

EVALUATION OF SEVERAL FULLY CONVOLUTIONAL NETWORKS IN SAR IMAGE CHANGE DETECTION

Linxia Ji¹, Zheng Zhao^{1*}, Wenhao Huo^{1,2}, Jinqi Zhao³, Rui Gao^{1,2}

¹ Chinese Academy of Surveying & Mapping, Beijing, 100830, China (zhaozheng@casm.ac.cn)

² Shandong University of Science and Technology, Qingdao, 266590, China

³ China University of Mining and Technology, Xuzhou, 221116, China

KEY WORDS: Change Detection, SAR, Siamese network, Encoder-Decoder, Transfer learning, Generalization.

ABSTRACT:

In recent years, the world is suffering from frequent natural disasters. Change detection (CD) technology can quickly identify the change information on the ground and has developed into an important means of disaster monitoring and assessment. Synthetic aperture radar (SAR) has the characteristics of periodic observation and wide coverage. Moreover, SAR has the advantages of penetrating, all-day and all-weather observation, which plays an important role in disaster monitoring. Due to the rapid development of satellite sensors, the available CD data has been greatly enriched. This situation provides an opportunity for deep learning change detection (DLCD) techniques. However, SAR data are affected by speckle noise and lack of available labeled samples, it remains challenging to precisely locate the change information with high efficiency. This paper focuses on several commonly used and outstanding networks in the DLCD field to evaluate their performance and develop them to SAR data. In addition, Transfer learning experiments are designed to evaluate the generalization performance of each network for the CD task. The experimental results show that the Siamese CD network encoding multi-temporal data separately has the best ability to detect changes and generalization performance. In addition, adding high quality explicit difference guidance information to the network is more specific for the CD task, which can further improve network performance and refine the boundaries of changed ground objects on change map.

1. INTRODUCTION

In recent years, earthquakes, floods, landslides, forest fires, and other natural disasters have occurred frequently around the world, which has brought terrible suffering to humanity and nature. By using multi-temporal remote sensing data to detect change in affected areas, decision makers can quickly obtain disaster information. Synthetic aperture radar (SAR), as active remote sensing technology, has the advantages of penetrability, all-weather all-day observation ability, and wide coverage. As a result, SAR can obtain ground information even in the harsh environment with disasters, which provides technical support for rapid emergency response and disaster rescue. Therefore, SAR image change detection technology is widely used in land use, disaster monitoring, ecological protection, and agricultural investigation (Erten et al., 2016; Manavalan, 2017).

The early research on change detection using SAR images is mainly based on manually designed features (Bovolo & Bruzzone, 2005, 2007). These methods are simple to implement, but require high image quality and lack robustness to speckle noise. The deep learning (DL) technique has achieved great success in many fields due to its powerful feature extraction ability. Some researchers have made many attempts to introduce DL into remote sensing image change detection (RSCD) and created many surprising outcomes.

According to different strategies of multi-temporal information fusion, the change detection based on the deep learning (DLCD) method can be divided into two types: (1) Early Fusion (EF): multi-temporal images are stacked together as different channels into the network to fuse temporal information. Simultaneously, other manually designed features can be added to increase the information for different tasks. For the SAR CD task, many

studies (Gong et al., 2016; R. Wang et al., 2021) first generated the difference image (DI) as a channel of input to increase explicit difference information. The deep belief network (DBN) and autoencoder require the input to be converted into a vector, and researchers usually connect the neighboring data of different times into a vector as the network input. These methods fuse multi-temporal information before the data is fed into the network. Therefore, similar methods are classified as Early Fusion, this term was first mentioned in Daudt et al., (2018), and this paper expands its meaning. (2) Siamese network: multi-temporal images are fed into different branches of the Siamese network to extract their high-level feature representations and then change detection is performed. Some existing studies use the Siamese network to separately extract features from multi-temporal images and detect change information according to the similarity measure of feature vectors (Chen et al., 2021; Zhan et al., 2017). Other researchers fed multi-temporal high-level features extracted from the Siamese network into a change decision network to output the final change map (M. Wang et al., 2020). In addition, some researchers use the Siamese network as an encoder to extract features from multi-temporal data, then fuse these features in the decoding stage (Caye Daudt et al., 2018; Fang et al., 2022). These two designs can be based on patches or images, and both of them produce many state-of-art CD algorithms. However, these methods based on patch sampling cannot make full use of the spatial information, and overlap between patches leads to high computational cost, which limits the efficiency improvement of the CD algorithm. While a series of image-based fully convolutional semantic segmentation networks, CD algorithms based on FCN (Song et al., 2018), Unet (Fang et al., 2022), DeepLab (Y. Wang et al., 2021) and their variants can overcome these shortcomings. These methods can accept input of any size and directly generate dense predictions corresponding to each input pixel, which is efficient and accurate,

* Corresponding author

and has become the most popular network of DLCD.

In this paper, SAR images are used to detect the change caused by natural disasters or environmental factors. The change detection based on Fully Convolutional Neural Network (FCNN) architectures are introduced and used in the multi-temporal SAR images. Firstly, we compare the Encoder-Decoder architecture represented by the Unet with the Siamese structure design and analyse the impact of these two structural designs on detection accuracy. Secondly, the DI image is added to the input of Encoder-Decoder architecture to evaluate the effect on the accuracy of the CD. In addition, the transfer learning experiment is designed to evaluate the ability of networks to detect change information. Finally, considering the scarcity of SAR datasets, we verify the generalization ability of the pre-trained model when applied to heterogeneous data.

