proceedings

MDPI

Proceeding Paper

Unsupervised Domain Adaptive Transfer Learning for Urban Built-Up Area Extraction †

Feifei Peng [1], Shuai Yao [2], Yixiang Chen [1,*] and Wenmei Li [1]

[1] School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210003, China; 1022173204@njupt.edu.cn (F.P.); liwm@njupt.edu.cn (W.L.)
[2] Huawei Technologies Co., Ltd., Nanjing Branch, Nanjing 210008, China; yaoshuai14@huawei.com
* Correspondence: chenyixiang@njupt.edu.cn
† Presented at the 31st International Conference on Geoinformatics, Toronto, ON, Canada, 14–16 August 2024.

Abstract: Built-up areas are the main gathering place for human activities. The widespread availability of various satellite sensors provides a rich data source for mapping built-up areas. Deep learning can automatically learn multi-level features of targets from sample data in an end-to-end manner, overcoming the limitations of traditional methods based on handcrafted features. However, existing deep-learning-based methods rely on the quantity and distribution of sample data, and the trained models often exhibit limited generalization ability when faced with image data from novel scenarios. To effectively tackle this issue, this study proposes an unsupervised domain adaptive transfer learning method based on adversarial machine learning. This method aims to utilize the feature information of the source domain to train a classifier suitable for target domain feature discrimination without requiring a target domain label, and achieve built-up area extraction of different sensor images. The model comprises a feature extraction module, a label classification module, and a domain discrimination module. Through adversarial training, the feature knowledge from the source domain is transferred to the target domain, achieving feature alignment and efficient discrimination of built-up areas. The Gaofen-2 (GF-2) and Sentinel-2 datasets were employed for experimental evaluation. The results show that the proposed method, trained on the GF-2 image dataset (source domain), can be transferred unsupervised to the Sentinel-2 image dataset (target domain), demonstrating robust detection performance. Further comparative experiments have also demonstrated the superiority of our method in extracting built-up areas through transfer learning.

Keywords: high-resolution sensors; built-up area; block-based processing; transfer learning

Citation: Peng, F.; Yao, S.; Chen, Y.; Li, W. Unsupervised Domain Adaptive Transfer Learning for Urban Built-Up Area Extraction. *Proceedings* **2024**, *110*, 10. https://doi.org/10.3390/proceedings2024110010

Academic Editors: Dongmei Chen, Yuhong He and Songnian Li

Published: 3 December 2024

1. Introduction

The spatial distribution information of urban built-up areas is pivotal for urban planning, land use analysis, and environmental monitoring. Satellite remote sensing, with its macroscopic and periodic observation capabilities, offers invaluable data sources for dynamic monitoring of built-up areas. Consequently, the extraction and mapping of built-up areas from satellite imagery have emerged as a topic of significant interest. Recent advancements in deep learning have markedly enhanced the delineation of urban built-up areas, excelling in learning intricate patterns from voluminous datasets and achieving fine-grained detection of complex spatial patterns in urban built-up areas, significantly improving the accuracy in built-up area extraction [1,2].

In the endeavor to extract built-up areas, tasks often necessitate cross-image and cross-regional operations. Yet, within the same geographic area, the stylistic characteristics of built-up areas can vary significantly when imaged by different sensors. While deep learning methods have improved accuracy, constructing separate label datasets for each image from various sensors or regions is laborious and expensive [3,4]. Therefore, transfer learning has gained prominence [5], as it facilitates the transfer of knowledge from one domain to another, expanding the applicability of the trained network [6].

A widely attractive technique of transfer learning is domain adaptation, which can be achieved through data transformation to harmonize styles across different domains [7,8]. To overcome the limitations of early methods, an unsupervised domain adaptation technique based on adversarial training has been developed. This technique narrows the gap between source and target domains by computing and minimizing their statistical distance. Common strategies involve employing Maximum Mean Discrepancy (MMD) [9] or CORAL LOSS [10] as loss functions, or implementing domain obfuscation through adversarial-based domain discriminators [11,12].

