RESEARCH ARTICLE

Urban surface classification using semi-supervised domain adaptive deep learning models and its application in urban environment studies

Xiaotian Ding[1,2,3] · Yifan Fan[1,2,3] · Yuguo Li[4] · Jian Ge[1,3]

Abstract

High-resolution urban surface information, e.g., the fraction of impervious/pervious surface, is pivotal in studies of local thermal/wind environments and air pollution. In this study, we introduced and validated a domain adaptive land cover classification model, to automatically classify Google Earth images into pixel-based land cover maps. By combining domain adaptation (DA) and semi-supervised learning (SSL) techniques, our model demonstrates its effectiveness even when trained with a limited dataset derived from Gaofen2 (GF2) satellite images. The model's overall accuracy on the translated GF2 dataset improved significantly from 19.5% to 75.2%, and on the Google Earth image dataset from 23.1% to 61.5%. The overall accuracy is 2.9% and 3.4% higher than when using only DA. Furthermore, with this model, we derived land cover maps and investigated the impact of land surface composition on the local meteorological parameters and air pollutant concentrations in the three most developed urban agglomerations in China, i.e., Beijing, Shanghai and the Great Bay Area (GBA). Our correlation analysis reveals that air temperature exhibits a strong positive correlation with neighboring artificial impervious surfaces, with Pearson correlation coefficients higher than 0.6 in all areas except during the spring in the GBA. However, the correlation between air pollutants and land surface composition is notably weaker and more variable. The primary contribution of this paper is to provide an efficient method for urban land cover extraction which will be of great value for assessing the urban surface composition, quantifying the impact of land use/land cover, and facilitating the development of informed policies.

Keywords Urban surface recognition · Deep learning · Domain adaptation · Semi-supervised learning · Urban environment · Air quality

✉ Yifan Fan
  yifanfan@zju.edu.cn

[1] College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, China

[2] Center for Balance Architecture, Zhejiang University, Hangzhou, China

[3] International Research Center for Green Building and Low-Carbon City, International Campus, Zhejiang University, Haining, China

[4] Department of Mechanical Engineering, The University of Hong Kong, Hong Kong SAR, China

Introduction

Different land cover types have a dramatic and distinct impact on the urban environment, including local climate and air quality because of their physical properties or as indicators of human activity (Fan et al. 2019). For instance, deforestation–afforestation can alter the local climate through strong modifications on evapotranspiration and urbanization directly affects local air pollution due to the increasing number of emission sources (Zhao et al. 2020a). High spatial resolution urban land cover information is required for investigating the influence of urban surfaces as well as other urban climate studies, transportation, and urban planning applications (Oke et al. 2017; Sha et al. 2018; Xu et al. 2019), urban ventilation and pollutant dispersion modeling (Yang et al. 2020; Zhao et al. 2020b), and carbon emission estimations (Zhang et al. 2014; Gao et al. 2014).

Previous studies (Du et al. 2016; Zheng et al. 2018; Xu et al. 2020) have aimed to investigate the influence of urban indicators on neighborhood-scale urban climate. These studies often relied on one or two case studies due to the scarcity of high-resolution land cover data, which typically have limited spatial and temporal coverage. Acquiring urban land cover data is challenging due to the intricate composition of urban surfaces, comprising various elements like buildings, roads constructed from different materials, and diverse impervious surfaces (Fan et al. 2021). While these data can be manually labeled or mapped, this process is labor-intensive and time-consuming. In recent years, deep learning methods, such as convolutional neural networks (CNNs), have recently been demonstrated to be an effective automated tool for complex urban surface classification at the pixel level (Volpi and Tuia 2017; Chen et al. 2020; Fan et al. 2021). It provides a promising tool for large-scale land cover extraction which will be of great value for urban environment studies at a larger scale, including the city, regional, or even global scale. However, deep learning methods are inherently data-driven and require a substantial volume of training examples, i.e., images, and corresponding manually labeled or annotated ground truth (Li et al. 2020). An additional challenge lies in achieving domain generalization, as models must acquire generalized and precise feature representations from a limited number of training data originating from the source domain to effectively perform on unseen target domains (Xu et al. 2022). The primary issue arises from data discrepancies between the training samples (source domain) and the target images (target domain), where domains often refer to different datasets or data distributions. It is common for the imagery used to have a different acquisition time, atmospheric condition, or even origination (from different satellite sensors), resulting in discrepancies between the source and target domains (Fan et al. 2021). Poor domain generalization ability has become a major hindrance to obtaining high-resolution land cover data at a larger regional, national, or even global scale.

Studies have examined the domain generalization ability and the reliance on training samples of deep learning methods (Hoffman et al. 2016; Scott et al. 2017; Li et al. 2019b). Transfer learning has emerged as a valuable tool for enhancing model generalization and has found application in image mapping tasks (Pan and Yang 2010; Fu et al. 2023b, a; Liu et al. 2023). A specific subset of transfer learning is domain adaptation (DA), which is specifically used to enhance the models trained on one domain (source domain) to perform well on a different domain (target domain) (Chen et al. 2018; Toldo et al. 2020; Wilson and Cook 2020; Peng et al. 2022). Transferring the images of different domains into a similar appearance or style is one of the effective ways to improve model generalization ability (Xu et al. 2022), such as the Cycle Generative Adversarial Network (Toldo

et al. 2020). To enhance the performance, pretrained semantic model can be used as a supervision model to encourage the model to maintain consistent semantic representations during translation (Hoffman et al. 2018). While DA models have been used in high-resolution image classification (Ji et al. 2021; Luo and Ji 2022), the primary focus has been on improving image translation performance. However, the translated images may still exhibit disparities with the target image, and this error can persist and propagate throughout subsequent training process if used as new training data. To address this problem, we employed the semi-supervised learning (SSL) (van Engelen and Hoos 2020; Jing and Tian 2021) technique, which can leverage the labeled data to guide the model's learning process, while also benefiting from the vast amount of unlabeled data. SSL can help reduce the influence of errors resulting from DA by extracting additional information from unlabeled data in the target domain, while also serving as an effective means to mitigate overfitting. Therefore, combining above-mentioned approaches will yield better results than using any single method alone (Zhang et al. 2022; Aryal and Neupane 2023).

In this study, our objective is to propose a domain adaptive land cover classification model for extracting land cover from urban surface imagery from different domains. The model combines the DA and SSL techniques and thus can be further exploited with existing annotated datasets and applied to different images from Google Earth platform [Google Inc., California, USA]. Then, the urban surface information, specifically the surface land cover types in urban areas, around weather stations and air quality monitoring stations was extracted from Google Earth images. This information was then employed to study the correlation between the urban surface types and the environmental parameters, i.e., air temperature and pollutant concentration.

The rest of the paper is structured as follows: "Data preparation" section presents the preparation of the source training dataset and the target images within study areas used for the training and testing process, as well as the meteorological and air pollutant data for further correlation analysis. The proposed domain adaptive urban surface recognition methods are described in "Methods" section. "Results and discussion" section presents results and a discussion of the domain adaptive image classification model and correlation analysis with local environmental parameters. In the end, the conclusions are drawn in "Conclusions" section.

Data preparation

Source dataset

To initiate the procedure, the GF2 dataset is chosen as the source dataset (X_S, L_S), which was built in our previous
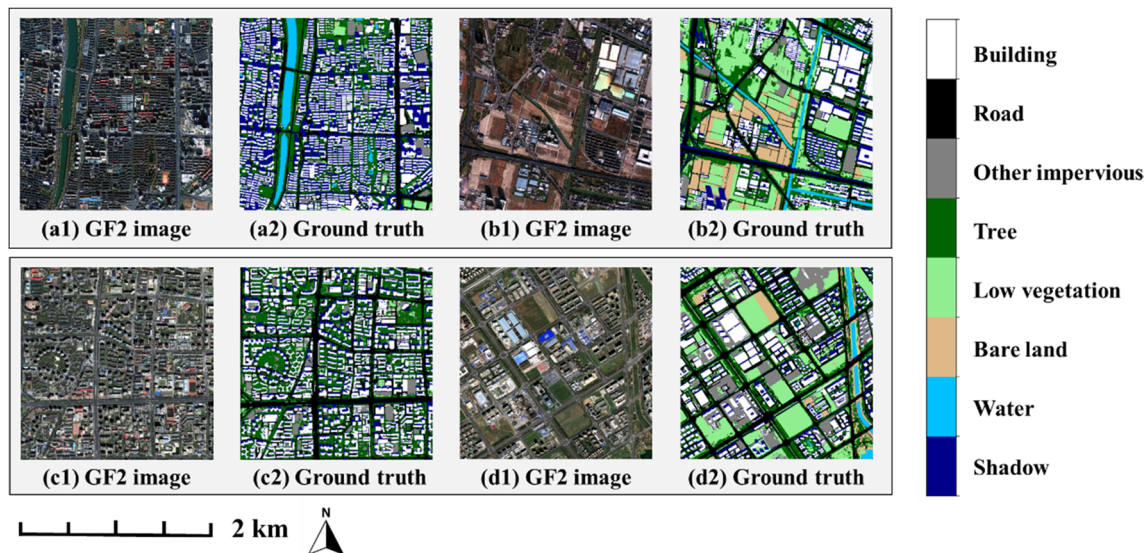
Fig. 1 Demonstration of the GF2 training dataset. Four sample images (2048 m×2048 m) from the source dataset were taken in Hangzhou (a1, b1) and Beijing (c1, d1). Corresponding ground truth (a2–d2)

study (Fan et al. 2021). The GF2 dataset is a high-quality 1-m resolution land cover dataset built with 41.9 km² of Gaofen2 (GF2) satellite imagery. Gaofen2 satellite is a sub-meter level optical Earth observation satellite, equipped with two high-resolution sensors (PMS) with a combined swath width of 45 km. The effective spatial resolutions for raw GF2 imagery are 4 m for multispectral bands (MSS) and 1 m for panchromatic band (PAN). The raw GF2 imagery has four spectral bands, including BLUE (450–520 nm), GREEN (520–590 nm), RED (630–690 nm), and Near-infrared (NIR, 770–890 nm). The Gram–Schmidt pan-sharpening method (Klonus and Ehlers 2009) was used to fuse MSS imagery (4 m) and PAN imagery (1 m) and produce 4-band multi-spectral images (RED, GREEN, BLUE, and NIR) with 1 m spatial resolution.