The rest of this paper is organized as follows: the principle and detailed procedures of the proposed method are described in Section 2. Section 3 provides the experimental results and analysis. Finally, our conclusions are drawn in Section 4.

2. METHOD

Based on the work of Caye Daudt et al. (2018), we design an unsupervised CD flow that automatically selects reliable training samples from SAR data to evaluate the performance of several different network structures for SAR image change detection. Furthermore, we are interested in the network's ability to detect change information. Therefore, we design transfer learning experiments to evaluate the generalization performance of these networks.

2.1 Sample Selection and Augmentation

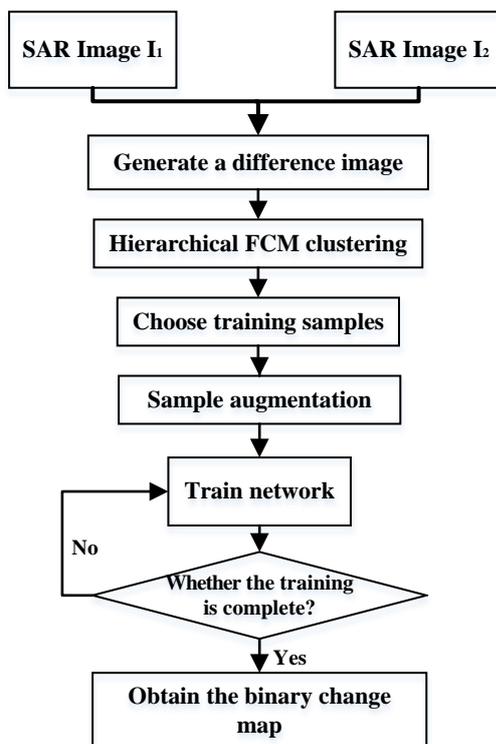


Figure 1. The flowchart of SAR change detection based on deep learning

The CD method in this paper is designed to work in an unsupervised manner via generating training samples with a pre-task (Qu et al., 2022). Given two SAR images acquired at different times in the same geographical area, the difference image (DI) is generated by using the log-ratio operator (Bovolo & Bruzzone, 2005). Hierarchical FCM (HFCM, Bazi et al., 2006) clustering is used to classify DI into three categories: changed class, unchanged class, and uncertain class. Pixels classified as changed and unchanged class can be considered as reliable samples with a high probability of having changed or no change.

Although the training samples have been obtained, only a few pixels are selected as samples, and others need to be further classified, so the training samples are still scarce. Therefore, we need to use data augmentation techniques that can augment samples to prevent overfitting and improve network generalization ability and performance (Krizhevsky et al., 2012; Zhou et al., 2018). we applied random crop with a 50% probability to expand training data. Then we use these samples to train the network. This CD flow is unsupervised even though the training of the network is supervised. The flowchart of this CD flow is shown in Figure. 1

2.2 Networks

Caye Daudt et al. (2018) proposed three Fully Convolutional Neural Network (FCNN) architectures to perform change detection on multi-temporal images of earth observation. We extend the other two structures based on their work and a total of five different networks for our experiments.

2.2.1 FC-EF: The first architecture, named Fully Convolutional Early Fusion (FC-EF), The term “Early Fusion (EF)” was proposed by Daudt et al. (2018). The EF network in the original paper is patch-based. Caye Daudt et al. (2018) extended it to an image-based full convolutional network and obtained the FC-EF (Figure 2 (a)). This network is directly based on the UNet (Ronneberger et al., 2015) model, but FC-EF only contains four layers of encoder-decoder, which is shallower than the U-net. The input of the network is stacking multi-temporal images.

2.2.2 FC-EF-DI: Based on FC-EF, we replaced the input with concatenating multi-temporal images and DI generated by the log-ratios operator, in other words, incorporating difference information at the beginning of the network. This design has been used in many studies, and we intend to examine whether this design could improve detection accuracy. This structure is named FC-EF-DI, it is the same as FC-EF except for the number of input channels (Figure 2 (a)).

2.2.3 FC-Siam-conc: The third architecture is a Siamese Network, also inspired by the work of Daudt et al. (2018). The basic idea is to process multi-temporal images in parallel through two branches with shared weights and fuse them at the output. Caye Daudt et al. (2018) extended it to the fully convolutional network with Decoder-Encoder architecture (Figure 2.(b)). The encoding module of the network is divided into two branches with shared weights for processing multi-temporal pairs of images respectively, and the features from the two branches were fused in the decoding module. Each decoding layer uses two skip connections, which concatenate the output of the decoding layer with features of the corresponding scale from two encoding streams. This network structure is named Fully Convolutional Siamese-concatenation (FC-Siam-conc).

