Deep Feature and Domain Knowledge Fusion Network for Mapping Surface Water Bodies by Fusing Google Earth RGB and Sentinel-2 images

Pu Zhou, Xiaodong Li, Giles M. Foody, *IEEE Fellow*, Doreen S. Boyd, Xia Wang, Feng Ling, Yihang Zhang, Yalan Wang, Yun Du

Abstract—**Mapping surface water bodies from fine spatial resolution optical remote sensing imagery is essential for the understanding of the global hydrologic cycle. Although satellite data are useful for mapping, the limited spectral information captured by some satellite systems can be sub-optimal for the task. For example, the very high resolution images of Google Earth (GE) only contain RGB bands, which often means many water bodies and land objects are confused. Sentinel-2 (S2) imagery have a spectral resolution more suitable for mapping water bodies, but its medium spatial resolution limits the ability for detailed mapping of water-land boundaries. This letter proposes a deep feature and domain knowledge fusion network (DFDKFNet) for mapping surface water bodies by fusing GE and S2 images while incorporating domain knowledge. DFDKFNet uses the remote sensing indices of normalized difference water index (NDWI) and normalized difference vegetation index (NDVI) derived from the S2 image as the representative domain knowledge to better extract water bodies from terrestrial features. A similar pixel-based approach is used to downscaling the NDWI and NDVI maps to match the spatial resolution between the GE and S2 images. The DFDKFNet uses the GE and downscaled NDWI and NDVI images to extract the deep semantic features of water bodies, which are fused with the domain knowledge extracted from the NDWI and NDVI images. DFDKFNet was compared with several state-of-the-art algorithms, and the results show that DFDKFNet can enhance water body mapping accuracy.**

Index Terms—**Water body, deep learning, domain knowledge.**

P. Zhou and Y. Wang are with the Key Laboratory for Environment and Disaster Monitoring and Evaluation, Hubei, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan 430077, China, and University of Chinese Academy of Sciences, Beijing 100049, China.

X. Li, F. Ling, Y. Zhang and Y. Du are with the Key Laboratory for Environment and Disaster Monitoring and Evaluation, Hubei, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan 430077, China (email: lixiaodong@whigg.ac.cn).

G. M. Foody and D. S. Boyd are with School of Geography, University of Nottingham, Nottingham, NG7 2RD, UK.

X. Wang is with CAS Key Laboratory of Aquatic Botany and Watershed Ecology, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China and with the Research Center for Environmental Ecology and Engineering, School of Environmental Ecology and Biological Engineering, Wuhan Institute of Technology, Wuhan 430205, China.

The code is available online https://github.com/zhoupu1997/DFDKFNet.git.

I. INTRODUCTION

Remote sensing can map the surface water bodies which are vital to environmental systems and processes [1, 2]. In recent years, shallow machine learning classifiers such as support vector machine have been used to map surface water from remote sensing imagery. Compared with the shallow classifiers, deep learning (DL) convolutional neural networks can extract inherent and deep-level features from a large amount of training data and have great potential in mapping water bodies. Many DL methods, including FCN8s [3], UNet [4, 5] DeepLabV3+ [6], and HRNet [7], as well as their derivative networks, have been widely applied for water body mappings from satellite images.

Commonly, DL methods are applied to map water bodies from fine-spatial-resolution multispectral remote sensing imagery [8]. One of the biggest challenges is that fine-resolution imagery have only limited spectral information for water body mapping [9]. Most fine resolution images such as PlanetScope do not supply bands such as the short-wave infrared band in which water and land are distinguishable. The limitation in spectral bands is more severe for DL methods when applying them to high spatial but low spectral resolution RGB images such as those obtained from GE [10] or captured from sensors carried on accessible unoccupied aerial vehicles [11]. In the GE RGB images, water bodies may sometimes be confused with terrestrial features (e.g. shadowed areas) [12]. In addition, different water bodies have different chemical (e.g. chlorophyll) concentrations and physical components (such as sediment content). Lastly, the sun glint effect gives rise to different RGB colors for water bodies with different surface water roughness in GE images. The data-driven DL methods require a large number of training samples, but the collection of representative samples considering the aforementioned aspects is usually difficult [13].

