FUSING MULTI-MODAL DATA FOR SUPERVISED CHANGE DETECTION

Patrick Ebel[1,], Sudipan Saha[1], Xiao Xiang Zhu[1,2*]

[1] Data Science in Earth Observation (SiPEO), Technical University of Munich (TUM), Munich, Germany - (patrick.ebel, sudipan.saha)@tum.de
[2] Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Wessling, Germany - xiaoxiang.zhu@dlr.de

Commission III, WG 6

KEY WORDS: change detection, multi-modal, fusion, synthetic aperture radar (SAR), optical, deep learning.

ABSTRACT:

With the rapid development of remote sensing technology in the last decade, different modalities of remote sensing data recorded via a variety of sensors are now easily accessible. Different sensors often provide complementary information and thus a more detailed and accurate Earth observation is possible by integrating their joint information. While change detection methods have been traditionally proposed for homogeneous data, combining multi-sensor multi-temporal data with different characteristics and resolution may provide a more robust interpretation of spatio-temporal evolution. However, integration of multi-temporal information from disparate sensory sources is challenging. Moreover, research in this direction is often hindered by a lack of available multi-modal data sets. To resolve these current shortcomings we curate a novel data set for multi-modal change detection. We further propose a novel Siamese architecture for fusion of SAR and optical observations for multi-modal change detection, which underlines the value of our newly gathered data. An experimental validation on the aforementioned data set demonstrates the potentials of the proposed model, which outperforms common mono-modal methods compared against.

1. INTRODUCTION

In a time of rapidly evolving urban landscapes and nature modified due to climate change our planet faces a rapid transformation of its surface area. This transformation is being continuously monitored by modern satellite systems like Copernicus' Sentinel mission. The automated recognition of changes observed by these repeated observations is the task of change detection (CD). CD is a prominent and long-standing challenge in remote sensing (Malila, 1980); on one hand because of the Earth's dynamic nature and the need to quantify change, on the other due to the variety of land cover and the persistent challenge of the task. Recent progress in deep learning greatly benefited previous application to change detection in satellite data (Ball et al., 2017) (Zhu et al., 2017), which is the approach followed as well in our work. However, most preceding publications do not consider the fusion of multiple sensors and thereby misses on the opportunity to utilize the variety of Earth observation data available. Our work specifically addresses the challenge of multi-modal bi-temporal change detection, where (multi-spectral) optical as well as ground range detected synthetic aperture radar (SAR) measurements are available at both considered time points. Fusing these two modalities poses a difficult problem, as both domains are very different from one another: First, in terms of viewpoint geometry—while our optical data is orthorectified, the SAR measurements are sideway-looking. Second, multi-spectral optical data provides a view on the surface characteristics of the target, whereas SAR observations provide information on its physical properties. Finally, SAR data is challenging to work with and contains speckle effects that a change detector must learn to interpret as noise and not raise any false alarms about. On the other hand, SAR as an active sensor does not suffer from drawbacks of optical imagery, such as sensitivity to light conditions or bad weather due as in e.g. the presence of clouds. Exemplary full-scene observations for one ROI are portrayed in Fig. 1.

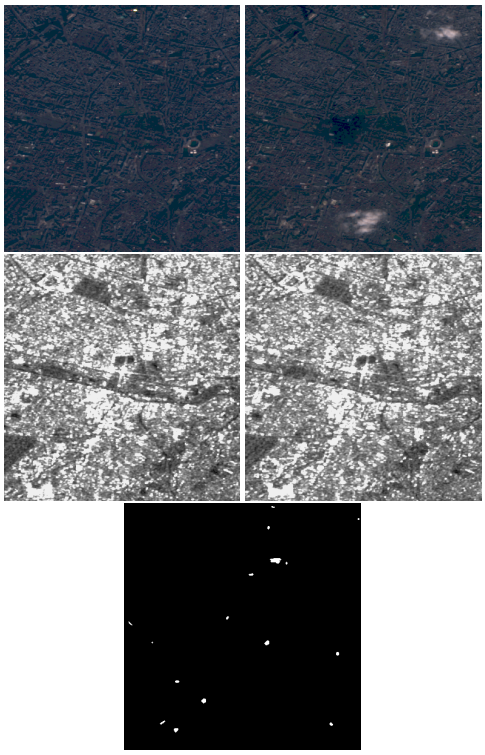

Figure 1. Multi-modal observations and change map for exemplary ROI 'Paris'. Rows: Sentinel-2 data (RGB channels). Sentinel-1 data (VV-polarized). Change maps. Columns: Time point 1. Time point 2. The images highlight the differences between both modalities and the potential complementary information they may provide to benefit change detection.