There are 10 sample images in the GF2 dataset covering both urban and sub-urban areas, and each sample image has a fixed size of approximately 4.2 km² (2048×2048 m), and four of them are illustrated in Fig. 1(a1–d1). Eight of the samples were taken in Hangzhou, China, while the remaining two images were from Beijing, China. In this study, Beijing and Hangzhou samples were used as the source training dataset to enhance generalization ability. All images were fully annotated with eight common land cover categories at pixel level using the proposed OSM-OBIA method (Fan et al. 2021). Considering the actual land cover behind the "Shadow" is not clear, the rest seven categories were grouped into two larger categories for further correlation analysis, i.e., the "Artificial" (including "Building", "Roads", and "Other impervious") and "Natural" (including "Tree", "Low vegetation", "Bare land", and "Water")

surface. The definition and fraction of each land cover category are listed in Table 5 in Appendix A.

Study area and target images

We investigated the impact of urban land cover in China's three most developed urban agglomerations: Beijing, Shanghai, and the Guangdong-Hong Kong-Macao Greater Bay Area (GBA). All those areas have an urban population of over 20 million (China Statistical Yearbook, 2015) (National Bureau of Statistics of China 2015). The urbanization rate in Beijing and Shanghai has exceeded 85%, including Guangzhou, Shenzhen in GBA. We collected land surface images from Google Earth around 33 meteorological monitoring sites and 107 air quality monitoring sites in those three regions, as listed in Table 1. The distribution of meteorological and air quality observation stations is shown in Fig. 9–10 in Appendix B.

To validate the model and obtain the urban surface information, the land surface images around meteorological and air quality stations are first downloaded from Google Earth

Table 1 Details of the environmental monitoring stations and observation data. (MT: meteorology, AQ: air quality)

Region name	Number of MT sites	Number of AQ sites	Time range of observation
Beijing	12	24	2021/03/21–2021/9/20
Shanghai	9	16	(Spring and Summer)
GBA	12	66	
Total	33	106	

Table 2 Detail information on the source and target images

Dataset name	Image origin	Acquisition time	Spatial resolution	Spectral bands
Source dataset	GF2 image	2019/12 (Hangzhou), 2019/08 (Beijing)	1 m	RGB
Target dataset	Google Earth	2017/01–2020/12	1 m	RGB

as target images, referred to as Google Earth images hereafter. The Google Earth is a computer program that maps the Earth by overlaying satellite images, aerial photographs, and other Geographic Information System (GIS) data. Consequently, it can be a repository of land surface images, typically acquired by diverse sensors under varying climate conditions and at different times. According to existing studies (Li et al. 2019a; Liu et al. 2021), a buffer zone of 2×2 km square area was selected for the urban surface recognition and correlation analysis related to environmental parameters. The corresponding meteorological or air quality monitoring sites are located in the center of the buffer zone. Thus, the target dataset consists of 556 km^2 of unlabeled Google images.

These selected images were chosen after the year 2017 to minimize the land surface difference in comparison to the source dataset. Then, images heavily obscured by cloud cover or mist were excluded by manual intervention. While the Google Earth platform can offer images with a spatial resolution as fine as 0.3 m, such high-resolution images were limited in availability. Therefore, images with a resolution of 0.6 m were chosen for our research., which were subsequently down-sampled to 1 m to be consistent with source images. The original GF2 imagery has four spectral bands, while images from Google Earth are RGB images that only contain three bands, i.e., RED, GREEN, and BLUE bands. For consistency, the NIR bands of the GF2 imagery in the training dataset were excluded. As a result, both source and target images are RGB with 1 m resolution. Detailed information of the source and target images is listed in Table 2. To further evaluate the model performance on Google Earth images, an additional number of images from Google Earth was obtained which correspond to the same ground truth with GF2 dataset, hereafter referred to as the Google dataset, as shown in Fig. 11 (a1–a3) in Appendix C. Misalignments exist between the Google Earth and the ground truth in GF2 dataset, and we have discussed the impact of these misalignments in detail in Appendix C.

Environmental data

To collect as many available samples as possible, meteorological and air quality data were collected over a two-season period in 2021 (Mar. 21, 2021, to Sep. 21, 2021) when a significant number of new air monitoring sites had been built and put into use. Meteorological data were collected from China National Meteorological Information Centre (http://data.cma.cn/). Mean values of air temperature (AT), relative humidity (RH), and Wind speed (WS) of the selected period were used for analysis in this study. Air quality monitoring data were obtained from the China National Environmental Monitoring Center (CNEMC, http://106.37.208.233:20035/). For air quality data, six pollutants ($PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, and O_3) were measured and recorded hourly according to China Environmental Protection Standards HJ 193–2013 and HJ 655–2013.[1] Detailed information about the data can be found in our previous study (Fan et al. 2020).

Methods

In this study, we initially combined the DA and SSL techniques to train a land cover classification model that can be applied on Google images, as shown in Fig. 2. Subsequently, the urban surface information around weather stations and air quality monitoring stations was extracted and was then used to study the correlation between the land surface types and the environmental parameters, e.g., air temperature.

Semi-supervised domain adaptive urban surface classification model

The overall workflow can be separated into two steps. In the first step, the source dataset (X_S, L_S) and target images (X_T) were used in the domain adaptation (DA) process to generate a translated GF2 dataset (X_S', L_S), as described in "Domain adaptation (DA)" section. The translated GF2 dataset is considered a training dataset built with target domain images. In the second step, the classification model is trained with the translated GF2 dataset (X_S', L_S) and the unlabeled target images (X_T) in semi-supervised learning (SSL) style, which is described in "Semi-supervised learning (SSL)" section. As a result, the trained classification model can be used in target Google Earth images.

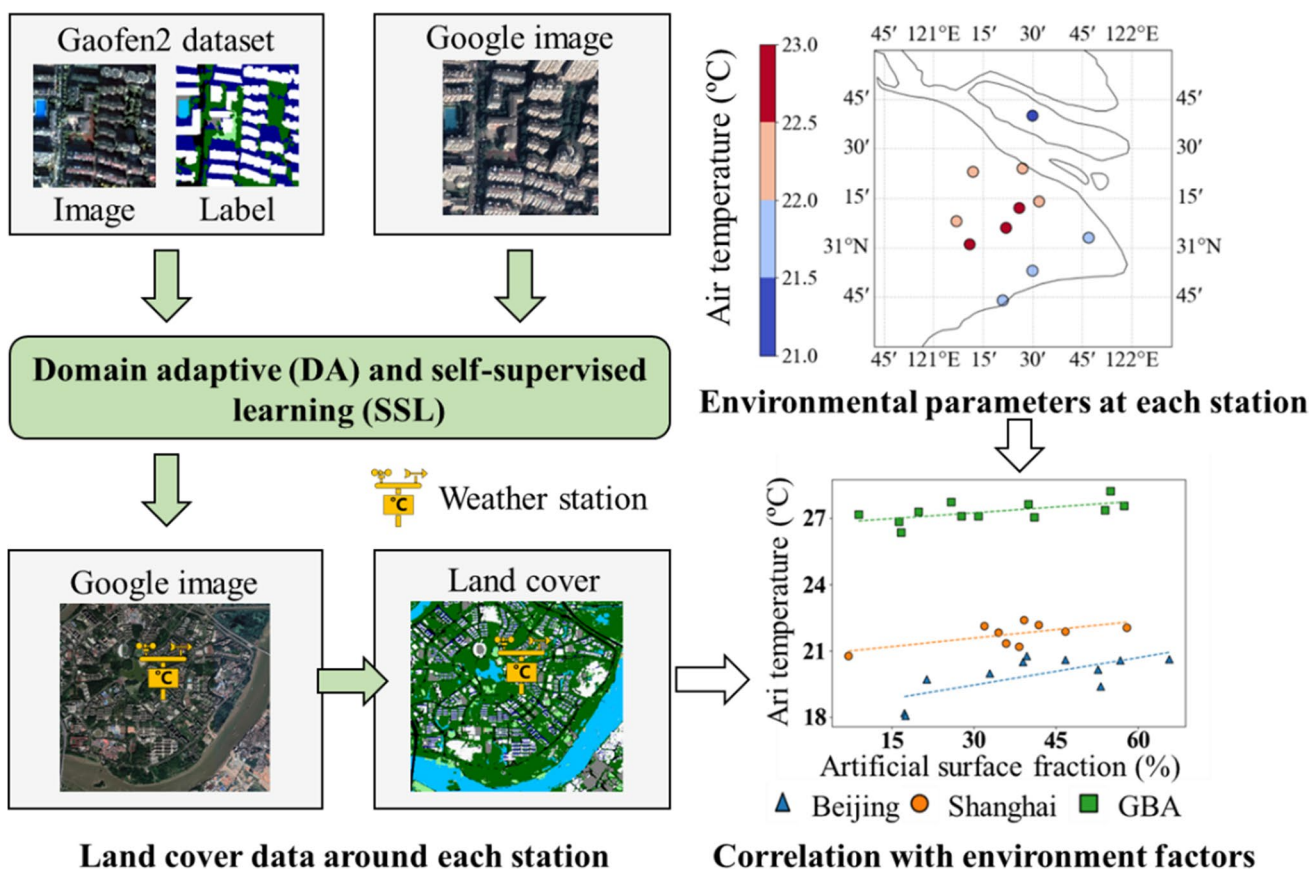[1] available at: http://english.mee.gov.cn/Resources/standards/Air_Environment/air_method/
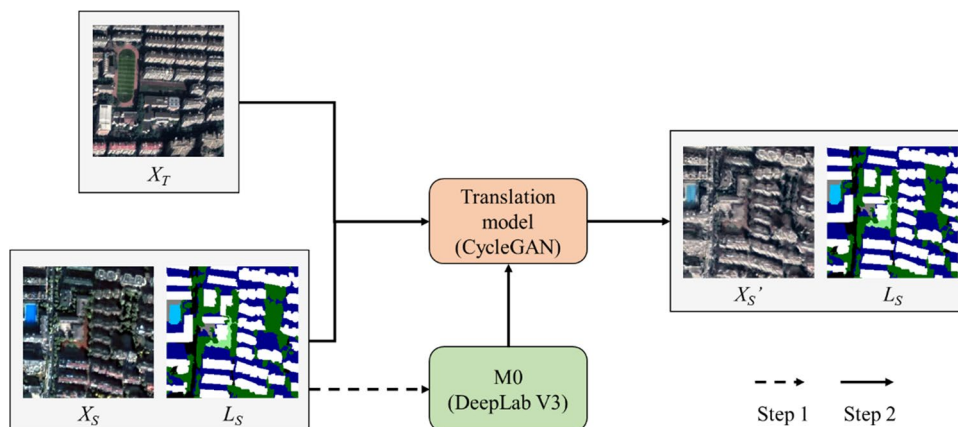
Fig. 2 Overview of the proposed workflow