In this study, we incorporate unsupervised domain adaptation technology to achieve migration extraction of built-up areas from satellite images of different sensors or regions. Our objective is to eliminate reliance on target domain labels, enabling efficient knowledge transfer between source and target domains at minimal cost, and improving the model's adaptability to target domain data distributions. This innovation paves the way for the automated, high-precision identification of urban built-up areas.

2. Methodology

2.1. Model Architecture

The proposed domain adaptive model, outlined in Figure 1, employs adversarial training to align source and target domain features, enhancing the detection of built-up areas in target images. It consists of three core modules: feature extraction, label classification, and domain discrimination. The feature extraction module leverages our previously proposed LMSFF-CNN model [13], which excels in extracting multi-scale fused features from both the source and target domain images. The extracted features are subsequently channeled into dual parallel branches. One branch is dedicated to label classification, tasked with discerning whether the image segment represents a built-up or non-built-up area. Concurrently, the other branch focuses on domain discrimination, identifying the origin of the features as either belonging to the source or target domain. Throughout the training phase, the efficacy of domain alignment is gauged by a domain discriminator in conjunction with a loss function, which quantifies the divergence between the feature distributions of the two domains.

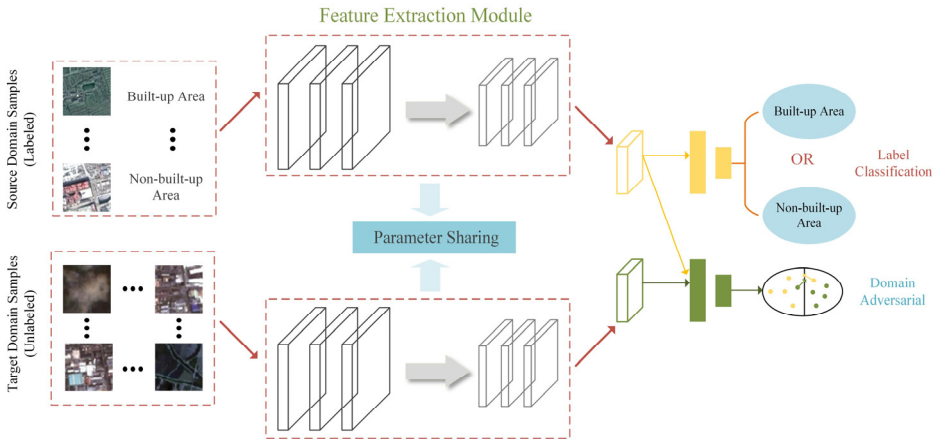

Figure 1. Proposed framework.

2.1.1. Feature Extraction Module

Feature extraction is pivotal for facilitating domain adaptive transfer learning. To attain effective knowledge transfer, it is imperative to extract comparable features—such as spectrum and texture—from both the source and target domains to enable feature alignment and domain confusion. Extracting single-level features can exacerbate the disparity between domains, hindering alignment efforts. Consequently, this study employs the LMSFF-CNN model for the feature extraction module, capable of fusing multi-scale features, thereby preparing them for subsequent processing stages aimed at feature alignment.

2.1.2. Label Classification Module

The label classification module constitutes a pivotal component dedicated to discriminating between the labels of features produced by the feature extraction module. Its primary function is to ascertain whether the image represents a built-up area or a non-built-up area. Initially, a global average pooling layer is utilized to transform input features into a one-dimensional feature, thereby diminishing the parameter count while computing output features across each dimension of the feature map. Subsequently, a dropout layer is appended to the global average pooling layer to mitigate overfitting by probabilistically neglecting certain features—a strategy particularly advantageous when confronted with scanty target domain data. Ultimately, a classification layer is constructed utilizing fully connected layers coupled with a sigmoid activation function to convert input features into output predictions, effectively discriminating between built-up and non-built-up areas.