Incorporating RGB images with multispectral images helps increase the surface water mapping accuracy of DL [9]. For instance, Yuan et al. [14] demonstrated that DL using both RGB and multispectral bands dealt better with the wide range of GE color shifts and outperformed the DL using only RGB bands. Although fusing RGB and multispectral images is promising in DL water mapping, challenges still exist.

First, current water mapping studies fuse RGB and multispectral bands from the same satellite sensor [14, 15], and the fusion from different satellite sensors has not been
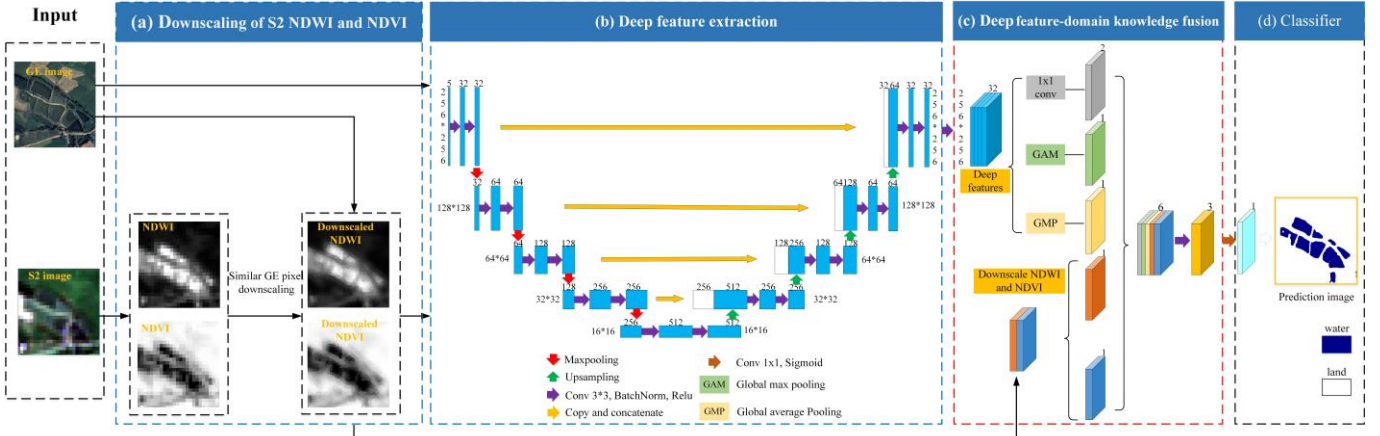
Fig. 1. Illustration of the DFDKFNet architecture. DFDKFNet has four components: (a) Downscaling of S2 NDWI and NDVI, (b) Deep feature extraction module, (c) Deep feature-domain knowledge fusion module, and (d) Classifier.

reported to the best of our knowledge. Among the range of remote sensing images available, the S2 multispectral image with near-infrared and shortwave infrared bands has great potential in the fusion with GE RGB images. The combination of GE with S2 images, which are acquired at the same or similar date, assuming there is no land cover change between them, could enhance the inter-class separability between water and terrestrial features in the GE image. However, there is a large gap in the spatial resolution between S2 (typically 10 m) and GE images (typically about 1 m). The simplest resampling algorithms (such as bilinear interpolation) may generate blurred boundaries if the scale factor between the original and output image is too large. It is, therefore, necessary to develop a new downscaling method before the fusion of GE and S2.