* Corresponding author

The central research question addressed in this work is whether and to which extent information from multiple sources benefits the detection of changes in remote sensing data. For this sake we build on preceding work in change detection (Daudt et al., 2018b)(Saha et al., 2019). We design a novel convolutional encoder-decoder architecture that fuses the multi-modal information and processes them in a supervised Siamese fashion. Furthermore, we collect a data set for multi-modal change detection and propose an experimental design to investigate the research question.

In sum, the contributions of our work are given by: First, we design a novel architecture that ingests both optical and SAR data and processes the multi-modal information through a Siamese network. Second, we collected SAR observations to complement an established data set of optical images, providing a multi-modal change detection data set. Third, we experimentally evaluate the proposed architecture, train it on the curated data set and test it to highlight the benefits of multi-modal data for change detection in remote sensing.

The remainder of the paper is organized as follows: The context of related work is provided in section 1.1. The methodology and the novel network architecture are introduced in section 2. The experimental design and results are reported in section 3. Finally, sections 4 and 5 close our work with a discussion and conclusion, respectively.

1.1 Related work

Change detection is a longstanding challenge in remote sensing and original methods date back accordingly into the past, constituting a long and rich body of literature on the problem. A classical approach, post-classification comparison, first semantically segments pre- and post-change images individually and then computes the change map as the difference of the labels. In comparison, direct-multidate classification follows a single-step approach via decision tree change detection on stacked features integrating across both spectral & temporal properties (Singh, 1989). A third classical technique is compound classification, which maximizes the posterior distribution of change and non-change assignments (Bruzzone et al., 1999)(Bruzzone et al., 2004) in the predicted change map.

With the advent of deep learning and its adoption by the remote sensing community (Ball et al., 2017) (Zhu et al., 2017), change detection in Earth observation took a paradigm shift from the classical methods described above and the application of deep neural networks became the leading approach. The seminal work of (Daudt et al., 2018b) collects data of Sentinel-2 observations and considers well-established neural network architectures for CD trained supervisedly on hand-annotated data. Our work builds upon this study by extending the the collected data with novel SAR observations and advances the methodology by proposing a novel data fusion neural network architecture for multi-modal change detection.

In terms of prior work on multi-modal change detection the following publications are of relevance. (Orsomando et al., 2007) detects change on a stack of SAR images and one multispectral observation. (Jiang et al., 2020) performs a transformation separating semantics and style to project SAR and optical images into a shared feature space. (Zhang et al., 2018) uses a Siamese architecture to detect building and tree changes between point cloud data and aerial observations (Chen et al., 2019) proposes a recurrent Siamese network for detecting change on multi-sensor very high resolution images of the same modality, and thus with relatively smaller domain differences.

Furthermore, the related work of (Liu et al., 2016), (Zhang et al., 2016), (Zhan et al., 2018) (Saha et al., 2019), (Ferraris et al., 2020) and (Saha et al., 2021) consider the challenging case where the pre-change observation may be captured by a sensor different from the one recording the post-change image. Whereas the first five consider an unsupervised training paradigm, (Saha et al., 2021) extends the prior work and proposes a method for self-supervised change detection between pairs of Sentinel-1 and Sentinel-2 imagery. Moreover, these methods have in common that representations of change are learned in scenarios where no sufficient amount of labeled training data is available, with potential effects on the quality of the learned features. While we also consider very heterogeneous pairings of SAR and optical data, our work differs in the sense that data is curated for our study to allow for a supervised training procedure. In addition, we focus on the data fusion case where both modalities are available as pre- and post-change inputs to the model.

These earlier contributions constitute fusion of multi-modal remote sensing data as a well-established research area in change detection and provides a vital starting point for our own contributions. To sum up, our work extends on the existing research by acknowledging the existence of very heterogeneous and more complex scenes, in which change may not be constrained to an individual class of land cover or objects in particular. Specifically, we build on the efforts of (Daudt et al., 2018b) and their hand-annotated data set of optical satellite observations to combine it with progress in data fusion for change detection. For this purpose, we curate co-registered and temporally aligned SAR observations for each of the bi-temporal change images, and demonstrate their purpose by introducing a novel Siamese network architecture for data fusion. The presence of a sufficiently large and hand-annotated data set allows for supervised training for bi-modal change detection—which is in contrast to most of the preceding work that, due to lack of training data, focuses on unsupervised methods. Taken together, these are the key characteristics that differentiate our contribution from prior work.