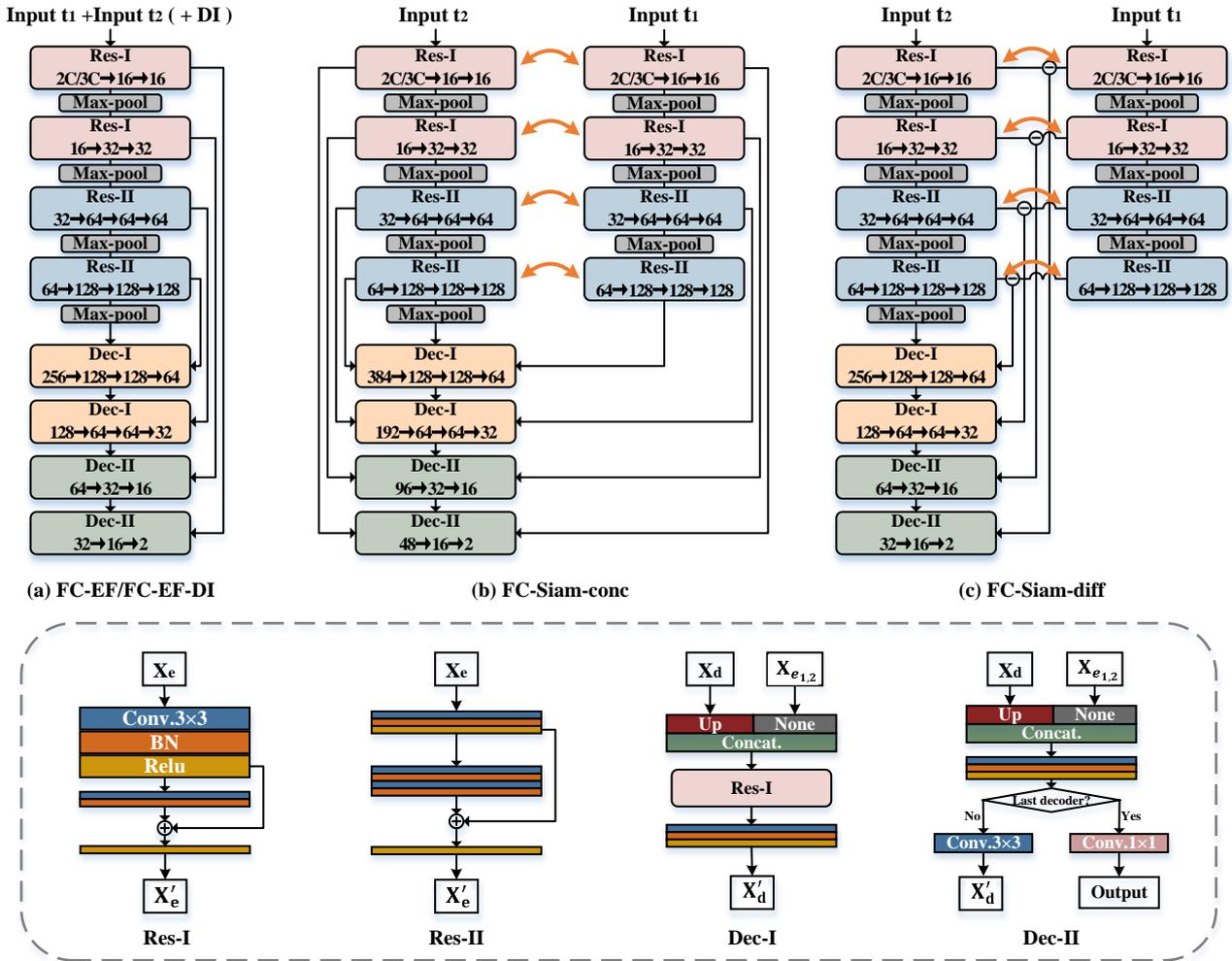


Figure 2. Our implementation of three different change detection frameworks proposed by Caye Daudt et al.(2018) . Two types of residual blocks and encoder modules are used. X_e represents the input features from the last layer and X'_e represents the output features at the encoding stage. X_d, X'_d has a similar denotation to X_e, X'_e but belong to the decoding stage. $X_{e_{1,2}}$ denotes the combination of encoder multi-temporal outputs. Orange arrows illustrate weight sharing.

2.2.4 FC-Siam-diff: This network differs from FC-Siam-conc only in how skip connections are used. Each decoding block using one skip connection obtains the absolute value of feature difference from the corresponding scale of the encoding stream (Figure 2.(c)). As detecting differences in the multi-temporal data is critical in the CD task, this design is intuitive.

2.2.5 FC-Siam-conc&diff: This structure is a combination of FC-Siam-diff and FC-Siam-conc, named FC-Siam-conc&diff. The original output of two encoding streams and the absolute value of their difference are concatenated simultaneously via skip connections. In this way, the network can maintain the original information of the high-level features from encoding streams, and explicit guidance of the difference information can be obtained. We are interested in whether a combination of the two structures could improve the network performance. Inevitably, the number of training parameters increases.

The above five networks have the same backbone. Caye Daudt et al. (2018) failed to provide further details about the implementation of the network. Our implementation refers to the work of Zhang et al., (2021), the details are shown in Figure 2. Each encoding block and decoding block uses the residual connection to prevent gradient disappearance to train a deeper

network, and the skip connection between the encoder and decoder is used to add more local information to high-level features and fuse multi-temporal information. Readers can refer to Figure 2 for more details.

