extremely useful for networks dealing with remotely sensed data.

It is thus clear with these sets of experiments that deep learning architectures can be augmented with image processing ideas such as edge detectors and MRA decompositions, and this offers a relatively simple way of improving their performance without introducing more number of parameters of complicated operations.

8 Code

The code associated with this work can be found at `https://github.com/RefineX/Change-Detection`

9 Contributions

The contributions of each member are as follows:

- Sarvesh Patil:
  - Implementing the early fusion, siamese and augmented model architectures
  - Evaluating models for noise robustness
- Pranjal Ranjan:
  - Curating and pre-processing the primary and alternate datasets
  - Evaluating models for dataset shift
- Badhrinarayan Malolan:

  - Implementing and experimenting with edge augmentation
  - Implementing and experimenting with MRA augmentation
- Ankit Parekh:
  - Implementing and experimenting with few-shot learning
  - Experimenting with feature augmentation

References

[1] Masroor Hussain et al. "Change detection from remotely sensed images: From pixel-based to object-based approaches". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 80 (2013), pp. 91–106. DOI: 10.1016/j.isprsjprs.2013.03.006.

[2] Jia Liu et al. "An end-to-end supervised domain adaptation framework for cross-domain change detection". In: *Pattern Recognition* 132 (2022), p. 108960. DOI: 10.1016/j.patcog.2022.108960.

[3] Daifeng Peng, Yongjun Zhang, and Haiyan Guan. "End-to-end change detection for high resolution satellite images using improved UNET++". In: *Remote Sensing* 11.11 (2019), p. 1382. DOI: 10.3390/rs11111382.

[4] Hao Chen and Zhenwei Shi. "A spatial-temporal attention-based method and a new dataset for Remote Sensing Image Change Detection". In: *Remote Sensing* 12.10 (2020), p. 1662. DOI: 10.3390/rs12101662.

[5] Rizwan Ahmed Ansari, Krishna Mohan Buddhiraju, and Rakesh Malhotra. "Urban change detection analysis utilizing multiresolution texture features from polarimetric SAR images". In: *Remote Sensing Applications: Society and Environment* 20 (2020), p. 100418. DOI: 10.1016/j.rsase.2020.100418.

[6] Sheng Fang et al. "SNUNet-CD: A densely connected siamese network for change detection of VHR Images". In: *IEEE Geoscience and Remote Sensing Letters* 19 (2022), pp. 1–5. DOI: 10.1109/lgrs.2021.3056416.

[7] Xueli Peng et al. "Optical remote sensing image change detection based on attention mechanism and image difference". In: *IEEE Transactions on Geoscience and Remote Sensing* 59.9 (2021), pp. 7296–7307. DOI: 10.1109/tgrs.2020.3033009.

[8] Hongjia Chen et al. "RDP-net: Region detail preserving network for change detection". In: *IEEE Transactions on Geoscience and Remote Sensing* (2022), pp. 1–1. DOI: 10.1109/tgrs.2022.3227098.

[9] Jie Chen et al. "DASNet: Dual attentive fully convolutional siamese networks for change detection in high-resolution satellite images". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021), pp. 1194–1206. DOI: 10.1109/jstars.2020.3037893.

[10] Chenxiao Zhang et al. "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 166 (2020), pp. 183–200. DOI: 10.1016/j.isprsjprs.2020.06.003.

[11] Zhuo Zheng et al. "Change is everywhere: Single-temporal supervised object change detection in remote sensing imagery". In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (2021). DOI: 10.1109/iccv48922.2021.01491.

[12] Hao Chen and Zhenwei Shi. "A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection". In: *Remote Sensing* 12.10 (2020). ISSN: 2072-4292. DOI: 10.3390/rs12101662. URL: https://www.mdpi.com/2072-4292/12/10/1662.

[13] Shunping Ji, Shiqing Wei, and Meng Lu. "Fully convolutional networks for Multisource building extraction from an open aerial and satellite imagery Data set". In: *IEEE Transactions on Geoscience and Remote Sensing* 57.1 (2019), pp. 574–586. DOI: 10.1109/tgrs.2018.2858817.

[14] Rodrigo Caye Daudt, Bertr Le Saux, and Alexandre Boulch. "Fully convolutional siamese networks for change detection". In: *2018 25th IEEE International Conference on Image Processing (ICIP)* (2018). DOI: 10.1109/icip.2018.8451652.