2.1.3. Domain Discrimination Module

The main task of the domain discrimination module is to distinguish whether the input features stem from the source domain or the target domain. This module adopts a configuration akin to the label classification module, with the critical distinction residing in the utilization of a full connection layer and a softmax function in the terminal layer. Through cleverly designed training methods, the domain discrimination module can achieve feature alignment and domain obfuscation, ensuring a more consistent representation of features from both source and target domains. This approach can significantly improve the model's robustness and adaptability, enhancing its performance on new target domains. Adding the domain discrimination module helps the model adjust to different data distributions, leading to better results in practical applications.

2.2. Training and Prediction

The annotated dataset within the source domain is represented as $\left\{\left(x_i^s, y_i^s\right)\right\}_{i=1}^{N_s}$, where x_i^s denotes the input data from the source domain and y_i^s represents the corresponding labels. The unannotated dataset in the target domain is designated as $\left\{x_j^t\right\}_{j=1}^{N_t}$, where x_j^t refers to the input data from the target domain. Throughout the training phase, the feature extraction module M_F, the label classification module M_L, and domain discrimination module M_D are concurrently optimized. The objective is to harmonize the characteristics of the source and target domains and diminish their disparities through the minimization of the loss function. The loss function governing the training process is delineated as follows:

$$\begin{aligned} \mathrm{L} &= L_l + \lambda L_d \\ &= \Sigma_{i=1}^{N_s} L_l(M_L(M_F(x_i^s)), y_i^s) \\ &+ \lambda \Sigma_{i=1/j=1}^{N_s/N_t} L_d(M_D(M_F(x_{i/j}^{s/t})), 1/0) \end{aligned} \tag{1}$$

In Formula (1), N_s and N_t denote the sample quantities of the source and target domains, respectively. L_l signifies the binary cross-entropy loss function of the label classification module M_L. This loss function measures the classification error based on the source domain labels. L_d refers to the categorical cross-entropy loss function used in the domain discriminator module M_D, which uses the loss function to assess its ability to distinguish between samples from the source and target domains:

$$L_l = -\frac{1}{N_s} \sum_{i=1}^{N_s} \{y_i^s log M_L(M_F(x_i^s)) + (1 - y_i^s) log(1 - M_L(M_F(x_i^s)))\} \tag{2}$$

$$\begin{aligned} L_d &= L_d(D_s, 1) + L_d(D_t, 0) \\ &= -\tfrac{1}{N_s} \sum_{i=1}^{N_s} (y_s^{'} = 1) log(M_D(M_F(x_i^s))) \\ &+ (-\tfrac{1}{N_t} \sum_{j=1}^{N_t} (y_t^{'} = 0) log(M_D(M_F(x_j^t)))) \end{aligned} \tag{3}$$

During the model training phase, the initial focus is on training the label classification module utilizing the source domain data. Subsequently, the training process for the domain discrimination module is divided into two steps. Initially, domain labels are attributed to both the target and source domain datasets, followed by freezing the label classification module while the feature extraction module is trained. In the subsequent step, the domain labels assigned in the preliminary stage are inverted; the label classification module remains frozen as the feature extraction module undergoes further training. By employing adversarial domain training, the model facilitates feature alignment, effectively mitigating the disparity between the domains and enhancing its proficiency in executing cross-domain tasks.

3. Experiments and Analysis

3.1. Study Area and Dataset

To verify the effectiveness of the proposed method, this study selected three typical cities in southeastern China, namely, Shenzhen, Fu'an, and Fuqing. Shenzhen (113°43′~114°38′ E, 22°24′~22°52′ N) is located in the southern part of Guangdong Province, with dense urban construction. The terrain features are mainly plains and hills, and the building styles are diverse and dense. Fu'an (119°38′~119°59′ E, 26°41′~27°24′ N) is located in the northeast of Fujian Province, with diverse terrain. The built-up areas are mainly concentrated in plains and river valleys, and there is a large amount of agricultural land and forest land around it. Fuqing (119°23′~119°49′ E, 25°18′~25°52′ N) is located along the southern coast of Fujian Province, with a terrain dominated by plains and hills. The city is densely built, with both industry and agriculture coexisting, exhibiting typical urban—rural integration characteristics.