2.3 Loss Function

As the distribution of changed and unchanged classes is often highly unbalanced in the CD task, the loss function will be biased to the class with large samples during training, resulting in low recognition accuracy for the class with small samples. Therefore, it is necessary to use the weighted cross-entropy function to train the network, which is defined as follows:

$$L_{CE}(\mathbf{P}, \mathbf{G}) = \frac{1}{N} \sum_i [-\omega_c G_i \log(P_i) - \omega_u (1 - G_i) \log(1 - P_i)] \quad (1)$$

where \mathbf{P} : change predicted map
 \mathbf{G} : ground truth
 i : pixel index
 ω_c : weights for the changed classes
 ω_u : weights for the unchanged classes

Dice Loss is the appropriate choice when the number of foreground and background pixels is unbalanced as in CD tasks. Dice similarity can measure the similarity of two sets, and its value is $[0, 1]$ (Equ.2). For the change predicted map \mathbf{P} and ground truth \mathbf{G} , the Dice Loss is defined as (Equ.3):

$$Dice = \frac{2|\mathbf{P} \cap \mathbf{G}|}{|\mathbf{P}| + |\mathbf{G}|} \quad (2)$$

$$L_{Dice}(\mathbf{P}, \mathbf{G}) = 1 - Dice \quad (3)$$

The Joint Loss Function is defined as follows (Equ.4), and we use it to train our CD networks.

$$L = L_{CE}(\mathbf{P}, \mathbf{G}) + L_{Dice}(\mathbf{P}, \mathbf{G}) \quad (4)$$

2.4 Transfer Learning

Transfer Learning generally refers to using the mastered knowledge to acquire new knowledge. The core of this concept is to find the similarity between the source domain and the target domain and use this similarity as a bridge to achieve the purpose of transfer learning. DL has a strong dependence on a large amount of training data, and we expect our network can understand the underlying nature of data. In the RSCD, although a large amount of observation data has been accumulated, available labeled data is still scarce, especially for SAR data. However, if we design a network with strong change detection ability, which is the most critical for the CD task, we use this network for different sources of data with different distributions, also can better complete the change detection task, we will consider that this network possesses good generalization performance for CD task. Briefly, designing networks with strong change detection capabilities is crucial for making full use of multi-source remote sensing CD datasets to accomplish more complex CD tasks with higher accuracy.

3. EXPERIMENTS AND RESULTS

This section is organized as follows. Section 2.1 describes the datasets; Section 2.2 introduces the evaluation criteria; In Section 2.3, comparative experiments are designed to evaluate the performance of five different change detection networks on SAR data. In Section 2.4, transfer learning experiments are designed to evaluate the generalization performance of each network for heterogeneous data.

3.1 Datasets

3.1.1 Ottawa dataset: The dataset was acquired by the Radarsat satellite over Ottawa in May 1997 and August 1997. The changes were caused by summer flooding.

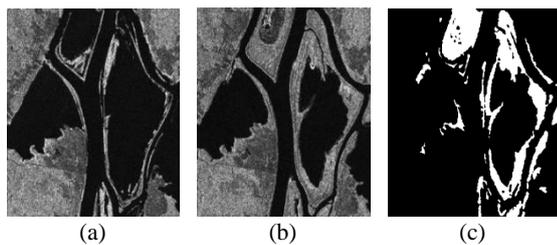


Figure 3. Ottawa dataset. (a) Image acquired in May 1997, during the summer flooding. (b) Image acquired in August 1997, after the summer flooding. (c) Ground truth.

3.1.2 Sulzberger dataset: The dataset was acquired by the Envisat satellite on March 11 and 16, 2011. Both the images show the process of sea ice breakup. When the Tohoku tsunami in the Pacific Ocean was triggered on March 11, 2011, the massive waves caused the ice shelf to flex and break (Gao et al., 2019).

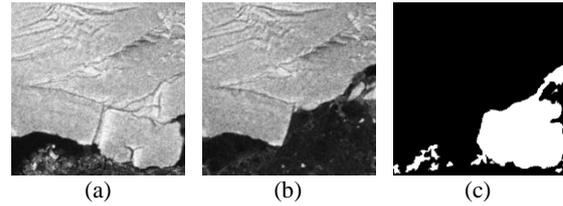


Figure 4. Sulzberger dataset. (a) Image acquired on March 11, 2011. (b) Image acquired in on March 16, 2011. (c) Ground truth.

3.1.3 Yellow River dataset: The dataset was acquired by Radarsat-2 satellite in June 2008 and June 2009 at the region of Yellow River Estuary in China. It is worth noting that the two images are single-look image and four-look image, respectively. This indicates they suffer from different levels of speckle noise. This dataset is associated with environmental change. We selected two typical changed areas, the YellowRiver -Inland water (Figure 6. second row) and YR-Farmland (Figure 6. third row).

3.1.4 Muraglia dataset: The dataset was acquired by the Landsat-5 satellite in September 1995 and July 1996. The single-band optical images show the changed lake in Muraglia, Italy (Figure 7).

Ottawa and Sulzberger Dataset were used to evaluate the performance of five different CD networks on SAR data, and Yellow River and Muraglia datasets were used to conduct transfer learning experiment.