Fig. 3 Illustration of the domain adaptation (DA) process

Domain adaptation (DA)

As shown in Fig. 3, the DA process contains two steps. The first step is to use the source dataset to train an initial classification model (*M0*). DeepLab V3 (Chen et al. 2017) was used for the training process, which has an encoder-decoder architecture (Ronneberger et al. 2015; Badrinarayanan et al. 2017). The ResNet101(He et al. 2016) was used as the backbone network for feature extraction as recommended in a previous study (Fan et al. 2021). The decoder part used to recover spatial details from the extracted feature map remains the same as in Li et al (2019b). In the second step, the image-to-image translation model, i.e., CycleGAN (Zhu et al. 2017), whose task is to translate the source image into the same style as the target image, is trained with the source dataset and the target data. Meanwhile, the training process

Fig. 4 Illustration of the semi-supervised learning (SSL) process
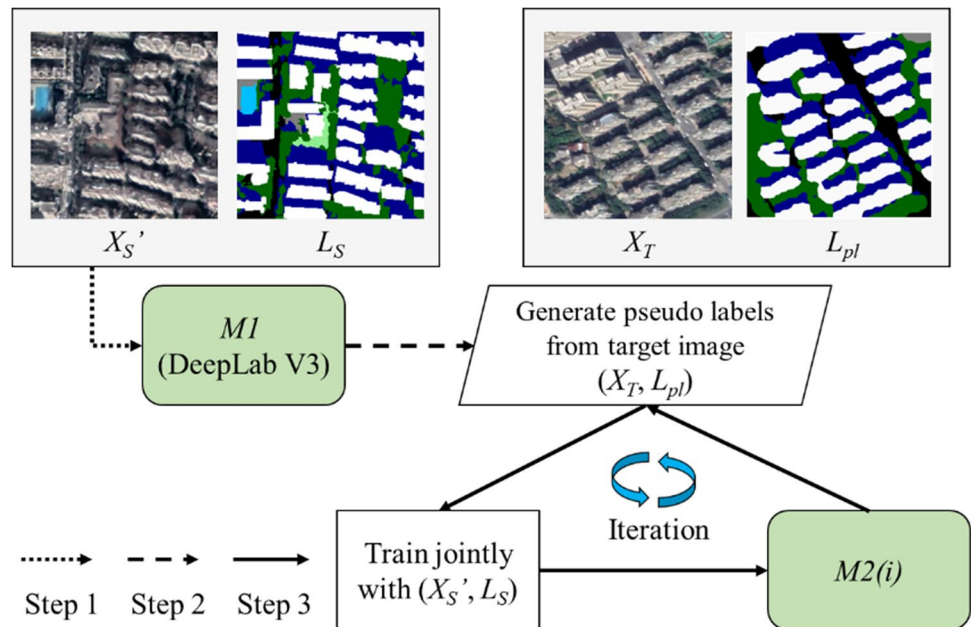
is conducted under the supervision of the initial model *M0* to improve the training performance (Hoffman et al. 2018). With the trained CycleGAN model, a translated GF2 dataset (X_S', L_S) is produced, which can be considered a dataset built from target images (X_T).

Semi-supervised learning (SSL)

After the domain adaptation, SSL was used to enhance the model performance on target images. A new classification model *M1* was trained with the translated GF2 dataset (X_S', L_S) (Fig. 4). Then, the *M1* model was applied to the target data (X_T) to generate segmentation results as pseudo labels. Only the pseudo labels (L_{pl}) with high quality or prediction confidence were selected for the next procedure. To obtain high-quality pseudo labels, the prediction confidence threshold is set as 0.9 (Li et al. 2019b). Following that, the translated GF2 dataset (X_S', L_S) and the target data with its corresponding pseudo labels (X_T, L_{pl}) were used to train the next generation of segmentation model *M2(i),* where *i* represents the i_{th} loop. The quality of pseudo labels (L_{pl}) can be improved and more high-quality pseudo labels can be obtained for the next training loop when a better *M2(i)* model is achieved. Therefore, the model with better performance is expected to be trained after each iteration. More detailed information for the models and configurations can be found in Appendix D.

Evaluation metric

In this study, we assess the model's performance in the target domain, specifically on both the translated GF2

dataset and the Google dataset. We employ two evaluation metrics: the *F1 score* and overall accuracy (*OA*), as defined in Eq. (1) and Eq. (2), respectively. The *F1 score* is a metric that combines *Precision* (Eq. 3) and *Recall* (Eq. 4) into a single value. *Precision* quantifies the proportion of accurately classified pixels within a specific category. *Recall*, conversely, gauges the accurate identification of pixels belonging to a particular category. Good classification performance is characterized by high scores in both *Precision* and *Recall*, making the *F1 score*, which represents the harmonic mean of *Precision* and *Recall*, a more comprehensive evaluation metric. Meanwhile, *OA* is computed by dividing the total count of correctly predicted pixels across all categories by the total number of pixels. In this study, we assessed per-category performance using the *F1 score* and overall performance using *OA*.

$$F1\ Score = 2 \times (Precision \times Recall)/(Precision + Recall) \tag{1}$$

$$OA = (TP + TN)/N_{all} \tag{2}$$

$$Precision = TP/(TP + FP) \tag{3}$$

$$Recall = TP/(TP + FN) \tag{4}$$

where True Positives (*TP*)/ True Negative (*TN*) are the number of pixels correctly classified as positive/positive; False Positives (*FP*)/ False Negative (*FN*) are the number of pixels wrongly classified as positive/positive; N_{all} is the total number of pixels.

Table 3 The overall accuracy of each trained model. *M0*: the model trained with the origin GF2 dataset. *M1*: the model trained with translated GF2 dataset. *M2(i)*: the model in the *i*th iteration of semi-supervised learning using translated GF2 dataset and the source dataset

Overall accuracy (OA)	GF2 dataset	Translated GF2 dataset	Google dataset
M0	**79.3%**	**19.5%**	**23.1%**
M1	**67.3%**	72.3%	58.1%
M2(1)	71.4%	74.2%	61.2%
M2(2)	72.0%	74.6%	**62.3%**
M2(3)	72.2%	**75.2%**	61.5%
M2(4)	72.0%	74.9%	61.4%

Note: The highest and lowest scores on each dataset are highlighted in bold

Correlation analysis

After the land cover information around each weather station and air pollutant monitoring station was derived, bivariate correlation analysis was conducted between the Artificial/Nature surface proportions and environmental data, i.e., meteorological and air quality data, over a two-season period (Mar. 21 to Sep. 20) in 2021. The Pearson's correlation coefficient ($-1 \leq r \leq 1$) is used to measure the linear correlation between land surface proportions and seasonal averaged environmental values at each site, which is calculated by Eq. (5). Following Dugord's work (Dugord et al. 2014), the significance of correlation was classified into three levels: strong correlation ($|r| \geq 0.6$), correlation ($0.3 \leq |r| < 0.6$) and weak correlation ($|r| < 0.3$). Bilateral t-test was conducted to evaluate the significance of the correlation coefficient.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (5)$$

where x_i, y_i denote value of the Artificial/Nature surface proportions and the environmental parameter at each monitoring station i, respectively; $\bar{x}$, $\bar{y}$ denote their average value within the city/region.

Results and discussion

Urban surface classification

The overall accuracy of the final classification model on each dataset is presented in Table 3. The performance of the initial classification model *M0*, which is trained on the original GF2 dataset (X_S, L_S) serves as the benchmark. Our objective is to achieve performance as close to the benchmark

as possible for the final model on the target domain, i.e., the translated GF2 dataset and Google dataset. The "*M1*" row represents the performance of the model trained on the translated GF2 dataset without using the SSL technique. The following rows show the performance of the model trained on the translated GF2 dataset at each iteration of SSL. The model's performance on each land cover category is assessed with the *F1 score*, as depicted in Fig. 5.

As listed in Table 3, the model *M0* achieved 79.3% of OA on the source GF2 dataset but failed on the translated GF2 and Google datasets (OA < 30%), which means that *M0* cannot be directly applied to land surface images from Google Earth. The model trained on the translated GF2 dataset could achieve 72.3% of OA using the domain adaptation technique. While using both DA and SSL techniques, the model performance on the translated GF2 dataset has a significant increase (72.3%–74.2%) at the first loop (Table 3). The improvement becomes smaller and reaches its highest OA (75.2%) on the translated GF2 dataset at the 3rd loop of SSL, i.e., *M2(3)*. This indicates that the quality of the extracted pseudo labels stops improving and the semi-supervised training process converges, which leaves limited space for further improvement (Li et al. 2019b) with more SSL loops. Besides, the achieved accuracy closely resembles the score attained by *M0*, indicating that the model is not suffering from underfitting. The final classification model [*M2(3)*] achieved a significant improvement on the Google dataset, as listed in Table 3, from 23.1% to 62.3%. This score is relatively lower compared with the one on the translated GF2 dataset (75.2%). This result may be attributed to the fact that the DA process might not entirely eliminate the domain differences. It could also be partially influenced by misalignments with ground truth due to changes in land cover and variations in camera angles, as illustrated in Fig. 12 in Appendix C.