2. METHOD

We build on recent work in change detection for remote sensing and propose a deep neural network that is capable of integrating data from multiple sources. Specifically, we consider a Siamese network (Chicco, 2021) with a U-net architecture (Ronneberger et al., 2015). In Section 2.1 we introduce the Siamese network. Section 2.2 briefly outlines the usage of Siamese network in homogeneous (single-sensor) change detection. Finally the network for multi-sensor change detection is detailed in Section 2.3. Triplet loss is also used (Dong and Shen, 2018).

2.1 Siamese network

Siamese networks were first proposed in context of image matching (Bromley et al., 1993). A Siamese network consists of twin networks (or parts thereof) that generally share weight yet accept different inputs of the same dimensionality. Weight sharing ensures that two similar inputs are mapped to alike representations in the feature space since they are processed through a shared set of non-linear functions. The outputs of the twin networks are processed through an energy function that computes

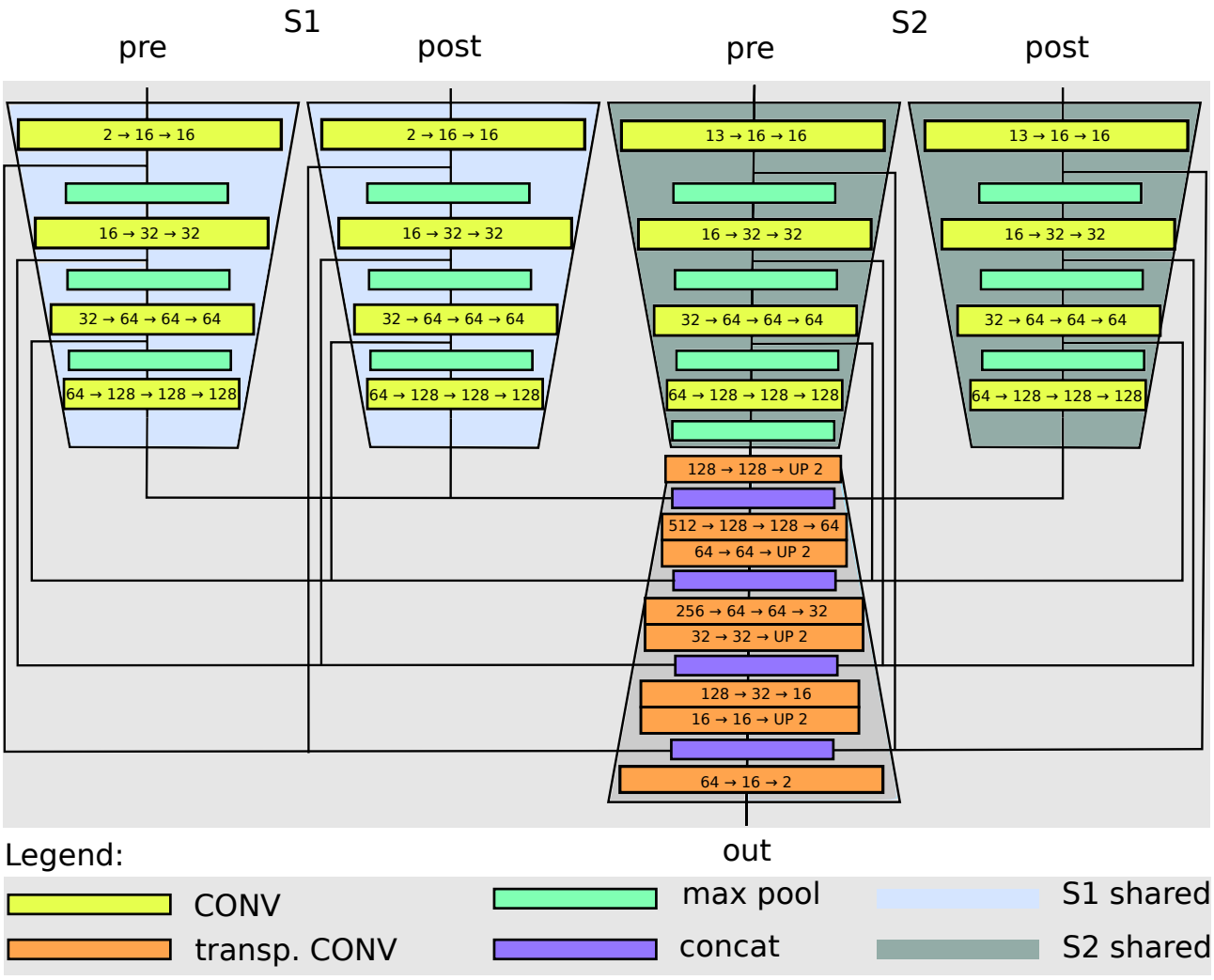

Figure 2. Proposed multi-modal Siamese architecture for CD. The network consists of two encoder branches for each sensore and a decoder part integrating the features from earlier layers. Each encoder branch processes its corresponding modality's bi-temporal samples, SAR and multispectral optical, in two passes. The extracted features get forwarded via skip connections in a U-Net like fashion and then concatenated. Figure style adopted from (Daudt et al., 2018a).

a similarity metric between the highest level feature representations for each of the inputs propagated through the model. Contrastive loss is generally used owing to its ability to increase the distance between dissimilar pairs and decrease the distance between similar pairs (Koch et al., 2015). In sum, Siamese networks are an appealing architecture for tasks that benefit from similarity of discrepancy-sensitive feature learning when comparing two or more input types.

2.2 Siamese network for homogeneous CD

The Siamese models used in the context of homogeneous CD benefit of the principles described above. Two weight-sharing networks (or components of a single network) are used for high-level feature extraction from the pre-change and post-change images. After, the inferred high-level features are processed through a decision network. The decision network segregates the changed pixels from the unchanged ones. Rather than just processing the highest-level feature from the last weight-sharing layer, multi-level features from multiple layers are often concatenated to obtain a multi-scale representation of the change information (Rahman et al., 2018). Considering the relatedness of the pixelwise change detection problem to semantic segmentation, U-net (Ronneberger et al., 2015) is generally used as