3.2 Evaluation Criteria

The evaluation indicators used in this paper are: overall accuracy (OA), precision (Pre), recall, F1-score, and Kappa, which are computed by:

$$OA = (TP + TN)/(TF + FP + TN + FN) \quad (5)$$

$$Pre = TP/(TP + FP) \quad (6)$$

$$Recall = TP/(TP + FN) \quad (7)$$

$$F1_score = \frac{2 \times Pre \times Recall}{Pre + Recall} \quad (8)$$

$$P_e = \frac{(TP + FP) \times (TP + FN) + (FN + TN) \times (FP + TN)}{(TP + TN + FP + FN)} \quad (9)$$

$$Kappa = \frac{OA - P_e}{1 - P_e} \quad (10)$$

where TP: the number of true positives
TN: the number of true negatives
FP: the number of false positives
FN: the number of false negatives

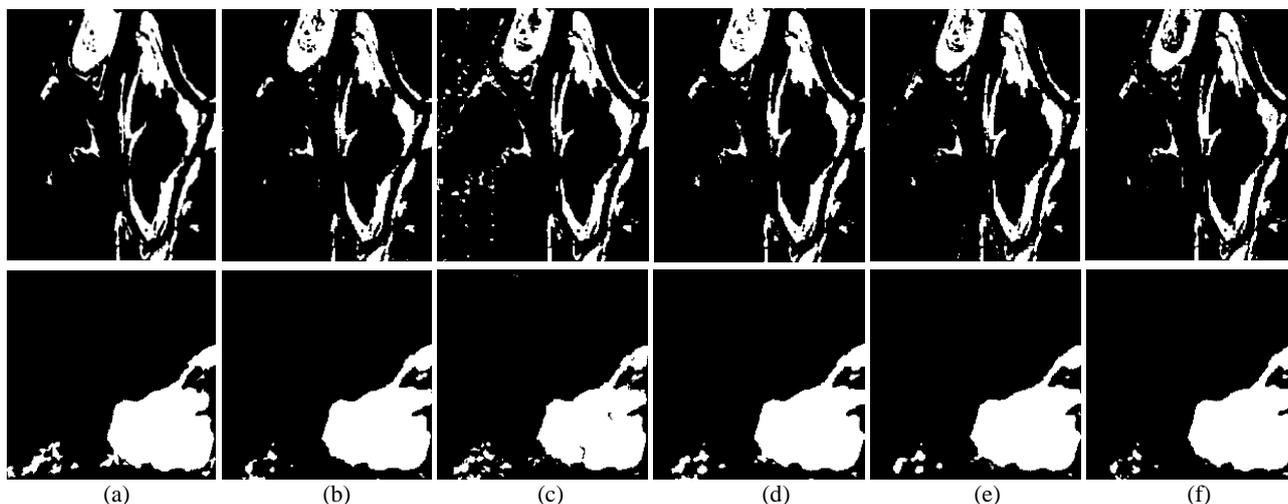


Figure 5. Visualized results for different change detection networks on the Ottawa dataset (first row) and Sulzberger dataset (second row). (a) Ground truth. (b) FC-EF. (c) FC-EF-DI. (d) FC-Siam-conc. (e) FC-Siam-diff. (f) FC-Siam-conc&diff.

Network	Results on Ottawa dataset					Results on Sulzberger dataset				
	OA	Pre	Recall	F1	Kappa	OA	Pre	Recall	F1	Kappa
FC-EF	<u>97.12</u>	96.51	84.83	90.29	88.61	97.53	95.23	91.78	93.47	91.95
FC-EF-DI	96.18	88.20	87.54	87.87	85.60	97.77	95.51	92.34	93.90	92.71
FC-Siam-conc	97.96	95.67	91.21	93.39	92.18	98.04	95.06	94.73	94.89	93.68
FC-Siam-diff	96.84	<u>96.13</u>	<u>88.22</u>	<u>92.01</u>	<u>90.58</u>	<u>97.92</u>	95.68	<u>93.39</u>	<u>94.52</u>	<u>93.34</u>
FC-Siam-conc&diff	96.20	92.78	82.40	87.28	85.06	97.64	<u>95.55</u>	92.03	93.76	92.30

Table 1. Change detection results for different networks on the Ottawa dataset (left) and Sulzberger dataset(right). Pre and F1 representative precision and F1-score respectively. The best results are marked in bold and the second-best results are underlined.

3.3 Comparative Experiments

This paper evaluated SAR image change detection networks on Ottawa and Sulzberger datasets. We implemented five networks in PyTorch, and the training was powered by an NVIDIA GeForce RTX 3060 Laptop GPU.

The hyperparameter setting is shown in Table 2. For all the datasets and networks, we use the Adaptive moment Estimation (Adam) optimizer to optimize the parameters and train 100 epochs. The initial learning rate is set to 0.001. We used the data augmentation strategy described above to augment the samples and prevent overfitting. The normalized transform is applied to the data. ω_c and ω_u in the weighted cross-entropy loss function are set as (0.6, 0.4).

Hyperparameter	Setting
ω_c, ω_u	0.6, 0.4
Initial learning rate	0.001
optimizer	Adam

Table 2. Hyperparameter setting.