Second, the DL fusion of RGB and multispectral bands may be not significantly superior to DL using only RGB bands, because DL mixes the RGB and multispectral values through concatenation without fully utilizing the independent information from each [14, 16]. The incorporation of domain knowledge is an efficient way to guide the DL training process and better use independent information [17]. The domain knowledge is represented in various forms such as prior information, but is usually unavailable or laborious to obtain from experts [18]. Remote sensing indices (e.g., NDVI), which are usually easily obtained from multispectral images, have been tested as representative domain knowledge for remote sensing image classification with high generalizability [17]. However, the use of this domain knowledge has not been reported in water body mapping from GE and S2 images.

In this letter, a novel deep feature and domain knowledge fusion network (DFDKFNet) is proposed to map water bodies by fusing GE RGB and S2 imagery. The NDWI and NDVI images, which help to distinguish water from background objects, are extracted from the S2 image as representative domain knowledge. Then, the 10 m NDWI and NDVI images are downscaled to 1 m using similar GE pixels. Lastly, the DFDKFNet, which contains a feature extraction module, a deep feature-domain knowledge fusion module, and a classifier, is constructed (Fig. 1). The novelty of this letter is that, unlike previous studies that fuse RGB and multispectral
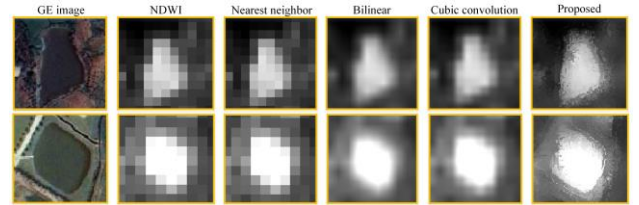

Fig. 2. Examples of different Sentinel-2 NDWI image downscaling results. The nearest neighbor, bilinear, and cubic convolution maps are blurred.

bands from the same satellite sensor with the same or similar spatial resolution, DFDKFNet fuses GE with S2 imagery from different satellite sensors with a large spatial resolution gap in surface water mapping. Moreover, unlike traditional DL methods that concatenate multisource images at the input level and feed them into the network directly, DFDKFNet extracts the domain knowledge from the NDWI and NDVI images to increase the accuracy with which water bodies are extracted. DFDKFNet was assessed and compared with several state-of-the-art DL methods.

II. METHODOLOGY

A. Downscaling S2 NDWI and NDVI using similar GE pixels

The Sentinel-2 and Google Earth images were geo-registration [19], and the NDWI and NDVI images were extracted from the S2 image at 10 m resolution. A similar GE pixel-based downscaling is used to downscale the NDWI and NDVI image to 1 m GE resolution to reduce the blocky effects compared with the nearest neighbor and bilinear interpolations (Fig. 2). First, the NDWI and NDVI images are resampled to 1 m resolution using the nearest neighbor method. Then, for each resampled pixel, the NDWI (or NDVI) value is defined as a weighted function based on similar GE pixels, according to the assumption that if the pixels have similar RGB values in a local window in the GE image, then they would probably belong to the same class and thus have similar NDWI (or NDVI) values in the local window. For the kth target GE pixel (i.e., a_k), a local window centered at a_k is with the window size W is defined. The similar GE pixels are selected according to the smallest difference in RGB values between the target pixel and the pixels within the local window in the GE image.

The difference in RGB values between a_k and a GE pixel a_n within the local window is calculated as:

$$D_n = \sum_{b=1}^{3} (|y_{k,b} - y_{n,b}| / y_{k,b}) \qquad (1)$$

where $y_{k,b}$ and $y_{n,b}$ are the digital number values for the kth target GE pixel a_k and the nth neighborhood GE pixel a_n within the local window centered at a_k, and D_n is the difference in RGB value between a_k and the nth GE pixel. b is the number of bands in GE image. A number of GE pixels with the smallest RGB difference are selected as the similar GE neighboring pixels for the target GE pixel a_k. The window size W is set as 10, and the number of similar GE pixels is set as 20 through many trials [20]. Finally, the NDWI (or NDVI) value for the target GE pixel a_k is calculated according to a weighting function of all the selected similar neighboring pixels:

$$I_k = \sum_{n=1}^{w^2} (w_n \times I_n) \qquad (2)$$

where I_n is the index value in the resampled NDWI or NDVI images. w_n is the weight of the nth similar neighborhood GE pixel, which is calculated based on the geographical distance between the target pixel a_k and the neighborhood pixel a_n [20].