Our dataset consists of two domains, the source domain and the target domain. The source domain data comes from GF-2 satellite imagery in Shenzhen, with a spatial resolution of 1 m. We randomly selected 300 annotated sample images covering as diverse scenes as possible, including 150 images of built-up areas and 150 images of non-built-up areas. The target domain data comes from Sentinel-2 images with a resolution of 10 m, and 300 unlabeled images were selected as the training data for the target domain. To better evaluate the performance of the model, in addition to the training data in the target domain, we annotated 70 images as the test set for the target domain.

3.2. Experimental Setup

The experiment was conducted on a high-performance computing platform, with the following specific configuration: a Dell workstation (manufactured by Dell Inc., Round Rock, TX, USA) equipped with an Intel Xeon E5-2620 v3 CPU, 32 GB of memory, and NVIDIA Quadro K620 graphics card, ensuring computing speed and data processing capabilities. The software environment includes the Windows operating system, Python 3.6 programming language, integrated deep learning frameworks TensorFlow 1.4, and Keras 2.3.1.

In terms of experimental parameter settings, the learning rate is initialized to 0.001 to promote rapid convergence of the model, the training epochs are set to 50, and the sample size for each batch is set to 10 to balance training speed and model update frequency, achieving efficient iterative optimization.

3.3. Performance Comparison of Different Methods

To evaluate the performance of the proposed method, we compared it with four alternative approaches: direct prediction using only the source domain dataset (Baseline1), a fine-tuned transfer learning method (Baseline2), and domain adaptive methods utilizing CORAL-LOSS (Baseline3) and MMD-LOSS (Baseline4) for feature alignment, respectively. The results are shown in Table 1. The direct prediction strategy, relying exclusively on source domain data to forecast target domain images, yielded unsatisfactory results. For example, the F1-Score for Shenzhen1 plummeted to a mere 0.2827. This indicates that

although the source domain and target domain data come from the same region, the data characteristics of different sensors are quite different. The fine-tuned transfer learning approach exhibited enhanced performance yet fell short of achieving the ideal precision.

Table 1. Accuracy Evaluation Results of Different Transfer Learning Methods.

Study Area	Method	P	R	F1-Score	IoU
Fu'an	Baseline1	0.5926	0.6162	0.6042	0.4329
	Baseline2	0.5921	0.8802	0.7080	0.5480
	Baseline3	0.9207	0.5180	0.6630	0.4959
	Baseline4	0.8274	0.6447	0.7247	0.5683
	Proposed	0.8712	0.8109	**0.8399**	**0.7241**
Fuqing	Baseline1	0.6042	0.8090	0.6917	0.5288
	Baseline2	0.5898	0.9143	0.7171	0.5590
	Baseline3	0.8743	0.7064	0.7814	0.6413
	Baseline4	0.8283	0.7399	0.7816	0.6415
	Proposed	0.7677	0.9249	**0.8390**	**0.7227**
Shenzhen 1	Baseline1	0.8511	0.1570	0.2652	0.1528
	Baseline2	0.8813	0.4891	0.6291	0.4589
	Baseline3	0.8929	0.7600	0.8211	0.6965
	Baseline4	0.9042	0.9138	0.9090	0.8332
	Proposed	0.9412	0.8930	**0.9165**	**0.8458**
Shenzhen 2	Baseline1	0.6437	0.7517	0.6935	0.5309
	Baseline2	0.5997	0.9609	0.7385	0.5855
	Baseline3	0.7189	0.9374	0.8137	0.6859
	Baseline4	0.7773	0.9570	0.8578	0.7511
	Proposed	0.8690	0.8978	**0.8832**	**0.7908**

By contrast, the proposed domain adaptive method showcased distinct advantages across the majority of the test images. Remarkably, the F1-Score for a specific image within Shenzhen soared to 0.9193, underscoring its robust consistency. However, the utilization of CORAL-LOSS and MMD-LOSS for feature alignment engendered results that fluctuated considerably, lacking the requisite stability. Figure 2 illustrates the extraction results of built-up areas using different methods. The proposed method achieves more complete and accurate extraction, highlighting its effectiveness.