The models exhibit performance imbalances on each land cover category, as shown in Fig. 5. The performance of model *M0* is relatively low for the "Low vegetation" and "Other impervious" categories, mainly due to the limited sample size of these two categories in the GF2 dataset (Table 5 in Appendix A) and high intra-class variances and interclass similarities (Liu et al. 2020). However, this issue is significantly mitigated while using *M2(3)*, as shown in Fig. 6, which has been trained additionally on substantial amount of unlabeled data. the *F1 score* on translated GF2 dataset are higher than that of *M1*, as well as *M0* on GF2 dataset. This indicates the SSL technique can alleviate performance imbalances across land cover categories resulting from uneven distribution of labeled training data (Yang and Xu 2020). While tested on Google dataset, the model experiences a general decline in performance across all categories, with the exception of "Water" and "Bare land". This phenomenon may be attributed to the fact that these two categories have smaller interclass differences compared to
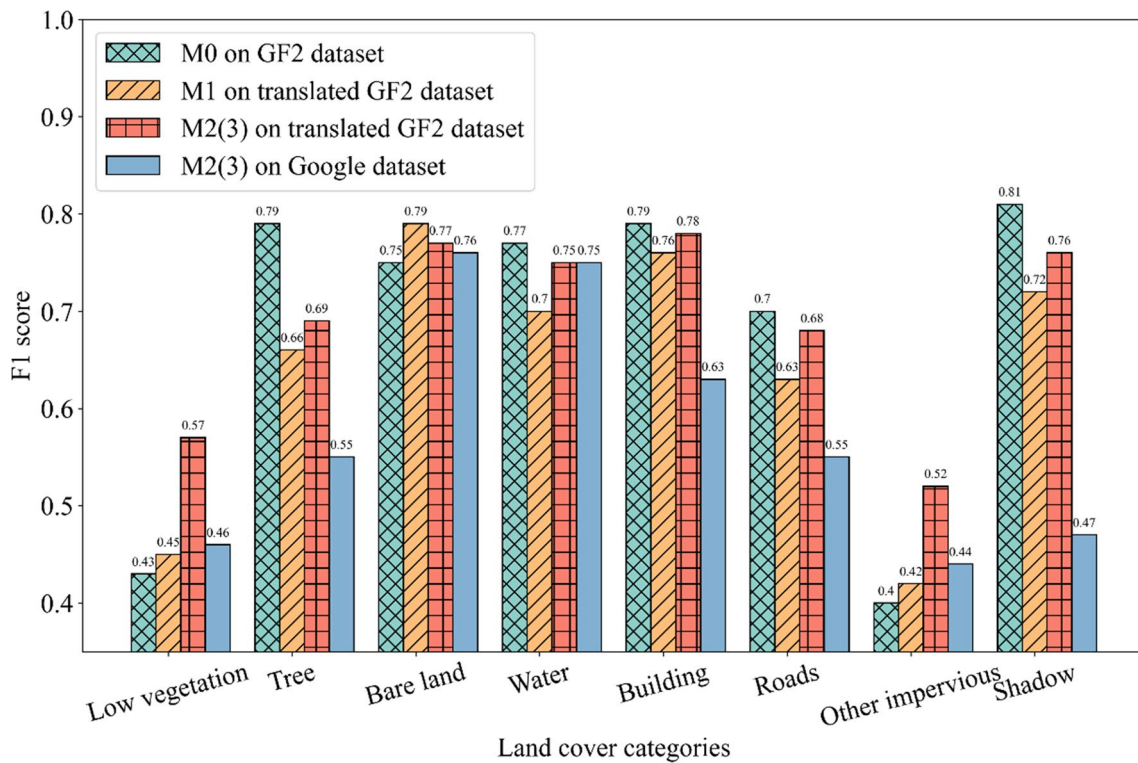
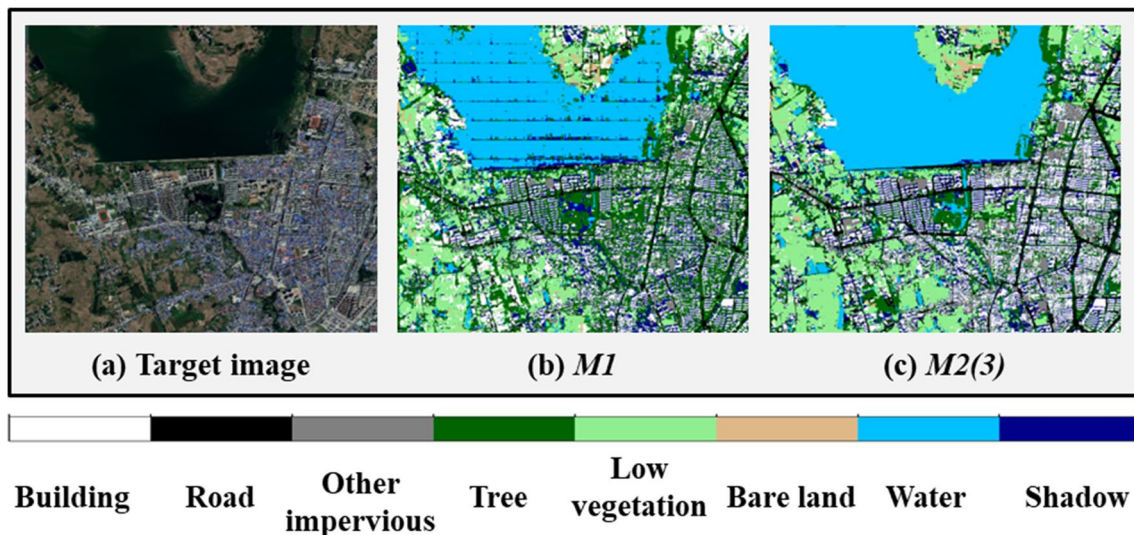**Fig. 5** Model's performance on each land cover category measured by *F1 score*

Fig. 6 The classification results on (**a**) the target image using (**b**) the *M1* model and (**c**) the *M2(3)* model

others, such as buildings and roads. Additionally, they are less susceptible to image misalignment caused by variations in view angles. The most significant performance difference is observed in the "Shadow" category, which is likely to be affected by climate condition and the solar angle at the time of photography.

The trained model has demonstrated performance improvement when tested on the translated GF2 dataset and the Google dataset. This method effectively enhances the model's generalizability across the two domains, making it suitable for application to satellite images exhibiting data discrepancies. However, the model's performance may
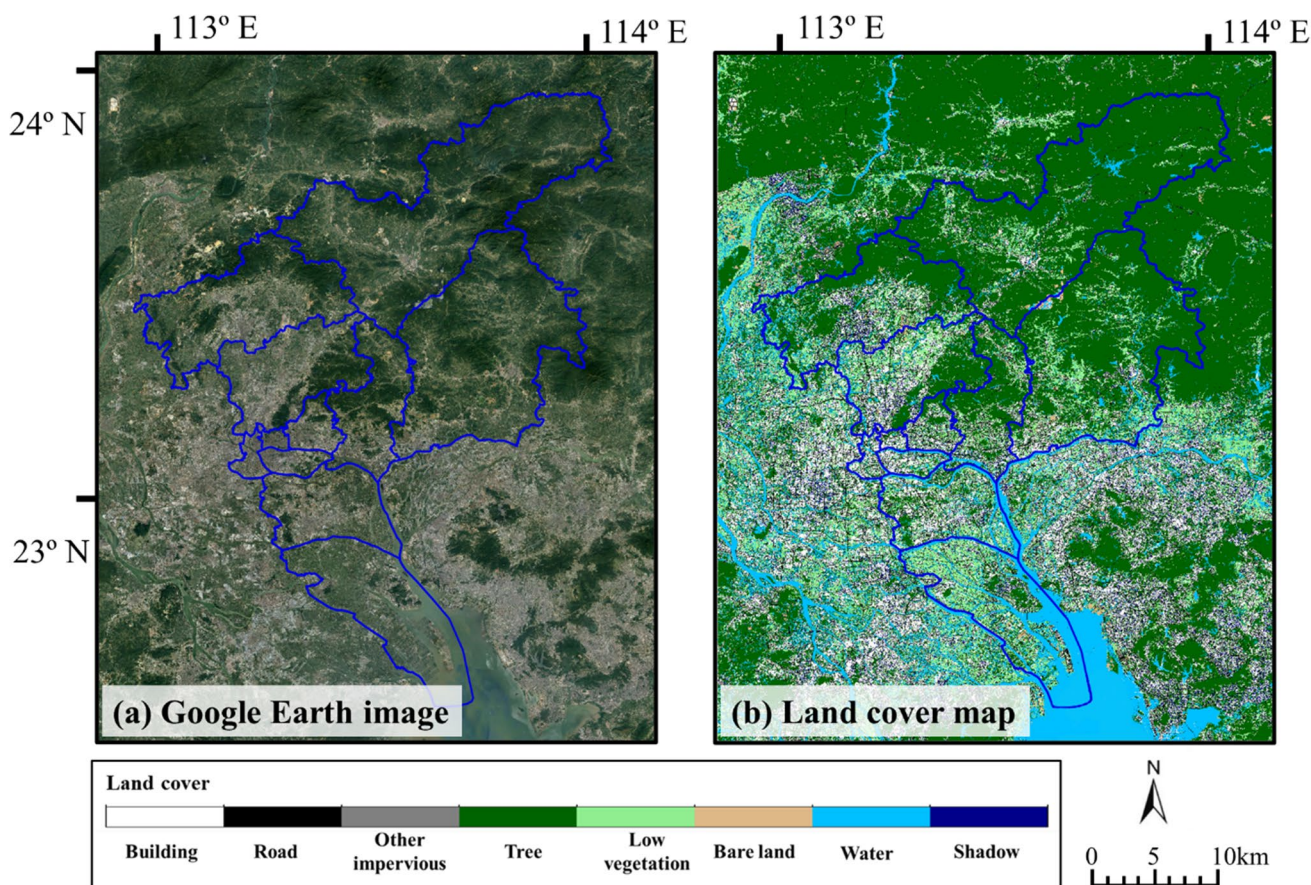
Fig. 7 The land cover map derived from Google Earth image. The blue solid line represents the administrative districts in Guangzhou, China

remain constrained by the limited diversity in the original training dataset, particularly concerning regional variations and spatial resolutions (Fan et al. 2021). The original GF2 dataset was built with satellite images captured in Chinese cities, where modern urban construction practices have led to similarities in building and road materials in recent decades. Consequently, the method may be applied to other similar cities such as Guangzhou, China, as the result shown in Fig. 7. But the model may not perform as well in areas characterized by diverse building characteristics, such as European regions with historic structures.