B. Feature Extraction Module (FEM)

The downscaled NDWI and downscaled NDVI images are concatenated with the GE RGB images, which were acquired at the same or similar date, to extract surface water body features using a FEM. The FEM is based on the structure of UNet, which has a simple structure and lightweight parameters and is efficient in image segmentation (Fig. 1 (b)) [8]. FEM contains an encoder and a decoder part. The encoder has the same structure as UNet, which contains five repeated blocks of convolutional layers, in which each block contains two convolutional layers with a 3×3 kernel size, and each layer is followed by a batch normalization layer and a rectified linear unit (ReLU). A 2×2 max pooling layer with stride 2 is added at the end of each block for downsampling except for the first block. The last layer of the origin UNet decoder is removed, and the number of feature channels after each block of the encoder is 32, 64, 128, 256, and 512. The decoder part consists of four repeated blocks. Different from the blocks in the encoder part, the blocks in the decoder part replace 2×2 max pooling layer with a 2×2 upsampling layer. In addition, the skip connection is adopted to concatenate the different level feature maps in the channel dimension between the encoder and decoder. Thus, the number of feature channels at the end of each block of the decoder is 256, 128, 64, and 32.

C. Deep Feature-Domain Knowledge Fusion Module (DFDKM)

In order to reduce the confusion between various semantic features and inefficient utilization of image information from the feature extraction module, a DFDKM is added to fuse the deep features from FEM and domain knowledge from downscaled NDWI and NDVI images in Fig. 1 (c). The deep features extracted from the FEM module are fed into a 3×3 convolution with 2 filters (and batch normalization and ReLU

layer) to capture local context features. Global max-pooling (GMP) and global average pooling (GAP) are applied in the channel dimension for capturing global context features, respectively. The downscaled NDWI and NDVI images are then concatenated with the local and global context feature maps. The concatenated feature maps are subjected into two 3×3 convolution layers with 6 and 3 filters (and batch normalization and ReLU layer) for fusing various receptive fields and spectral information. The concatenated feature maps are input into a batch normalization layer before they are fed into the next convolution blocks to make different features have the same distribution and converge faster. The DFDKM could be regarded as a kind of ensemble learning that combines the downscaled NDWI and NDVI images with the outputs through GAP, GMP, and 1×1 convolution layers by learning the weights and bias of convolution kernels automatically.

D. Classifier

The classifier is made up of a 1×1 convolution with 1 filter and a sigmoid function to produce the binary prediction (Fig. 1 (d)). The classifier loss function is defined as follows:

$$loss = -\frac{1}{N} \sum_{i=1}^{N} [p_i \cdot \log \sigma(q_i) + (1-p_i) \cdot \log(1-\sigma(q_i))] \quad (3)$$

where p_i is the reference label of pixel i, $p_i = 0$ indicates land class and $p_i = 1$ indicates water class; q_i is the probability of the ith pixel belonging to the water class; N is the total number of pixels and σ is the sigmoid function.

III. EXPERIMENTS

A. Data

A surface water body dataset containing both GE and corresponding S2 images, which were acquired at the same or similar date, was constructed. The GE and S2 images used for training, validation, and testing were cloud free. All the images were taken from urban and rural areas in China and acquired at different seasons. These images contain various types of water bodies, including ponds, rivers, paddies, and lakes. The NDWI and NDVI images are calculated and downscaled to 1 m resolution using the similar GE pixel method. The GE and the corresponding downscaled NDWI and NDVI images were randomly cropped to patches of 256×256 pixels to get a total of 11343 GE/S2 patches. Data augmentation operations including scaling and rotations were applied to generate a total of 22648 GE/S2 patches used for training and validation. Besides, another three GE/S2 image pairs were selected to test DFDKFNet (Fig. 3). Each selected image has an area of larger than 100 km^2. For each GE/S2 image pair, the acquisition times between GE and S2 are shorter than 10 days to reduce the impact of land cover change. The reference water map was produced by visual interpretation of the GE images (Fig. 3).