backbone architecture for Siamese CD. U-net is appealing for both tasks as the architecture is composed of processing information across two distinct pathways, one that preserve resolution while being relatively shallow (processing information about the *where* of content) and the other being deep and wide but with less spatial resolution (focusing on the *what* of content). The Fully Convolutional Siamese - Concatenation (Siamese (S2 only)) architecture of (Daudt et al., 2018b) uses U-net as backbone with a 10 layer encoder to process the pre-change and post-change observations via two passes of one Siamese branch. Skip connections in the network's decoder component concatenate multi-scale information coming from the two encoding streams of the pre- and post change images. Our proposed method builds on and extends the Siamese (S2 only) network to a multi-modal model, as detailed next.

2.3 Multi-sensor Siamese architecture

The Siamese network architecture proposed in this work (abbreviated as: ours) consists of two encoder branches, one per considered sensor type, and a decoder part integrating the features from the preceding layers on a multi-scale basis. Each encoder branch processes its modality's bi-temporal samples, SAR and multispectral optical, in two passes. The extracted features get

forwarded via skip connections in a U-Net like fashion and then concatenated at the individual levels of depth. Fig. 2 depicts the proposed Siamese neural network architecture.

Similar to the Siamese (S2 only) baseline, the encoder part consists of 10 convolutional layers (of $3 \times 3 px^2$ kernel size, a stride of $1\ px$ and a padding of $1\ px$), each followed by operators of batch normalisation, rectified linear units (ReLU) and dropout ($p = 0.2$). In the encoder, every convolution blocks indicated in Fig. 2 is followed by a layer of max pooling (of $2 \times 2\ px^2$ kernel size and a stride of $2\ px$). interleaved max pooling operations. The network's decoder component consists of 14 transposed convolution layers (of $3 \times 3\ px^2$ kernel size, a stride of $2\ px$, a padding of $1\ px$ and an padding of $1\ px$), followed by operators of batch normalisation, ReLU and dropout ($p = 0.2$). At the end of each of each differentiable upsampling block follows a layer of replication padding as well as a concatenation layer stacking earlier decoder features in a U-Net manner.

The described encoder-decoder architecture follows a conventional hourglass style with the bottleneck being the widest part and fewer kernels at the start as well as the end of the network. The last decoding layer reduces the features into a change map of only two bands representing changed and unchanged pixels accordingly, with a log softmax nonlinearity appplied. The predicted class, i.e. whether change or non-change, is then given by taking the maximum value across both bands in the output.