Overall, the proposed domain adaptation method performed the best in experiments and has significant advantages over traditional transfer learning methods, especially in the absence of annotated samples, enhancing its robustness and applicability.

3.4. Generalized Application of the Model

To further validate the feasibility of the proposed domain adaptive method for extracting built-up areas in a large area, we selected Sentinel-2 satellite image data covering three major cities, Shenzhen, Zhuhai, and Xiamen, for experimental evaluation. As depicted in Figures 3–5, the spatial distribution of built-up areas within these municipalities is heterogeneous, and the building types exhibit considerable variability. Despite these challenges, the resultant maps for each region accurately displayed the location and extent of built-up areas, thereby demonstrating the effectiveness of the proposed method.

Although the delineation of urban areas is predominantly precise, discrepancies are observed in non-built-up areas due to significant disparities between source and target domain characteristics. This observation underscores the necessity for refining the domain adaptive techniques to mitigate such inaccuracies.

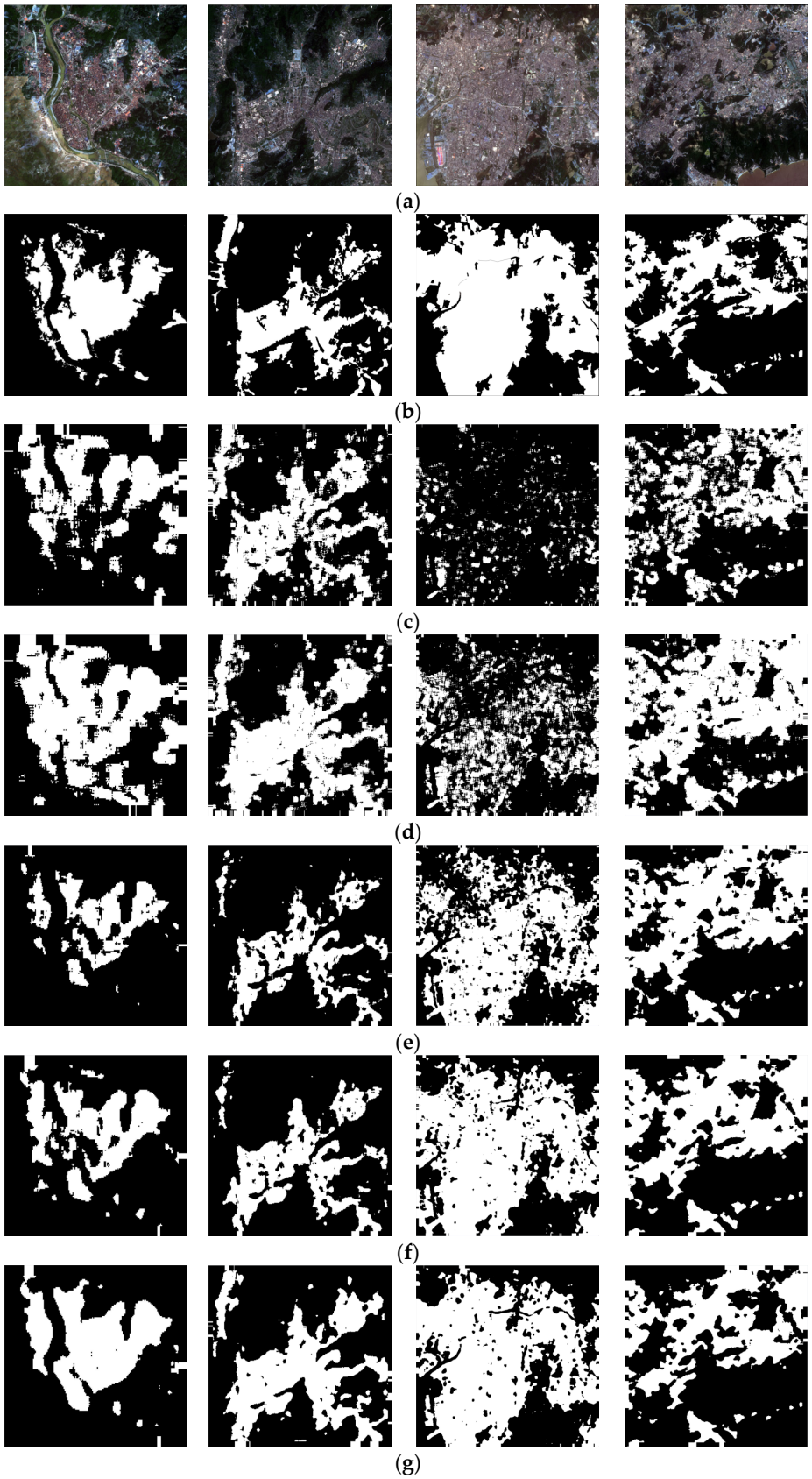