B. Comparison Methods and Result Assessment

The proposed DFDKFNet was compared with several state-of-the-art DL methods, including UNet, FCN8s, HRNet, and DeepLabV3+. The performance of the comparison methods using GE RGB image (namely UNet_GE,

FCN8s_GE, HRNet_GE, and DeepLabV3+_GE) and using both GE and downscaled S2 NDWI and NDVI images (UNet_GE&S2, FCN8s_GE&S2, HRNet_GE&S2, and DeepLabV3+_GE&S2) were assessed. While the DFDKFNet uses domain information from S2 data, the comparator DL methods using both GE and S2 images simply concatenated the GE RGB bands with downscaled NDWI and NDVI bands as network input. Ablation experiments were also conducted for DFDKFNet. The DFDKFNet using S2 and GE image but without using the DFDKM model is the same as the UNet_GE&S2, and the DFDKFNet using only GE image but without using the DFDKM model is the same as UNet_GE. Five indexes, including overall accuracy (OA), F1 score (F1), Intersection over Union (IoU), Precision, and Recall, were used to assess the accuracy of different methods.

C. Implementation Platform and Parameters

Python 3.6 and the open-source deep learning framework PyTorch was used. The GPU is NVIDIA 2060 which has 6 GB of RAM and uses cuDNN 10.0 for acceleration. Mini-batch stochastic gradient descent (SGD) was used for training. The initial learning rate and weight decay were set as 1e-4 and 1e-5. The binary cross-entropy in PyTorch was used as the loss function, and the Adam optimization algorithm was used for gradient descent. A five-folds cross-validation was used to partition the datasets of 22648 GE/S2 patches and evaluate the performance of each method. Each model was trained for 200 epochs and the training weights that resulted in the highest validation accuracy were saved for that model. We averaged the scores of the five cross-validation folds for the accuracy assessment of each method.

IV. RESULTS

According to the accuracy metrics in Table I, all the DL methods using only the GE images are inferior to those incorporating the S2 images to exploit additional surface water information. In general, the DL methods using both GE and S2 increased OA, F1, precision, recall, and IoU by about 0.02, 0.21, 0.28, 0.03, and 0.27, respectively, compared with DL using only the GE image. This finding verifies that fusing GE RGB images with S2 images can enhance the performance of DL in surface water mapping. Among the DL methods using both GE and S2 imagery, DFDKFNet typically generated higher accuracies than FCN8s_GE&S2, DeepLabV3+_GE&S2, and HRNet_GE&S2. This finding shows that, in mapping water bodies, the incorporation of domain knowledge from the NDWI and NDVI images through DFDKM can typically increase the accuracy of the DL models which simply concatenate the downscaled NDWI, NDVI, and GE images. DFDKFNet generated the highest OA and IoU, showing the highest percentage of pixels correctly classified and the highest overlap between predictions and the ground truth, respectively. DFDKFNet generated the highest F1, showing the method well-balanced recall and precision.

Fig. 3 shows the DFDKFNet water map in three study areas. The outlines of water bodies in the DFDKFNet are similar to the reference maps in Fig.3. Fig. 4 shows the visualization of
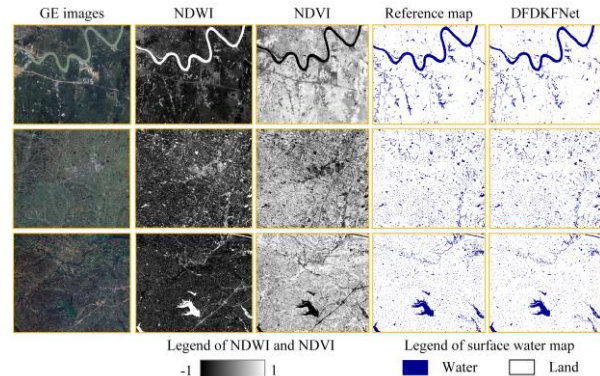