3. EXPERIMENTS AND ANALYSIS

3.1 Data

To conduct experiments we collect and process a multi-modal data set and acquire SAR observations as follows: The geospatial locations of each ROI and the acquisition dates of their original observations Sentinel-2 multi-spectral observations are read from the meta information of the ONERA CD data set (Daudt et al., 2018b). The (ascending orbit) Sentinel-1 SAR observations are downloaded via Google Earth Engine (Gorelick et al., 2017) and coordinate-transformed via GDAL (Warmerdam, 2008) to match the coordinate system of the original optical data. Exemplary full-scene observations for one ROI are illustrated in Fig. 1. Finally, all full-scene images are sliced online into patches of sizes 96 pixels $\times$ 96 pixels with a stride of 1 pixel between spatially adjacent patches. The train and test splits of the data set are as defined in (Daudt et al., 2018b). The SAR and multi-spectral optical patches are value-clipped in the intervals [-25, 0] and [0, 1000], respectively. Finally, patches are normalized to the unit range for input to the networks. For the baselines this is done via z-standardizing each patch individually (such that the extreme values are guaranteed to be taken per patch) as suggested in (Daudt et al., 2018a). For our proposed Siamese fusion network we experimentally observed that standardizing in absolute terms outperforms z-scoring, so patches are rescaled according to their modality's theoretically obtainable range rather than z-scoring, preserving band-wise information in absolute terms across observations.

To facilitate future research in remote sensing on bi-temporal change detection of multi-modal satellite observations we wish to share our data with the scientific community. Our SAR observations specifically collected and preprocessed for this study can be found online on `https://github.com/PatrickTUM/multimodalCD_ISPRS21`.

3.2 Experiments & Results

To address the research question stated in section 1 we train the neural network proposed in section 2 on the data set introduced in section 3.1 and compare it against baseline models utilizing just a single sensor as well as networks utilizing both sensors but exercising less guidance on the fusion process. The baselines compared against are given as follows:

1. *Siamese (S2 only)* that corresponds to the FC-Siam-conc setup from (Daudt et al., 2018a) as detailed in Section 2.2 and relies on S2 inputs only.

2. *Siamese (S1+S2)* that follows the Siamese (S2 only) architecture but with stacked S1 and S2 observations combined into a singly input tensor and processed jointly. That is, other than our model, no explicitly separate processing of modalities is taking place.

3. *U-Net (S2 only)* that stacks the channels from pre-change and post-change images into a single image and then processes it through a U-Net treating change detection as a semantic segmentation task.

4. *U-Net (S1+S2 only)* that works similarly as U-Net (S2 only) but in addition to stacking pre- and post-change S2 images it also combines S1 and S2 cross-modality into a singly input tensor and handles everything jointly without imposing further constraints on the structure of information processing.

All networks considered are trained in a supervised manner via the ADAM optimizer (Kingma and Ba, 2014) with a weight decay of $1e{-}4$ and an exponential decay learning rate scheduler on a cross-entropy loss on the collected data set. The cross-entropy cost function is weighted according to

$$2 \times \lambda_{FP} \times \frac{n_{positive}}{n_{total}}, 2 \times \frac{(n_{total} - n_{positive})}{n_{total}}$$

for the respective classes, where we set $\lambda_{FP} = 10$ as a parameter and n_{total} and $n_{positive}$ denote the number of total and positively labeled pixels in the training split, respectively. The networks are trained on batches of 32 samples. Random rotations (in steps of 90 degrees) or mirroring (on the vertical axis) are applied as data augmentation steps at equally distributed chances to synthetically increase the training set size.

The goodness of predictions are evaluated in terms of the well-established metrics of precision, recall and F1 score, given by

$$precision = \frac{TP}{TP + FP},$$

$$recall = \frac{TP}{TP + FN},$$

$$F1\ score = \frac{2 \times precision \times recall}{precision + recall},$$

where TP, FP and FN denote true positives, false positives and false negatives, respectively. The F1 score is the harmonic

model	precision	recall	F1 score
ours	0.602	**0.561**	**0.581**
Siamese (S2 only)	0.680	0.494	0.573
Siamese (S1+S2)	0.699	0.412	0.519
U-Net (S2 only)	**0.762**	0.394	0.519
U-Net (S1+S2)	0.562	0.255	0.351

Table 1. Performance of the evaluated change detection models on the ONERA test split. The proposed multi-modal network outperforms the considered baseline and is strongest in terms of both recall and F1 score.

mean of both precision as well as recall and provides a summary statistics of a considered method's overall accuracy.