Figure 2. Extraction results of different methods: (**a**) The test images from left to right are Fu'an, Fuqing, Shenzhen1, and Shenzhen2; (**b**) Ground truths; (**c**) Baseline1; (**d**) Baseline2; (**e**) Baseline3; (**f**) Baseline4; (**g**) Proposed method.

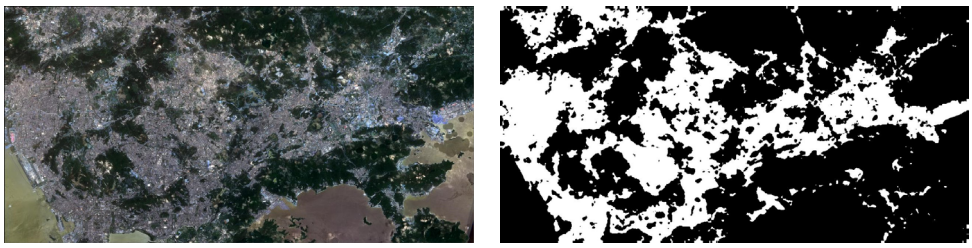

Figure 3. Sentinel-2 image and extraction results in Shenzhen City.

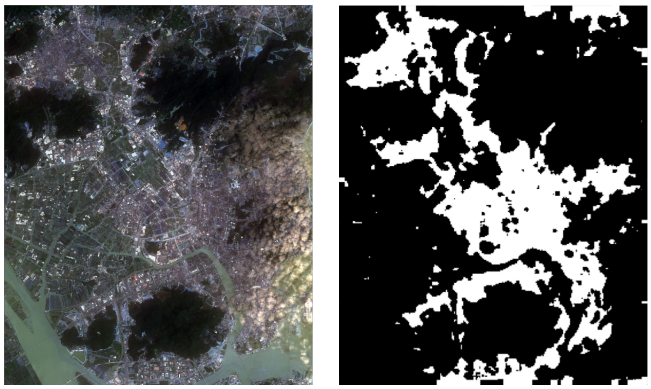

Figure 4. Sentinel-2 image and extraction results in Zhuhai City.

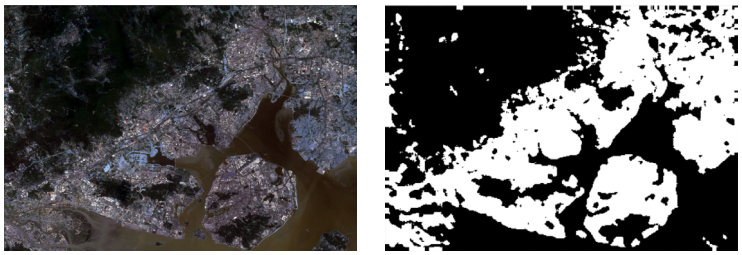

Figure 5. Sentinel-2 image and extraction results in Xiamen City.

4. Conclusions

For the task of transfer learning in remote sensing images from different sensors, this study introduces an unsupervised domain adaptive method. The method proves effective in addressing the complexities associated with identifying built-up areas across varying images and geographical locales without necessitating a lot of annotation work from the target domain, thereby markedly reducing the labor cost. A comparative evaluation against four prevailing transfer learning methods reveals that the proposed technique excels in terms of both stability and accuracy, especially in image prediction of large-scale areas such as Shenzhen and Zhuhai. This showcases the method's strong generalization ability. Prospective research endeavors will concentrate on enhancing this technique and broadening its application in remote sensing analyses to facilitate urban planning and environmental management.