Fig. 3. GE, S2 NDWI and NDVI images, reference map, and the DFDKFNet prediction map.

five zoom areas predicted from different methods. In zoom areas I and II, the ponds highlighted with red ellipses and circles resemble dense vegetation in the GE image, but are very distinguishable in the S2 NDWI and NDVI images. The DL methods using only the GE image, including UNet_GE, FCN8s_GE, HRNet_GE, and DeepLabV3+_GE, failed or partly mapped these ponds, while the DL methods using both GE and S2 imagery better detected these ponds. In zoom area III, the dark bareland highlighted with a red ellipse resembles the pond in the GE image, but it has a low NDWI value and is dissimilar to water in the NDWI image. The DL methods using only the GE image incorrectly mapped the bareland as water, while the DL methods using both GE and S2 imagery correctly mapped it as land. These findings show that incorporating the S2 image in DL can reduce both the omission and commission errors in water mapping. In zoom area IV, most water maps from the comparison methods contain the linear river with a disconnected shape, while the DFDKFNet map is more similar to the reference. In zoom area V, DFDKFNet predicted water-land boundaries better than the comparison DL methods that combined GE and S2. In all the zoom areas, the DFDKFNet maps are more similar to the reference maps than the comparison maps, showing the DL using domain knowledge from S2 outperforms the DL that simply concatenates the GE and S2 bands.

V. CONCLUSION

A new DFDKFNet that combines GE and S2 to more fully utilize the complementary information of the two datasets was proposed for surface water body mapping. While the GE RGB imagery is frequently used in mapping surface water at a very fine spatial resolution, we show that the fusion of S2 images in DL can effectively improve the water body mapping accuracy, even if the fused images are from different satellite sensors and have a large spatial resolution gap. Experimental results show that DL applied to only the GE image resulted in omission errors in regions where water bodies resemble dense vegetation, and resulted in commission errors in regions where dark land objects are present. The DL methods that fuse GE with S2 reduced the degree of confusion between water and land which are distinguishable in the S2 NDWI and NDVI images in water body mapping. The proposed method uses DFDKM to incorporate deep semantic features of water

TABLE I

PERFORMANCE COMPARISON OF DIFFERENT METHODS TO THE THREE TEST AREAS. HIGHEST VALUES ARE HIGHLIGHTED IN BOLD.