Results are reported in Table 1 and show that the proposed model outperforms the considered baselines. It is evident from the reported numbers that our fusion-based method outperforms the other S2-only models, accomplishing a considerably improved recall score and an overall increase in terms of F1 metric as well. Remarkably, solely feeding S1 and S2 combined inputs to the standard Siamese and U-Net architectures does not guarantee any increase in performance but may even be detrimental. This may indicate that fusing diverse modalities such as S1 and S2 together necessitates more guidance (as provided by our proposed architecture) than merely stacking them together. The predictions of the proposed model and the second best method, the Siamese (S2 only) baseline, on data of three exemplary ROI are displayed in Fig. 3. The results show that both models share many of the correctly predicted changes, indicating that these pixels may exhibit change that is clearer to detect than more ambiguous change in other parts of the scenes. Interestingly, our proposed model has a tendency to correctly detect more change, but it may also be more prone to false alarms—a circumstance that is discussed further in section 4.

4. DISCUSSION

Change detection in remote sensing poses a challenging task as the typical scenes considered by practitioners are very complex. The images utilized for training and testing in this study are constituted by spatial arrangements of many objects which are themselves often not constrained or clearly defined in terms of their land cover or object class. While complementary views on the scene (as given by multi-modal observations) can ease this uncertainty and be of benefit, bridging the difference between two very heterogeneous domains and integrating sensor information is all by itself a nontrivial task. One contribution of our work is to provide the scientific community with the needed data to collectively address this challenge, as well as proposing a novel deep neural network architecture demonstrating the benefits of multi-modal change detection.

Our work builds on the original data set of (Daudt et al., 2018a) and its high-quality annotations hand-labeled by introspecting Sentinel-2 data. While the presence of labels allows for supervised training and competitive performances, it may as well pose a limitation to our study as the provided supervision may not always capture change perfectly. An example is given in Fig. 4, where the upper right image quartile displays clear change between the pre- and post-change Sentinel-1 observations but the subtle differences are barely visible in the RGB plots of the Sentinel-2 data and consequently not annotated in the labels. This may lead to predicted change where there is none annotated, eventually raising the false positives (compared

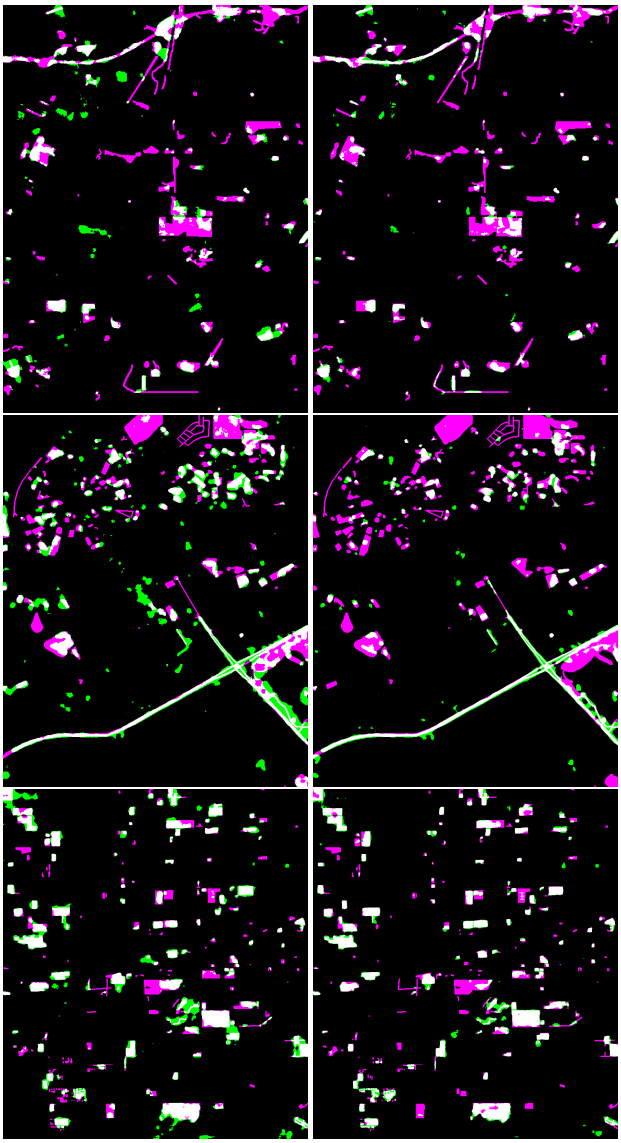

Figure 3. Exemplary change predictions. Rows: Proposed method and optical-only Siamese baseline model. Column: Three different ROI. Chonqqing, Duba and Las Vegas. Labels: White (true positive), green (false positive), violet (false negative). The results show that the proposed model is more sensitive to change, yet may also be more prone to false alarms, as compared to the baseline.