Figure 5 shows the CD results using different networks. The quantitative results are listed in Table 1. For the Ottawa dataset, FC-Siam-conc and FC-Siam-diff based on Siamese network achieve the best and second-best results respectively. These results suggested that the design of Siamese structure is beneficial to improve the accuracy of change detection. The Siamese network separately extracts the high-level features of multi-temporal images, then integrates them in the high-dimensional space. We infer that these high-level features contain less noise, therefore, Siamese structure design is a better

fusion strategy for multi-temporal information. Furthermore, FC-Siam-conc outperforms FC-Siam-diff. The former has a larger capacity, and it's skip connection retains more original information about the multi-temporal high-level features. FC-EF, as a shorter version of the Unet, also perform well. However, FC-EF-DI show poor performance. We estimated that the poor quality of the DI adds a noisy difference guide. The results of Sulzberger dataset also confirm the above conclusion. The results of both datasets demonstrated that FC-Siam-conc&diff achieves bad results, so this design leading to redundant information is unnecessary.

3.4 Generalization performance evaluation

The Siamese network proposed by Caye Daudt et al.(2018) is aimed at the CD task. It can use the available multi-source CD dataset to train without the addition of other types of data. Therefore, these networks should possess a better ability to detect change information but not others. We designed transfer learning experiments to evaluate the generalization performance of FC-EF, FC-Siam-conc, FC-Siam-diff, and FC-Siam-conc&diff.

This part of the experiment still follows the unsupervised flow mentioned above. The Ottawa dataset is used to select reliable samples and train these networks to get the pre-trained models. These models are used to perform change detection on other datasets to evaluate the generalization performance of each network.

First, we applied these pre-trained models to other SAR datasets which are acquired from different sensors and have different causes of change. Furthermore, the unsupervised flow in this paper has yielded poor results when applied to these datasets.

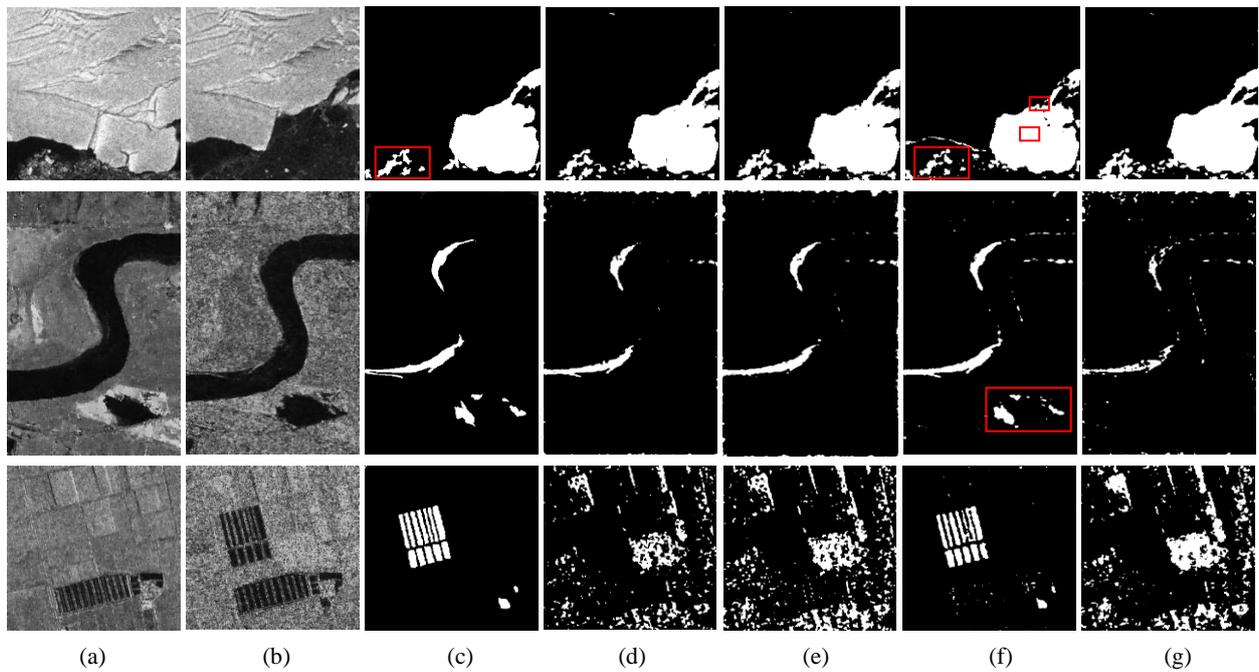


Figure 6. Visualized change detection results for different pre-trained networks on the SAR dataset, Sulzberger dataset (first row), YellowRiver-Inland water (second row), YellowRiver-Farmland (third row). (a) Image captured at t_1 . (b) Image captured at t_2 . (c) Ground truth. (d) FC-EF. (e) FC-Siam-conc. (f) FC-Siam-diff. (g) FC-Siam-conc&diff.

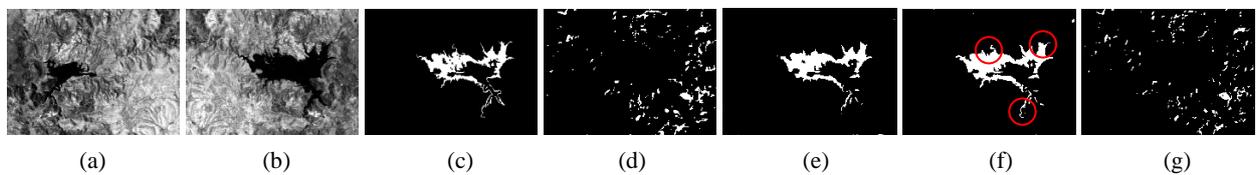


Figure 7. Visualized change detection results for different pre-trained networks on the optical Muragia dataset. (a) Image captured at t_1 . (b) Image captured at t_2 . (c) Ground truth. (d) FC_EF. (e) FC_Siam_conc. (f) FC_Siam_diff. (g) FC_Siam_conc&diff.