| Methods	Test area A					Test area B					Test area C				
	OA	F1	Precision	Recall	IoU	OA	F1	Precision	Recall	IoU	OA	F1	Precision	Recall	IoU
UNet_GE	0.9655	0.7594	0.6398	0.9494	0.6180	0.9636	0.7198	0.5939	0.9290	0.5675	0.9593	0.6636	0.5232	0.9259	0.5015
UNet_GE&S2	0.9834	0.8988	0.8430	**0.9633**	0.8166	0.9833	0.8900	0.8407	**0.9465**	0.8024	0.9818	0.8746	0.8124	**0.9483**	0.7778
FCN8s_GE	0.9512	0.6144	0.4765	0.9194	0.4611	0.9481	0.5485	0.4073	0.8791	0.3877	0.9450	0.4890	0.3452	0.8701	0.3300
FCN8s_GE&S2	0.9830	0.8968	0.8430	0.9582	0.8131	0.9816	0.8790	0.8318	0.9323	0.7843	0.9790	0.8541	0.7870	0.9344	0.7457
HRNet_GE	0.9690	0.7920	0.6937	0.9317	0.6623	0.9672	0.7634	0.6628	0.9046	0.6200	0.9626	0.7091	0.5876	0.9010	0.5517
HRNet_GE&S2	0.9846	0.9087	**0.8758**	0.9448	0.8328	0.9833	0.8927	0.8605	0.9287	0.8063	0.9816	0.8755	0.8293	0.9293	0.7789
DeepLabV3+_GE	0.9645	0.7529	0.6238	0.9548	0.6058	0.9558	0.6427	0.4979	0.9159	0.4756	0.9558	0.6374	0.4944	0.9256	0.4766
DeepLabV3+_GE&S2	0.9811	0.8821	0.8247	0.9529	0.7918	0.9803	0.8672	0.8078	0.9400	0.7671	0.9778	0.8419	0.7639	0.9427	0.7290
DFDKFNet	**0.9849**	**0.9090**	0.8633	0.9607	**0.8332**	**0.9846**	**0.9004**	0.8608	0.9451	**0.8189**	**0.9836**	**0.8892**	0.8396	0.9465	**0.8005**
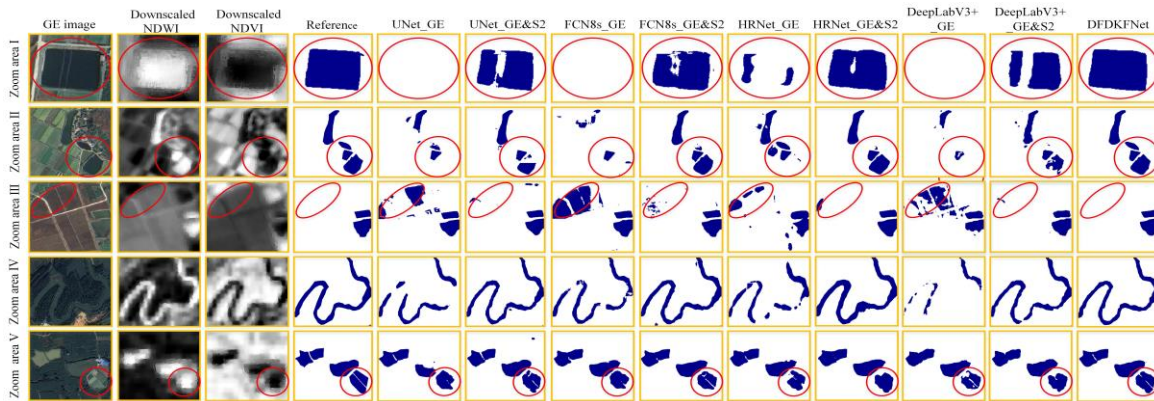
Fig. 4. Visual comparisons of different methods in five zoom areas. In the water body maps, dark blue indicates water and white indicates land.

bodies with domain knowledge from the S2 NDWI and NDVI images, and is superior to the state-of-the-art DL methods that simply concatenate different input data in the fusion. Results show that DFDKFNet can not only enhance the mapping accuracy for water bodies than the comparison DL methods, but it can also improve the spatial detail in water mapping. Further research can focus on incorporating RGB-based indices to enrich the spectral information of the input data, and using synthetic aperture radar images and gap-filling methods [21] to reduce the impact of clouds that may exist in the S2 image in water mapping based on DFDKFNet.

REFERENCES

[1] J.-F. Pekel, A. Cottam, N. Gorelick, and A. S. Belward, "High-resolution mapping of global surface water and its long-term changes," *Nature,* vol. 540, no. 7633, pp. 418-422, Dec. 2016.

[2] X. Li *et al.*, "Monitoring high spatiotemporal water dynamics by fusing MODIS, Landsat, water occurrence data and DEM," *Remote Sens. Environ.,* vol. 265, p. 112680, Nov. 2021.

[3] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell. ,* vol. 39, pp. 640–651, 2014.

[4] Y. Ren, X. Li, X. Yang, and H. Xu, "Development of a dual-attention U-Net model for sea ice and open water classification on SAR Images," *IEEE Geosci. Remote Sens. Lett.,* vol. 19, pp. 1-5, 2022.