to the optical-only baseline) as exemplified in Fig. 3. While this point highlights the complementing nature of both modalities, it may put SAR data at a disadvantage when evaluating on labels driven by optical information, raising the more principal question of what modification of pixel intensities in which modality should actually count as a change in the ground truth.

Furthermore, the results presented in section 3.2 demonstrated the benefits of multi-modal data in combination with our designed architecture, but the relative improvement over the strong second best method is not very large. This may reflect the order of improvement provided by the proposed neural network. Alternatively, a saturation effect of the F1 score just around 0.6 may be natural on the considered data set, the given training split size and the challenging test split scenes. Similarly

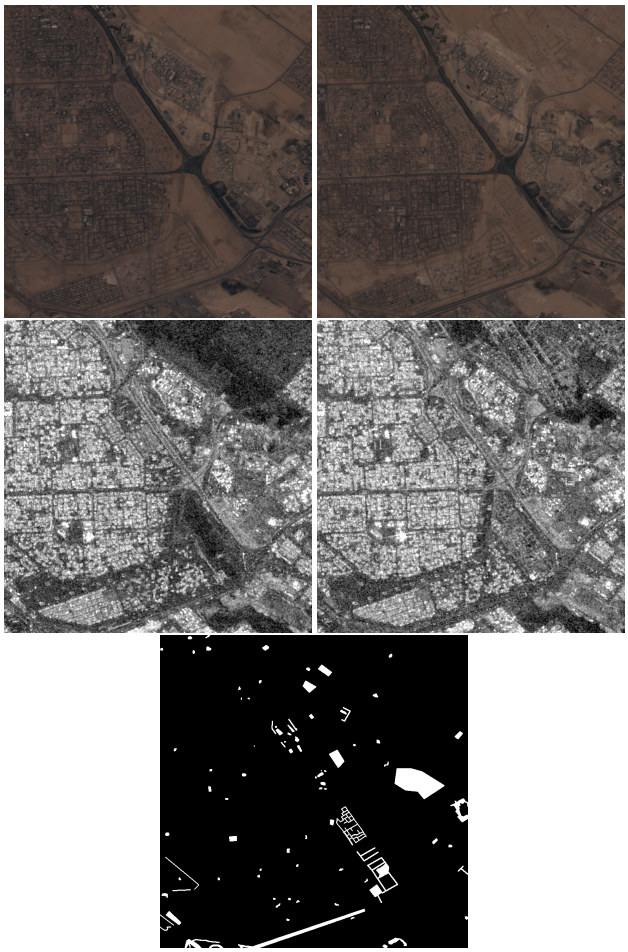

Figure 4. Multi-modal observations and change map for exemplary ROI 'Abu Dhabi'. Rows: Sentinel-2 data (RGB channels). Sentinel-1 data (VV-polarized). Change maps. Columns: Time point 1. Time point 2. The images indicate discrepancies between SAR versus multispectral data and the change labels, raising a question about the semantics of change.

(Daudt et al., 2018a) reported considerable gains in the range of poorer performances, but the degree of improvements as well decreased around the marks observed in our study. In sum, we are positive that future research building on our data will provide valuable insights with respect to this point and advance the state of the art of multi-modal change detection.

5. CONCLUSION

This work addressed the challenge of multi-modal bi-temporal change detection. We investigated the central research question of whether multi-modal fusion approaches benefit bi-temporal change detection. To evaluate this question on a substantially large amount of train and test images, we extended an existing and well-established single-sensor S2 data set by complementing it with corresponding S1 images curated for this study. Furthermore, we proposed a novel architecture for bi-modal fusion based change detection that integrates information from both SAR as well as optical sensors. The results show that bi-modal fusion improves result over single-sensor approach. Though the improvement is not large, further improvement in multi-modal fusion architecture can potentially improve the result. We emphasize that the contribution of this work is not only limited to devising a novel approach for multi-sensor change detection, but further opens up the research towards this direction by making available a novel data set and encouraging further research in the direction. In addition to improving the proposed architecture, our future work will focus on extending the proposed method for integrating more than two modalities. We will also extend the proposed method for more challenging problem of multi-class or semantic change detection.