Correlation analysis

The classification model [*M2(3)*] derived the land cover information around each environmental monitoring station, including weather stations and air pollutant monitoring stations. The correlation analysis results are listed in Tables 4 and 7 in Appendix F. The correlations between the land cover fraction of a certain type and mean values of the weather data and air pollutant data are shown in

Fig. 8, 13 and 14 in Appendix F. The meteorological data in 2020 is also used for analysis and the result is shown in Table 6 in Appendix E.

Correlation of land cover with meteorological parameters

As shown in Fig. 8(a, d) and Table 4, there is a strong positive correlation (0.60 to 0.73) between artificial surfaces and air temperature in all three areas. All correlation coefficients are significant except for those in GBA. This result is in line with previous case studies (Yan and Dong 2015). Zhang et al. (2020) found 70% of the variance in daytime air temperature can be explained by building footprint ratio. Air temperature can be affected by changed physical land surface properties and anthropogenic heat emissions (Stewart and Oke 2012). A larger proportion of artificial surfaces means more natural land cover is reformed into buildings, roads, and other impervious surfaces. Therefore, the increased heat storage, anthropogenic heat emissions, and decreased water storage capacity result in a warmer environment with reduced relative humidity (Oke 1982). Those correlations are ubiquitous in all three areas and have small seasonal
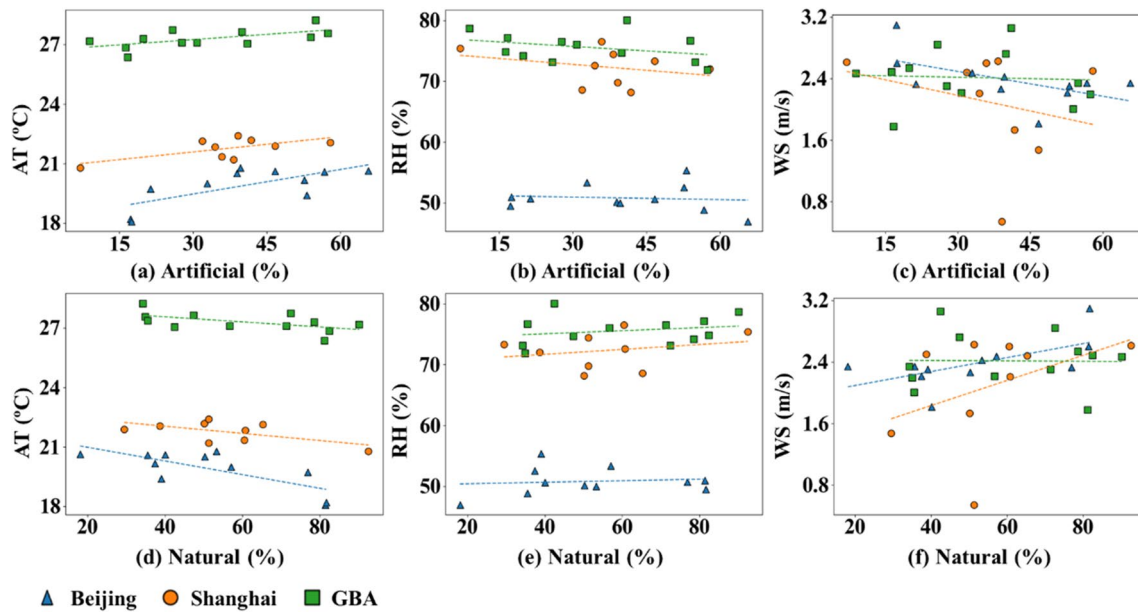
Fig. 8 Relationship between urban land cover and meteorological indicators (two-season average values) in three regions

Table 4 Pearson correlation coefficients between land cover and meteorological factors for each season in 2021. Bold values (with |r|≥ 0.3) indicate a correlation. (AT: air temperature, RH: relative humidity, WS: wind speed)

		Artificial			Natural		
	Season	Beijing	Shanghai	GBA	Beijing	Shanghai	GBA
AT	Spring (Mar. 21 to Jun. 20)	**0.71**[**]	**0.68**[**]	**0.66**	**−0.75**[**]	**−0.62**[**]	**−0.62**[**]
	Summer (Jun. 21 to Sep. 20)	**0.73**[**]	**0.66**[*]	**0.60**	**−0.75**[**]	**−0.60**[**]	**−0.41**[*]
	Overall (Mar. 21 to Sep. 20)	**0.72**[*]	**0.65**[**]	**0.60**	**−0.75**[**]	**−0.60**[**]	**−0.54**[**]
RH	Spring (Mar. 21 to Jun. 20)	−0.09	**−0.33**[**]	**−0.43**[**]	0.11	0.28[**]	**0.31**[**]
	Summer (Jun. 21 to Sep. 20)	**−0.45**[**]	**−0.34**[**]	**−0.33**[**]	**0.47**[**]	0.10[**]	−0.10[**]
	Overall (Mar. 21 to Sep. 20)	**−0.23**[*]	**−0.36**[**]	**−0.32**[**]	0.25	0.21[**]	0.20[**]
WS	Spring (Mar. 21 to Jun. 20)	**−0.60**[**]	−0.28[**]	−0.10[**]	**0.62**[**]	**0.43**[**]	0.03[**]
	Summer (Jun. 21 to Sep. 20)	**−0.40**[**]	−0.24[**]	0.07[**]	**0.44**[**]	**0.40**[**]	−0.15[**]
	Overall (Mar. 21 to Sep. 20)	**−0.53**[**]	−0.26[**]	−0.02[**]	**0.57**[**]	**0.41**[**]	−0.05[**]

[**]. Significant correlation at 0.01 level (bilateral, $p < 0.01$)

[*]. Significant correlation at 0.05 level (bilateral, $p < 0.05$)

variations, which are also supported with data from the year 2020, as shown in Table 6 in Appendix E.

There is a negative correlation (−0.32 to −0.43) between relative humidity and impervious land cover in Shanghai and GBA (Fig. 8(b, e) and Table 4). Lin et al. (2020) also found a negative correlation between urbanization and humidity in the urban areas of Guangdong, which is the so-called urban dry island effect (Lokoshchenko 2017). Given the definition of relative humidity, the urban–rural difference in RH is governed by thermal and moisture differences (Oke et al. 2017). The thermal difference between artificial and natural areas has a significant contribution to the RH difference. However, the emission of water vapor from industrial sources and transpiration may offset this phenomenon. This likely explains the weak correlation between artificial surface and RH in Beijing, except for summer (Table 4), This can also be witnessed in the analysis result in 2020 (Table 6 in Appendix E). Compared with Shanghai and GBA, Beijing has a drier climate with less rainfall, except for its rainy season (summer) (Liu et al. 2009). In dry seasons, the irrigation of urban green spaces, additional water vapor from human activities, and the evaporation of water bodies become the dominant contributor to air humidity in urban areas, causing a similar or even higher RH than rural areas, i.e., an urban wet island effect (Liu et al. 2021). In this scenario, the correlation between RH and the fraction of artificial surface can be weakened significantly.

There is a negative correlation (-0.40 to -0.60, Beijing) and a weak negative correlation (-0.24 to -0.28, Shanghai) between wind speed and artificial surfaces fraction, as high-rise buildings and their complex layouts reduce mean wind speed (Fan et al. 2019; Zhao et al. 2020b). This phenomenon has been intensively observed in previous studies (Du et al. 2016; Tao et al. 2018). This relationship is not obvious in GBA. The meteorology stations in GBA are more scattered than in the other two cities (Fig. 10 in Appendix B). The different background wind conditions at these stations could be the dominant influencing factor. For instance, the wind speed in an urban area near the coast may be higher than in rural areas inland due to the sea breezes.

Correlation of land cover with air pollutant concentrations

The results of correlation with air pollutant concentrations are detailed in Appendix F, with the corresponding discussion presented here. The correlations between most air pollutants and land cover are not as strong as those of meteorological indicators, as shown in Table 7 in Appendix F. One reason is that the emission or generation of each air pollutant is complex and can be affected by multiple factors ranging from the building scale, neighborhood scale, city-scale to regional scale (Tao et al. 2018; Fan et al. 2020). Most air pollutants are related to human activities, e.g., traffic, agriculture, and industry, and have different formation mechanisms (Fan et al. 2020). For instance, traffic is a major emission source for NO_2 (Lin and Cheng 2007) and CO (Hrebtov and Hanjalić 2019). The NO_2 concentration is also affected by industrial emission and photochemical reactions with VOCs (Liu et al. 2013). This all causes the heterogeneous distribution of air pollution within a region. Solely land cover information does not contain land use information, traffic density and the location emission source. For example, residential and industrial land use have distinct impacts on local air quality, but such information cannot be recognized by this model at present. Besides, the impervious land cover or the recognized road cannot reflect the traffic volume, a large proportion of road cover in the sub-urban industrial area may have small traffic flows. Therefore, the impact of human activity on air pollution may be more clearly revealed if more impact factors can be included in the analysis. Another limitation of this analysis is that all monitoring stations in Shanghai are located in areas with a similar impervious surface fraction, ranging from 25 to 55%. In contrast, the range for Beijing is 0% to 70%, and for the Great Bay Area (GBA), it is 5% to 80%. The absence of data points in rural and highly densely built areas may introduce deviations into the current results. Therefore, our future work will prioritize including a more diverse set of sample points.