[5] P. Qin, Y. Cai, and X. Wang, "Small waterbody extraction with improved U-Net using Zhuhai-1 hyperspectral remote sensing images," *IEEE Geosci. Remote Sens. Lett.,* vol. 19, pp. 1-5, 2022.

[6] Chen *et al.* (2017). "Rethinking atrous convolution for semantic image segmentation." [Online]. Available: https://arxiv.org/abs/1706.05587

[7] J. Wang *et al.*, "Deep high-resolution representation learning for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 43, no. 10, pp. 3349-3364, Oct. 2021.

[8] W. Feng, H. Sui, W. Huang, C. Xu, and K. An, "Water body extraction from very high-resolution remote sensing imagery using deep U-Net and a superpixel-based conditional random field model," *IEEE Geosci. Remote Sens. Lett.,* vol. 16, no. 4, pp. 618-622, April. 2019.

[9] Y. Li, B. Dang, Y. Zhang, and Z. Du, "Water body classification from high-resolution optical remote sensing imagery: Achievements and perspectives," *ISPRS J. Photogramm.,* vol. 187, pp. 306-327, May. 2022.

[10] W. Li, R. Dong, H. Fu, J. Wang, L. Yu, and P. Gong, "Integrating Google Earth imagery with Landsat data to improve 30-m resolution land cover mapping," *Remote Sens. Environ.,* vol. 237, Feb. 2020.

[11] Z. Zhang, M. Lu, S. Ji, H. Yu, and C. Nie, "Rich CNN features for water-body segmentation from very high resolution aerial and satellite imagery," *Remote Sens.,* vol. 13, no. 10, May. 2021.

[12] Q. Hu *et al.*, "Exploring the use of Google Earth imagery and object-based methods in land use/cover mapping," *Remote Sens.,* vol. 5, no. 11, Nov. 2013.

[13] C. Shen, "A transdisciplinary review of deep learning research and its relevance for water resources scientists," *Water Resour. Res.,* vol. 54, no. 11, pp. 8558-8593, Aug. 2018.

[14] K. Yuan, X. Zhuang, G. Schaefer, J. Feng, L. Guan, and H. Fang, "Deep-learning-based multispectral satellite image segmentation for water body detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.,* vol. 14, pp. 7422-7434, 2021.

[15] S. K. Ahmad, F. Hossain, H. Eldardiry, and T. M. Pavelsky, "A fusion approach for water area classification using visible, near infrared and synthetic aperture radar for South Asian conditions," *IEEE Trans. Geosci. Remote Sens.,* vol. 58, no. 4, pp. 2471-2480, 2020.

[16] Benjamin, MatvL, midaha, PGibson, RMcKinlay, and W. Kan. (2016). "Dstl satellite imagery feature detection." [Online]. Available: https://kaggle.com/competitions/dstl-satellite-imagery-feature-detection

[17] Y. Li, Y. Zhou, Y. Zhang, L. Zhong, J. Wang, and J. Chen, "DKDFN: Domain Knowledge-Guided deep collaborative fusion network for multimodal unitemporal remote sensing land cover classification," *ISPRS J. Photogramm.,* vol. 186, pp. 170-189, April. 2022.

[18] W. Cui *et al.*, "Knowledge and spatial pyramid distance-based gated graph attention network for remote sensing semantic segmentation," *Remote Sens.,* vol. 13, no. 7, Mar. 2021.

[19] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," *Opt. Lett.,* vol. 33, no. 2, pp. 156-158, Jan. 2008.

[20] X. Li *et al.*, "SFSDAF: An enhanced FSDAF that incorporates sub-pixel class fraction change information for spatio-temporal image fusion," *Remote Sens. Environ.,* vol. 237, p. 111537, Feb. 2020.

[21] Q. Wang, L. Wang, X. Zhu, Y. Ge, X. Tong, and P. M. Atkinson, "Remote sensing image gap filling based on spatial-spectral random forests," *Science of Remote Sensing,* vol. 5, p. 100048, 2022.