ACKNOWLEDGEMENTS

The work is supported by the German Federal Ministry of Education and Research (BMBF) in the framework of the international future AI lab "AI4EO – Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond", Grant number: 01DD20001.

REFERENCES

Ball, J. E., Anderson, D. T., Chan Sr, C. S., 2017. Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. *Journal of Applied Remote Sensing*, 11(4), 042609.

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., Shah, R., 1993. Signature verification using a" siamese" time delay neural network. *Advances in neural information processing systems*, 6, 737–744.

Bruzzone, L., Cossu, R., Vernazza, G., 2004. Detection of land-cover transitions by combining multidate classifiers. *Pattern Recognition Letters*, 25(13), 1491–1500.

Bruzzone, L., Prieto, D. F., Serpico, S. B., 1999. A neural-statistical approach to multitemporal and multisource remote-sensing image classification. *IEEE Transactions on Geoscience and remote Sensing*, 37(3), 1350–1359.

Chen, H., Wu, C., Du, B., Zhang, L., Wang, L., 2019. Change detection in multisource VHR images via deep Siamese convolutional multiple-layers recurrent neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 58(4), 2848–2864.

Chicco, D., 2021. Siamese neural networks: An overview. *Artificial Neural Networks*, 73–94.

Daudt, R. C., Le Saux, B., Boulch, A., 2018a. Fully convolutional siamese networks for change detection. *2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, 4063–4067.

Daudt, R. C., Le Saux, B., Boulch, A., Gousseau, Y., 2018b. Urban change detection for multispectral earth observation using convolutional neural networks. *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 2115–2118.

Dong, X., Shen, J., 2018. Triplet loss in siamese network for object tracking. *Proceedings of the European conference on computer vision (ECCV)*, 459–474.

Ferraris, V., Dobigeon, N., Cavalcanti, Y., Oberlin, T., Chabert, M., 2020. Unsupervised change detection for multimodal remote sensing images via coupled dictionary learning and sparse coding. *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 4627–4631.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202, 18–27.

Jiang, X., Li, G., Liu, Y., Zhang, X.-P., He, Y., 2020. Change detection in heterogeneous optical and SAR remote sensing images via deep homogeneous feature fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 1551–1566.

Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Koch, G., Zemel, R., Salakhutdinov, R., 2015. Siamese neural networks for one-shot image recognition. *ICML deep learning workshop*, 2, Lille.

Liu, J., Gong, M., Qin, K., Zhang, P., 2016. A deep convolutional coupling network for change detection based on heterogeneous optical and radar images. *IEEE transactions on neural networks and learning systems*, 29(3), 545–559.

Malila, W. A., 1980. Change vector analysis: an approach for detecting forest changes with landsat. *LARS symposia*, 385.

Orsomando, F., Lombardo, P., Zavagli, M., Costantini, M., 2007. SAR and optical data fusion for change detection. *2007 Urban Remote Sensing Joint Event*, IEEE, 1–9.

Rahman, F., Vasu, B., Van Cor, J., Kerekes, J., Savakis, A., 2018. Siamese network with multi-level features for patch-based change detection in satellite imagery. *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, IEEE, 958–962.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, Springer, 234–241.

Saha, S., Bovolo, F., Bruzzone, L., 2019. Unsupervised multiple-change detection in vhr multisensor images via deep-learning based adaptation. *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 5033–5036.

Saha, S., Ebel, P., Zhu, X. X., 2021. Self-supervised Multisensor Change Detection. *arXiv preprint arXiv:2103.05102*.

Singh, A., 1989. Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, 10(6), 989–1003.

Warmerdam, F., 2008. The geospatial data abstraction library. *Open source approaches in spatial data handling*, Springer, 87–104.

Zhan, T., Gong, M., Jiang, X., Li, S., 2018. Log-based transformation feature learning for change detection in heterogeneous images. *IEEE Geoscience and Remote Sensing Letters*, 15(9), 1352–1356.

Zhang, P., Gong, M., Su, L., Liu, J., Li, Z., 2016. Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116, 24–41.

Zhang, Z., Vosselman, G., Gerke, M., Tuia, D., Yang, M. Y., 2018. Change detection between multimodal remote sensing data using siamese CNN. *arXiv preprint arXiv:1807.09562*.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36.