Conclusions

In this study, we combined the domain adaptive (DA) and semi-supervised learning (SSL) techniques to achieve domain adaptive land cover classification. With a small labeled dataset built from Gaofen2 (GF2) satellite imagery, the trained model can be applied to images from Google Earth. The model was jointly trained with a significantly larger amount of unlabeled data. Compared to the result using only DA, this study utilizes the SSL technique to further improve the model's performance and mitigate performance imbalances across land cover categories that often arise from an uneven distribution of labeled training data in a small dataset. As a result, the model's performance on Google Earth images was improved significantly. The best performance was achieved after three SSL iterations, represented by the *M2(3)* model. With this model, the overall classification accuracy on the translated GF2 dataset was improved from 19.5% to 75.2%. Although the performance of the Google dataset is underestimated due to misalignments between image and ground truth, the overall accuracy can be improved from 21.3% to 62.3%.

The classification model [*M2(3)*] was used to derive land cover data for analyzing the relationship between urban surface information and environmental parameters in the three most developed cities/areas in China: Beijing, Shanghai, and GBA. The results provide indicate a strong positive correlation ($|r| \geq 0.6$) between air temperature and artificial land surface. The relative humidity is negatively related ($|r| \geq 0.3$) to the artificial land surface except for Beijing. The wind speed in Beijing is negatively related ($|r| \geq 0.3$) to the artificial land surface, but the correlation is weak in Shanghai and GBA. In terms of air pollutants, we found most correlations between air quality parameters and land cover are weaker than those among meteorological parameters, and show a regional difference. Detailed land use information, traffic volume, and location of emission source are needed for investigating the distribution of air pollutants in future studies.

This study makes contributions to both high-resolution land surface classification in complex urban areas and understanding the influence of land surface on urban climate. However, the model's performance remains constrained by the origin training dataset, which only contains training samples in Chinese cities. For future studies, integrating diverse sources of training samples into the scheme to build a more generalized model. The correlation analysis also provides useful insights for the parameterizing the impact of land cover in urban environment modelling. The high-resolution land cover information will be important data for assisting urban climate studies. Coupling high-resolution land cover data with other impact factors and more sophisticated urban climate models will be of great interest in the future.

Appendix A

Table 5 Detailed information of the eight land cover categories in the GF2 dataset

No	Category	Description	Fractions in the GF2 dataset
1	"Building"	Buildings, including residential, public, and industrial buildings, as well as buildings under construction	24.4%
2	"Road"	Traffic roads, including cement and asphalt road	15.0%
3	"Other impervious"	All impervious surfaces except building and road, e.g., parking lots, construction area, and sports fields	6.6%
4	"Tree"	Trees of all species with dense canopy, including forest and street trees	19.5%
5	"Low vegetation"	Grass and shrubs, agricultural areas with low crops or vegetables	6.8%
6	"Bare land"	Unpaved natural surfaces with bare land or soil with no vegetation	1.8%
7	"Water"	Open natural or man-made water bodies, including rivers, streams, ponds, lakes, and outdoor swimming pools	2.6%
8	"Shadow"	Areas shaded by shadows of tall objects like buildings and trees	23.3%

Appendix B. The locations of meteorological and air quality monitoring stations

The locations of meteorological and air quality monitoring stations in Beijing, Shanghai, and GBA are demonstrated in Fig. 9, and 10. The meteorological monitoring stations are national surface meteorological monitoring stations and location information is provided by China National Meteorological Information Centre. The air quality monitoring network is built by The China National Environmental Monitoring Center (CNEMC).
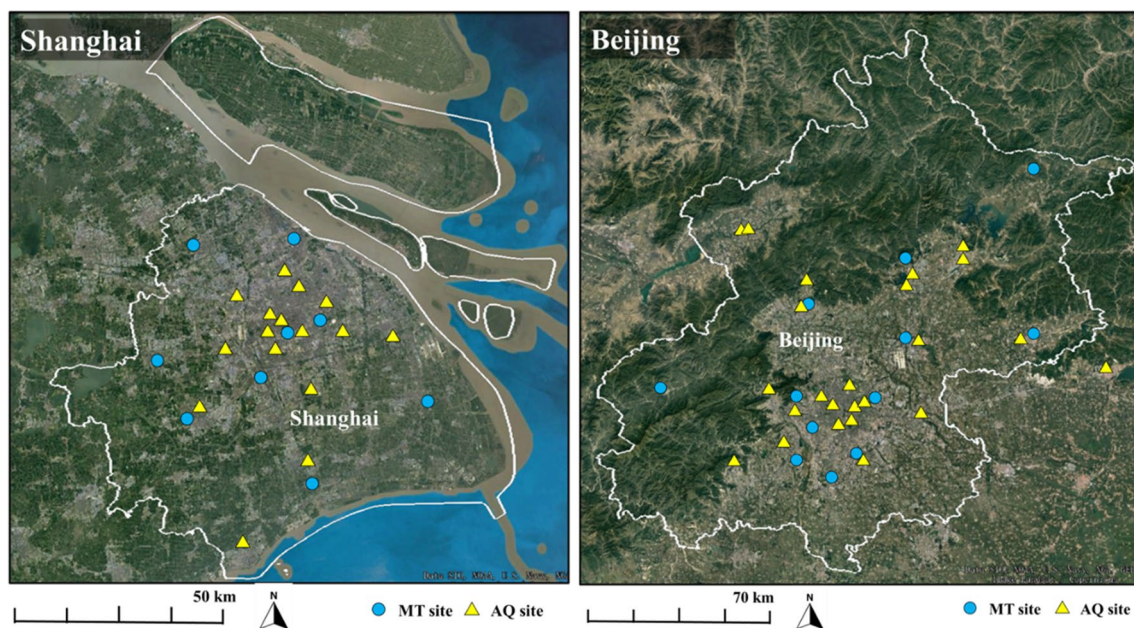

Fig. 9 The location of meteorological monitoring sites (9 MT sites) and air quality monitoring sites (16 AQ sites) in Shanghai, and 12 MT sites and 24 AQ sites in Beijing
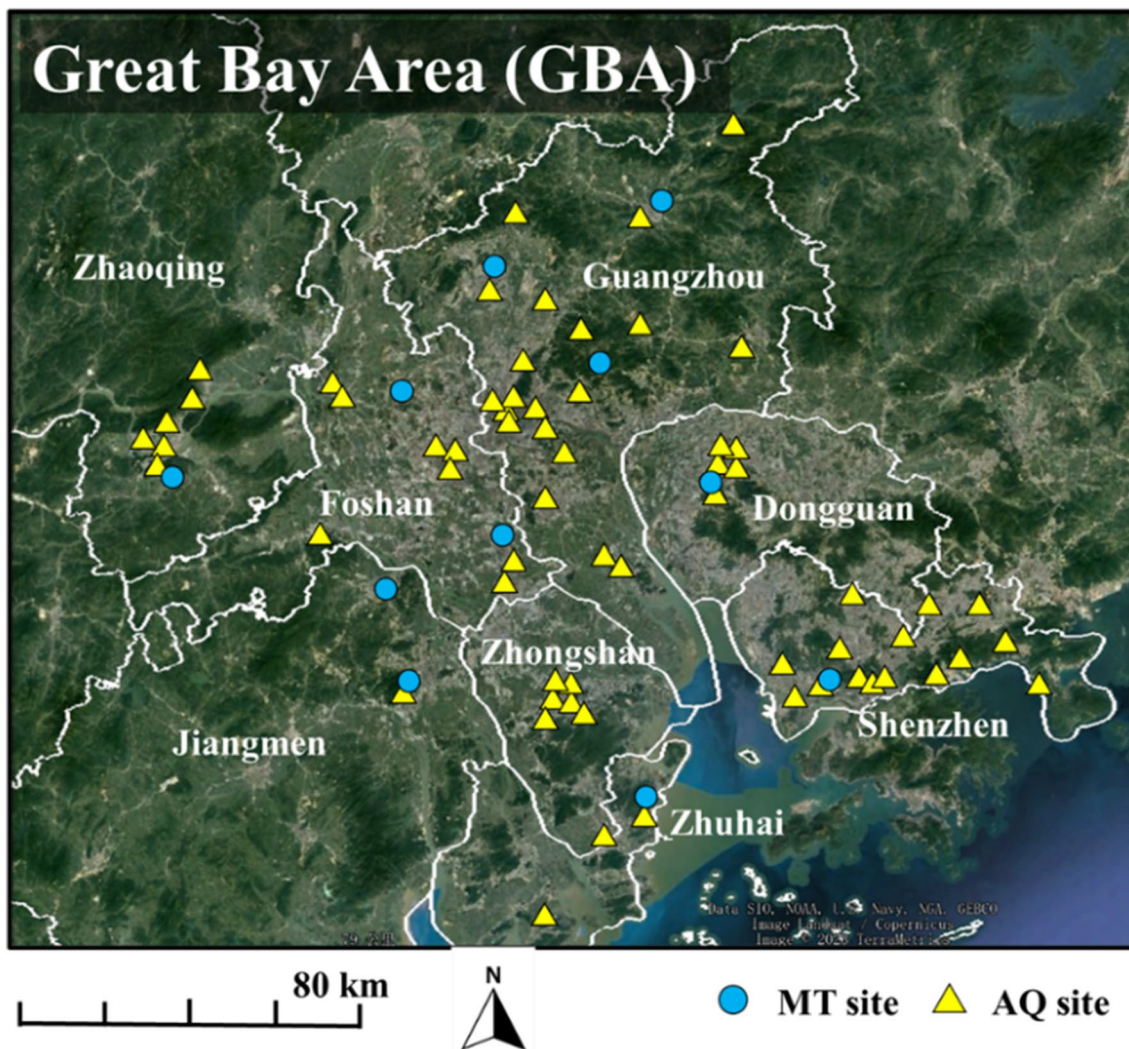
Fig. 10 The location of meteorological monitoring sites (12 MT sites) and air quality monitoring sites (66 AQ sites) in GBA