Author Contributions: F.P., S.Y. and Y.C. conceived and designed the model. F.P. performed the experiments. S.Y. analyzed the data. F.P. wrote the paper. Y.C. and W.L. reviewed and revised the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Postgraduate Research and Practice Innovation Program of Jiangsu Province (Grant No. KYCX23_1070).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data presented in this study are not publicly available due to privacy restrictions related to sensitive geographic information.

Acknowledgments: We appreciate the helpful comments of the reviewers.

Conflicts of Interest: The authors declare that they have no conflicts of interest. Author Yao Shuai is currently employed by Huawei Technologies Co., Ltd. However, this research was conducted entirely prior to their employment at Huawei Technologies and is not associated with their current position. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Chen, Y.; Yao, S.; Hu, Z.; Huang, B.; Miao, L.; Zhang, J. Built-Up Area Extraction Combing Densely Connected Dual-Attention Network and Multiscale Context. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 5128–5143. [CrossRef]
2. Chen, Y.; Cheng, H.; Yao, S.; Hu, Z. Building Extraction from High-Resolution Remote Sensing Imagery Based on Multi-Scale Feature Fusion and Enhancement. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *XLIII-B3-2022*, 55–60. [CrossRef]
3. Iqbal, J.; Ali, M. Weakly-supervised domain adaptation for built-up region segmentation in aerial and satellite imagery. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 263–275. [CrossRef]
4. Chen, Y.; Dang, X.; Zhu, D.; Huang, Y.; Qin, K. Urban functional zone mapping by coupling domain knowledge graphs and high-resolution satellite images. *Trans. GIS* **2024**, *28*, 1510–1535. [CrossRef]
5. Liu, W.; Qin, R. A MultiKernel Domain Adaptation Method for Unsupervised Transfer Learning on Cross-Source and Cross-Region Remote Sensing Data Classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 4279–4289. [CrossRef]
6. Zhang, W.; Zhu, Y.; Fu, Q. Semi-Supervised Deep Transfer Learning-Based on Adversarial Feature Learning for Label Limited SAR Target Recognition. *IEEE Access* **2019**, *7*, 152412–152420. [CrossRef]
7. Benjdira, B.; Bazi, Y.; Koubaa, A.; Ouni, K. Unsupervised Domain Adaptation Using Generative Adversarial Networks for Semantic Segmentation of Aerial Images. *Remote Sens.* **2019**, *11*, 1369. [CrossRef]
8. Tasar, O.; Happy, S.L.; Tarabalka, Y.; Alliez, P. ColorMapGAN: Unsupervised Domain Adaptation for Semantic Segmentation Using Color Mapping Generative Adversarial Networks. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7178–7193. [CrossRef]
9. Wang, J.; Jordan, M.I.; Long, M.; Cao, Y. Learning Transferable Features with Deep Adaptation Networks. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 6–11 July 2015.
10. Sun, B.; Saenko, K. Deep CORAL: Correlation Alignment for Deep Domain Adaptation. In Proceedings of the ECCV 2016, Amsterdam, The Netherlands, 8–10 and 15–16 October 2016; pp. 443–450.
11. Ganin, Y.; Lempitsky, V. Unsupervised Domain Adaptation by Backpropagation. *arXiv* **2014**, arXiv:1409.7495.
12. Tsai, Y.H.; Hung, W.C.; Schulter, S.; Sohn, K.; Chandraker, M.J.I. Learning to Adapt Structured Output Space for Semantic Segmentation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018. [CrossRef]
13. Chen, Y.; Peng, F.; Yao, S.; Xie, Y. Lightweight Multilevel Feature-Fusion Network for Built-Up Area Mapping from Gaofen-2 Satellite Images. *Remote Sens.* **2024**, *16*, 716. [CrossRef]