Network	Results on Sulzberger dataset					Results on YellowRiver-Inland water dataset				
	OA	OA	OA	OA	OA	OA	Pre	Recall	F1	Kappa
FC-EF	<u>97.77</u>	<u>97.74</u>	<u>97.74</u>	<u>97.74</u>	<u>97.74</u>	<u>97.74</u>	78.77	89.61	<u>83.84</u>	<u>83.62</u>
FC-Siam-conc	97.39	97.27	97.27	97.27	97.27	97.27	<u>84.08</u>	83.20	83.64	83.43
FC-Siam-diff	98.01	98.77	98.77	98.77	98.77	98.77	88.32	<u>84.42</u>	86.33	86.15
FC-Siam-conc&diff	96.89	97.00	97.00	97.00	97.00	97.00	0.00	0.00	NaN	1.62

Table 3. Change detection results for different pretrained networks on the SAR dataset, Sulzberger dataset (left) and YellowRiver-Inland dataset (right). Pre and F1 representative precision and F1-score respectively. The best results are marked in bold and the second-best results are underlined.

Network	Results on YellowRiver-Farmland dataset					Results on Muragia dataset				
	OA	Pre	Recall	F1	Kappa	OA	Pre	Recall	F1	Kappa
FC-EF	<u>83.32</u>	0.00	0.00	NaN	8.27	89.49	0.00	0.00	NaN	5.37
FC-Siam-conc	78.54	0.00	0.00	NaN	9.38	98.40	86.69	<u>87.49</u>	87.09	86.24
FC-Siam-diff	98.63	94.06	82.03	87.63	86.91	<u>98.25</u>	<u>81.12</u>	93.42	<u>86.84</u>	<u>85.90</u>
FC-Siam-conc&diff	78.05	<u>0.09</u>	<u>0.25</u>	<u>0.13</u>	<u>9.32</u>	90.75	0.00	0.00	NaN	4.28

Table 4. Change detection results for different pretrained networks, YellowRiver-Farmland dataset (left) and optical Muragia dataset (right). Pre and F1 representative precision and F1-score respectively. The best results are marked in bold and the second-best results are underlined.

Figure 6 shows the visualization results obtained by the four pre-trained networks, and Table 3 show their quantitative results. For the Sulzberger dataset (Figure 6.first row), the results are almost comparable to the above, and the boundary of the part marked by the red box is closer to the Ground truth than the above results. FC-Siam-diff had the best performance and the least false alarm. FC-EF and FC-Siam-conc achieved the second-best result, and FC-Siam-conc&diff got the worst result. For the Yellow River dataset, it is challenging to detect changes due to different levels of noise in image pairs. For Inland water (Figure 6.second row), only FC-Siam-diff can detect the changes in the red box. For the Farmland data (Figure 6. third Row), only FC-Siam-diff effectively detects changes, all other networks achieved messy results.

Therefore, a structure design like FC-Siam-diff adding an effective difference information guide to the network, is more capable of detecting changes. Even if its pretrained model is used for data with different sensors, different resolutions, and different change causes, it can achieve better results, so we consider that this kind of network has a stronger generalization ability.

We further applied the pre-trained model based on the Ottawa dataset to optical CD datasets, which is more challenging as their imaging mechanism is completely different. Figure 7 shows the visualized results, and Table 6 shows their quantitative comparison. For the optical Muraglia Dataset, FC-Siam-conc and FC-Siam-diff win by a large margin. FC-Siam-diff is best in preserving change details although the accuracy metrics are slightly worse than FC-Siam-conc. The other two structures fail to detect any useful information.

Therefore, the structure design like FC-Siam-conc and FC-Siam-diff that uses different encoding streams to process the multi-temporal data separately and then fuses them in an abstract high-dimensional space, is beneficial to improve the ability of the network to detect changes. Moreover, this ability can be maintained even for transfer learning between heterogeneous data.

4. CONCLUSIONS

In this paper, we evaluate the performance of several network structures commonly used in the field of change detection in SAR images. In addition, we designed transfer learning experiments to evaluate the generalization performance of each network on the CD task. The experimental results show that the network design similar with FC-EF (the most basic and widely used one) is simple and efficient for CD. However, FC-EF-DI based on FC-EF is generated by adding DI to the input. Although more information is added to the network, the poor quality of DI also brings noisy difference guidance information to the network, which often leads to worse CD results. The network of FC-Siam-conc and FC-Siam-diff based on the Siamese structure is better at detecting change information and generalization performance, and FC-Siam-diff is best at preserving the boundaries of the changed areas. However, FC-Siam-conc&diff, the combination of them with larger model capacity, has the worst performance unexpectedly for the data used in this paper. Therefore, if readers consider using it for other data, it is necessary to implement several experiments to test the effect. In conclusion, this study demonstrates the ability and generalization performance of various CD algorithms in SAR image change detection. We hope that these conclusions will aid researchers to design more effective CD algorithms in the future.