Appendix C. the impact of the misalignment in Google dataset

The Google dataset is built with image from Google Earth and ground truth from GF2 dataset, as shown in Fig. 11(a1–a3). The misalignments exist between the Google Earth and GF2 images mainly because they were obtained at different acquisition times and off-nadir angles. Therefore, after being clipped into smaller tiles, tiles with severe misalignment (Fig. 11(b1–b3)) caused by the large land cover change were excluded. As a result, the target dataset now consists of 31 km^2 of Google images with ground truth. Misalignments caused by different shading or off-nadir view angle are hardly evitable, as shown in the red frame in Fig. 11(c1–c3). Therefore, the classification accuracy evaluated on the target dataset will be underestimated and can only be used for reference.

With minor land cover change and camera angle difference the classification overall accuracy can reach a high level (Fig. 12 (a1–a4)). However, severely underestimated score will occur when there is a land cover change or camera angle difference, even if the model has correctly classified the images. As shown in Fig. 12(b1–b4) and Fig. 12(c1–c4), the underlying land cover has been well recognized, but the overall accuracy score is significantly low because of the misalignment between Google image and ground truth caused by land cover change and difference caused by off-nadir view angles. Although we have excluded image tiles with significant land cover changes, misalignment caused by different camera angles still exists in most test images. This is because buildings in the study areas were mostly high-rise buildings with more than seven floors. Therefore, the evaluation result on Google dataset is underestimated.
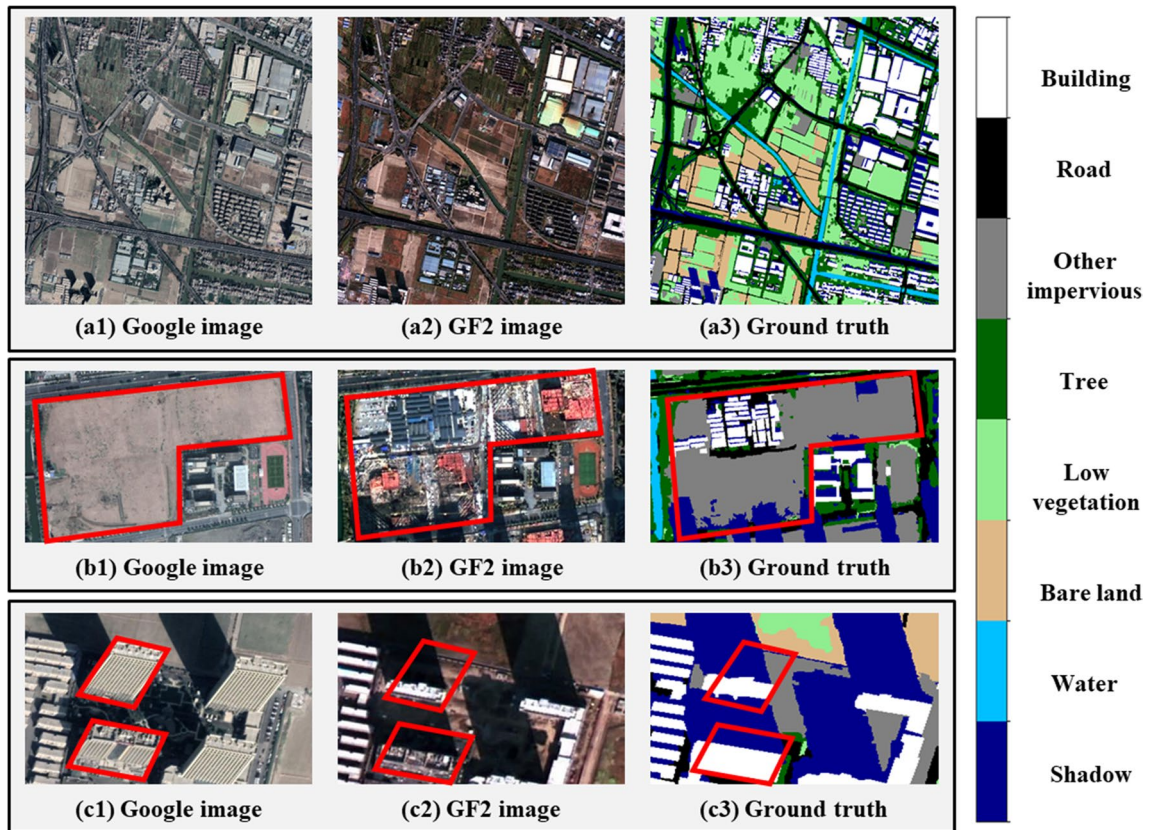
Fig. 11 Difference between the Google dataset and the source dataset. The differences are highlighted using red frames. (a1–a2) Google image and GF2 image of the same area. (b1–b2) Google image and GF2 image with large land cover changes. (c1–c2) Google image and GF2 image has large difference caused by different off-nadir view angle. (a3–c3) The ground truth corresponds to the GF2 image
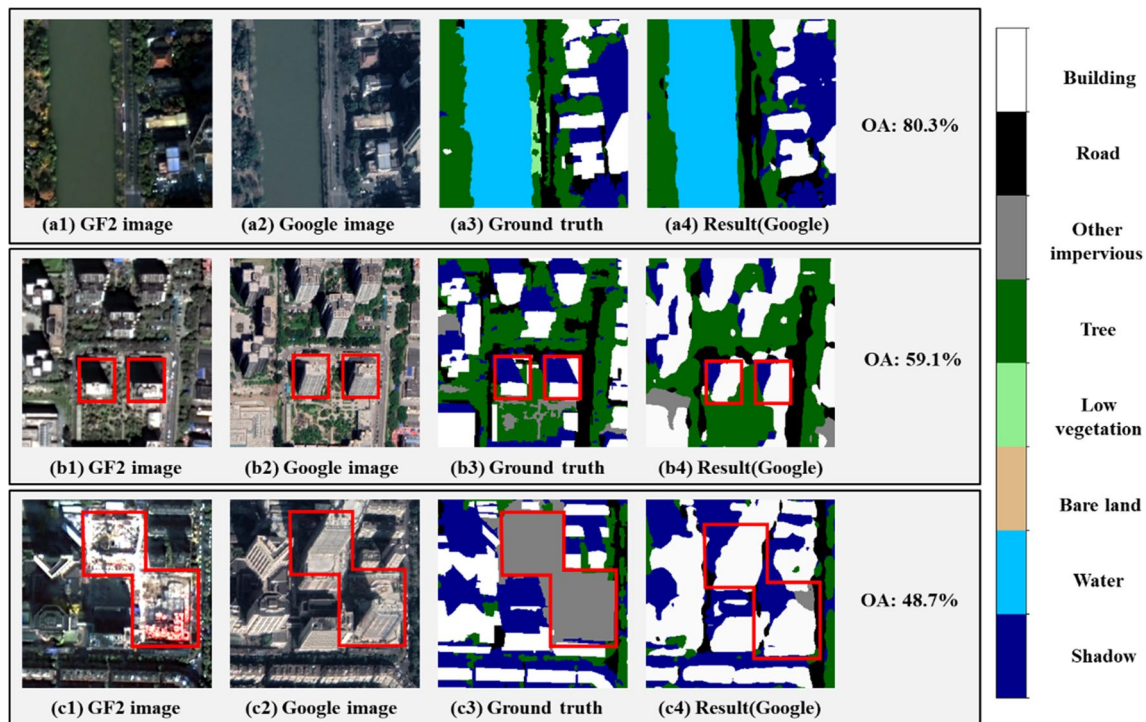
Fig. 12 The classification result on the Google image of the *M2(3)* model. Image differences are highlighted using red frames. (a1-a4) classification result on Google image with minor difference. (b1-b4) classification result on Google image with the difference caused by the off-nadir view angle. (c1-c4) classification result on Google image with large land cover change

Appendix D. Detailed settings for the training of deep learning models

The whole process was coded in the PyTorch (v1.10) (Mazza and Pagani 2020) framework. To reduce the dependence on the computer memory size, each training sample was furtherly divided into 64 patches [256m×256m for one patch and 8×8 patches (2048/256=8)]. 80% of the training dataset was used for training and the remaining 20% was reserved for performance evaluation. Data augmentation is a common strategy used in deep learning training (Wong et al. 2016; Diakogiannis et al. 2020), especially when the training dataset is not sufficient. In this study, training samples were augmented with the transform modules implemented in *torchvison*(Mazza and Pagani 2020) including RandomHorizontalFlip, RandomVerticalFlip, ColorJitter, RandomAdjustSharpness, and RandomAutocontrast. Following Li's work (Li et al. 2019b), the initial learning rate for training the classification model is 2.5×10^{-4} and decreased with a 'poly' learning rate policy with power as 0.9. The batch size is set to 4 and the CycleGAN

model was trained for 15 epochs. The training process was conducted on a computer platform with an Intel(R) Core i7 9700K CPU, 64 GB of RAM, and two Nvidia RTX 2080Ti graphics processing units (GPUs).