ACKNOWLEDGEMENTS

Thanks to my teachers, my family and friends. Thanks to the researchers who were willing to share their data and code.

REFERENCES

- Bazi, Y., Bruzzone, L., & Melgani, F. (2006). Automatic identification of the number and values of decision thresholds in the log-ratio image for change detection in SAR images. *IEEE Geoscience and Remote Sensing Letters*, 3(3), 349–353. <https://doi.org/10.1109/LGRS.2006.869973>
- Bovolo, F., & Bruzzone, L. (2005). A detail-preserving scale-driven approach to change detection in multitemporal SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 43(12), 2963–2972. <https://doi.org/10.1109/TGRS.2005.857987>
- Bovolo, F., & Bruzzone, L. (2007). A Theoretical Framework for Unsupervised Change Detection Based on Change Vector Analysis in the Polar Domain. *IEEE Transactions on Geoscience and Remote Sensing*, 45(1), 218–236. <https://doi.org/10.1109/TGRS.2006.885408>
- Caye Daudt, R., Le Saux, B., & Boulch, A. (2018). Fully Convolutional Siamese Networks for Change Detection. 2018 25th IEEE International Conference on Image Processing (ICIP), 4063–4067. <https://doi.org/10.1109/ICIP.2018.8451652>
- Chen, J., Yuan, Z., Peng, J., Chen, L., Huang, H., Zhu, J., Liu, Y., & Li, H. (2021). DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 1194–1206. <https://doi.org/10.1109/JSTARS.2020.3037893>
- Daudt, R. C., Le Saux, B., Boulch, A., & Gousseau, Y. (2018). Urban Change Detection for Multispectral Earth Observation Using Convolutional Neural Networks. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2115–2118. <https://doi.org/10.1109/IGARSS.2018.8518015>
- Erten, E., Lopez-Sanchez, J. M., Yuzugullu, O., & Hajnsek, I. (2016). Retrieval of agricultural crop height from space: A comparison of SAR techniques. *Remote Sensing of Environment*, 187, 130–144. <https://doi.org/10.1016/j.rse.2016.10.007>
- Fang, S., Li, K., Shao, J., & Li, Z. (2022). SNUNet-CD: A Densely Connected Siamese Network for Change Detection of VHR Images. *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5. <https://doi.org/10.1109/LGRS.2021.3056416>
- Gong, M., Zhao, J., Liu, J., Miao, Q., & Jiao, L. (2016). Change Detection in Synthetic Aperture Radar Images Based on Deep Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 27(1), 125–138. <https://doi.org/10.1109/TNNLS.2015.2435783>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (Vol. 25). Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>

- Manavalan, R. (2017). SAR image analysis techniques for flood area mapping—Literature survey. *Earth Science Informatics*, 10(1), 1–14. <https://doi.org/10.1007/s12145-016-0274-2>
- Qu, X., Gao, F., Dong, J., Du, Q., & Li, H.-C. (2022). Change Detection in Synthetic Aperture Radar Images Using a Dual-Domain Network. *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5. <https://doi.org/10.1109/LGRS.2021.3073900>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28
- Song, A., Choi, J., Han, Y., & Kim, Y. (2018). Change Detection in Hyperspectral Images Using Recurrent 3D Fully Convolutional Networks. *Remote Sensing*, 10(11), 1827. <https://doi.org/10.3390/rs10111827>
- Wang, M., Tan, K., Jia, X., Wang, X., & Chen, Y. (2020). A Deep Siamese Network with Hybrid Convolutional Feature Extraction Module for Change Detection Based on Multi-sensor Remote Sensing Images. *Remote Sensing*, 12(2), 205. <https://doi.org/10.3390/rs12020205>
- Wang, R., Wang, L., Dong, P., Jiao, L., & Chen, J.-W. (2021). Graph-Level Neural Network for SAR Image Change Detection. 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 3785–3788. <https://doi.org/10.1109/IGARSS47720.2021.9555003>
- Wang, Y., Gao, L., Hong, D., Sha, J., Liu, L., Zhang, B., Rong, X., & Zhang, Y. (2021). Mask DeepLab: End-to-end image segmentation for change detection in high-resolution remote sensing images. *International Journal of Applied Earth Observation and Geoinformation*, 104, 102582. <https://doi.org/10.1016/j.jag.2021.102582>
- Zhan, Y., Fu, K., Yan, M., Sun, X., Wang, H., & Qiu, X. (2017). Change Detection Based on Deep Siamese Convolutional Network for Optical Aerial Images. *IEEE Geoscience and Remote Sensing Letters*, 14(10), 1845–1849. <https://doi.org/10.1109/LGRS.2017.2738149>
- Zhang, H., Lin, M., Yang, G., & Zhang, L. (2021). ESCNet: An End-to-End Superpixel-Enhanced Change Detection Network for Very-High-Resolution Remote Sensing Images. *IEEE Transactions on Neural Networks and Learning Systems*, 1–15. <https://doi.org/10.1109/TNNLS.2021.3089332>
- Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J. M. R. S. Tavares, A. Bradley, J. P. Papa, V. Belagiannis, J. C. Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Greenspan, & A. Madabhushi (Eds.), *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (pp. 3–11). Springer International Publishing. https://doi.org/10.1007/978-3-030-00889-5_1