Appendix E. Correlation analysis result in 2020

Correlation analysis was conducted using meteorological parameters from 2020. Seasonal and annual average values were used for analysis. Additional analysis for air quality parameters hasn't been conducted, because the available air quality stations in Shanghai are all located in compact urban areas which are inappropriate for analysis. As listed in Table 6, the results indicate a small seasonal variation for most of the correlation coefficients except for two cases: first, the relative humidity is correlated with artificial surfaces in the summer of Beijing which is not shown in other seasons; Second, in the autumn of Beijing and GBA, the correlation coefficients is different from other seasons.

Table 6 Correlation coefficients between land cover and meteorological factors for each season in 2020. Bold values (with |r|≥ 0.3) indicate a linear correlation. Spring: Mar. 21 to Jun. 20; Summer: Jun. 21 to Sept. 20; Autumn: Sept. 21 to Dec. 20; Winter: Dec. 21 to Mar. 20

	Season	Artificial			Natural		
		Beijing	Shanghai	GBA	Beijing	Shanghai	GBA
AT	Spring	**0.70**[**]	**0.64**[**]	**0.60**	**−0.75**[**]	**−0.62**[**]	**−0.60**[**]
	Summer	**0.70**[**]	**0.67**[*]	**0.60**	**−0.73**[**]	**−0.64**[**]	**−0.52**[**]
	Autumn	**0.68**[**]	**0.83**[**]	**0.59**[**]	**−0.73**[**]	**−0.81**[**]	**−0.66**[**]
	Winter	**0.66**[**]	**0.71**[**]	**0.60**[**]	**−0.71**[**]	**−0.75**[**]	**−0.64**[**]
	Annual	**0.70**[**]	**0.73**[**]	**0.66**	**−0.75**[**]	**−0.73**[**]	**−0.68**[*]
RH	Spring	−0.24	**−0.33**[**]	**−0.41**[**]	**0.31**	0.28[*]	**0.31**[**]
	Summer	**−0.41**[**]	**−0.41**[**]	**−0.39**[**]	**0.45**[**]	**0.34**[**]	0.25[**]
	Autumn	−0.03[**]	**−0.49**[**]	**−0.64**[**]	0.06	**0.44**[*]	**0.59**
	Winter	−0.02	**−0.41**[**]	**−0.55**[**]	0.05	**0.39**[*]	**0.49**
	Annual	−0.16[**]	**−0.42**[**]	**−0.53**[**]	0.20	**0.37**[**]	**0.44**
WS	Spring	**−0.47**[**]	−0.29[**]	−0.20[**]	**0.50**[**]	**0.44**[**]	0.14[**]
	Summer	**−0.38**[**]	−0.21[**]	−0.05[**]	**0.42**[**]	**0.37**[**]	−0.02[**]
	Autumn	−0.26[**]	−0.19[**]	**−0.41**[**]	0.27[**]	**0.34**[**]	**0.30**[**]
	Winter	**−0.36**[**]	−0.27[**]	−0.26[**]	**0.36**[**]	**0.40**[**]	0.19[**]
	Annual	**−0.39**[**]	−0.25[**]	−0.25[**]	**0.41**[**]	**0.39**[**]	0.18[**]

[**]. Significant correlation at 0.01 level (bilateral)

[*]. Significant correlation at 0.05 level (bilateral)

Appendix F. Correlation analysis result for air pollution

The $PM_{2.5}$ and PM_{10} concentrations are positively correlated with artificial land cover fraction in Shanghai, whereas this relationship is insignificant in Beijing and GBA (Table 7). The strong background PMs pollution and regional transport of pollutants can weaken the significance of intra-urban differences (Wang et al. 2014; Tao et al. 2017; Liu et al. 2017; Fan et al. 2020). Studies have shown that the air pollution in Beijing and GBA is significantly affected by its surrounding area (Liu et al. 2013; Xue et al. 2016; Tao et al. 2017). The local industrial activities and vehicle emission are the major contributors to PMs in Shanghai (Wang et al. 2014), which may explain its stronger correlation than the other two areas.

There is a ubiquitous correlation (0.26 to 0.38) between NO_2 concentrations and artificial land cover in all areas (Table 7). The traffic in urban areas with dense populations is usually heavier than that in rural areas as a major source for NO_2 in urban areas (Lee et al. 2014; Fan et al. 2020; Xue et al. 2020). The NO_2 concentrations of those sites vary significantly even with a similar artificial surface fraction (Fig. 13(c, e)). This can be explained by the fact that the traffic intensity can be significantly different even when the fraction of artificial land cover is similar. Furthermore, other local factors like ship emissions can be an important factor affecting the correlation result (Xue et al. 2020).

Traffic is also a major factor in total CO emissions (Hrebtov and Hanjalić 2019). The results show that the CO concentration also has a distinct variation even in areas with a similar artificial surface fraction (Fig. 14 a, d). However, the correlation between CO concentration and artificial land cover in GBA is weaker (−0.02 to 0.23) than that of NO_2. This is probably because the concentration of CO is also affected by agricultural and/or residential heating-related biomass burning in rural areas (Sokhi et al. 2021).

A negative correlation (−0.42 to −0.46) exists between the fraction of artificial surface and SO_2 concentrations in the summer of Shanghai and Beijing. The major sources of SO_2 are heavy industries and coal-fired power stations (Yoo et al. 2015; Xue et al. 2016, 2020; Li et al. 2017), and these heavy industries are normally located in suburban or rural areas. Therefore, the SO_2 emissions in rural industrial land can be larger than in urban residential areas. For instance, with a heavier industrial intensity, the SO_2 emission in the northern region of Shanghai is larger than in the denser urban central region (Xue et al. 2020). The local concentration of SO_2 also depends on the wind direction. In the summer of Shanghai, the prevailing southeast monsoon from the sea brings cleaner air, causing a lower SO_2 concentration in urban areas (Xue et al. 2020). However, in other seasons, the north wind from more polluted inland may lead to a similar SO_2 concentration throughout the city, making it a regional pollution problem and reducing the intra-urban variability. Therefore, the correlation in Spring can be weak. The correlation between SO_2 and artificial surfaces in GBA is weaker than in Shanghai and Beijing. Unlike Shanghai and Beijing,

there still are heavy industries and electric power stations (Hu et al. 2021) in the center of GBA (Foshan city), which can weaken urban–rural differences.

There is no clear correlation shown for the O_3 concentration, which may be because O_3 has a more complex

photochemical generation mechanism than other pollutants (Ren et al. 2021). The formation of O_3 can be affected by factors like air temperature and reduced solar radiation in rainy seasons. Therefore, the O_3 concentration shows a strong seasonal variation (Fan et al. 2020).

Table 7 Correlation coefficients between land cover and air quality factors for two seasons in 2021. Bold values (with |r|≥ 0.3) indicate a correlation

		Artificial			Natural		
	Season	Beijing	Shanghai	GBA	Beijing	Shanghai	GBA
$PM_{2.5}$	Spring (Mar. 21 to Jun. 20)	0.00	**0.35**[**]	0.11[**]	−0.13[**]	**−0.44**	−0.08[**]
	Summer (Jun. 21 to Sep. 20)	0.15[**]	0.22[**]	0.13[**]	−0.28[**]	**−0.31**[**]	−0.13[**]
	Overall (Mar. 21 to Sep. 20)	0.06[**]	0.28[**]	0.13[**]	−0.19[**]	**−0.39**[**]	−0.11[**]
PM_{10}	Spring (Mar. 21 to Jun. 20)	0.17[**]	**0.44**[**]	0.19	−0.25[**]	**−0.43**[**]	−0.17[**]
	Summer (Jun. 21 to Sep. 20)	0.10	**0.44**[**]	0.15[**]	−0.22[**]	**−0.45**[**]	−0.16[**]
	Overall (Mar. 21 to Sep. 20)	0.15[**]	**0.44**	0.18[**]	−0.24	**−0.45**[**]	−0.17[**]
NO_2	Spring (Mar. 21 to Jun. 20)	**0.34**[**]	**0.38**[**]	0.26[**]	**−0.49**[**]	**−0.30**[**]	**−0.30**[**]
	Summer (Jun. 21 to Sep. 20)	**0.31**[**]	**0.35**[**]	0.28[**]	**−0.48**[**]	−0.18[**]	**−0.31**[**]
	Overall (Mar. 21 to Sep. 20)	**0.33**[**]	**0.38**[**]	0.27[**]	**−0.49**[**]	−0.27[**]	**−0.31**[**]
CO	Spring (Mar. 21 to Jun. 20)	−0.03[**]	0.12[**]	0.23[**]	−0.03[**]	**−0.31**[**]	−0.21[**]
	Summer (Jun. 21 to Sep. 20)	0.05[**]	0.05[**]	0.20[**]	−0.16[**]	−0.18[**]	−0.17[**]
	Overall (Mar. 21 to Sep. 20)	0.01[**]	0.09[**]	0.22[**]	−0.12[**]	−0.27[**]	−0.20[**]
SO_2	Spring (Mar. 21 to Jun. 20)	−0.02[**]	0.13[**]	−0.18[**]	−0.05[**]	−0.23[**]	0.14[**]
	Summer (Jun. 21 to Sep. 20)	**−0.42**[**]	**−0.46**[**]	−0.10[**]	**0.42**[**]	**0.42**[**]	0.04[**]
	Overall (Mar. 21 to Sep. 20)	**−0.31**[**]	−0.29[**]	−0.14[**]	0.27[**]	0.21[**]	0.09[**]
O_3	Spring (Mar. 21 to Jun. 20)	−0.29[**]	−0.05[**]	−0.05[**]	**0.38**[**]	0.07[**]	0.13
	Summer (Jun. 21 to Sep. 20)	0.24[**]	0.10[**]	0.00[**]	**−0.30**[**]	−0.13[**]	0.01
	Overall (Mar. 21 to Sep. 20)	−0.03[**]	−0.01[**]	−0.03[**]	0.05[**]	0.00[**]	0.08

[**]. Significant correlation at 0.01 level (bilateral)

[*]. Significant correlation at 0.05 level (bilateral)
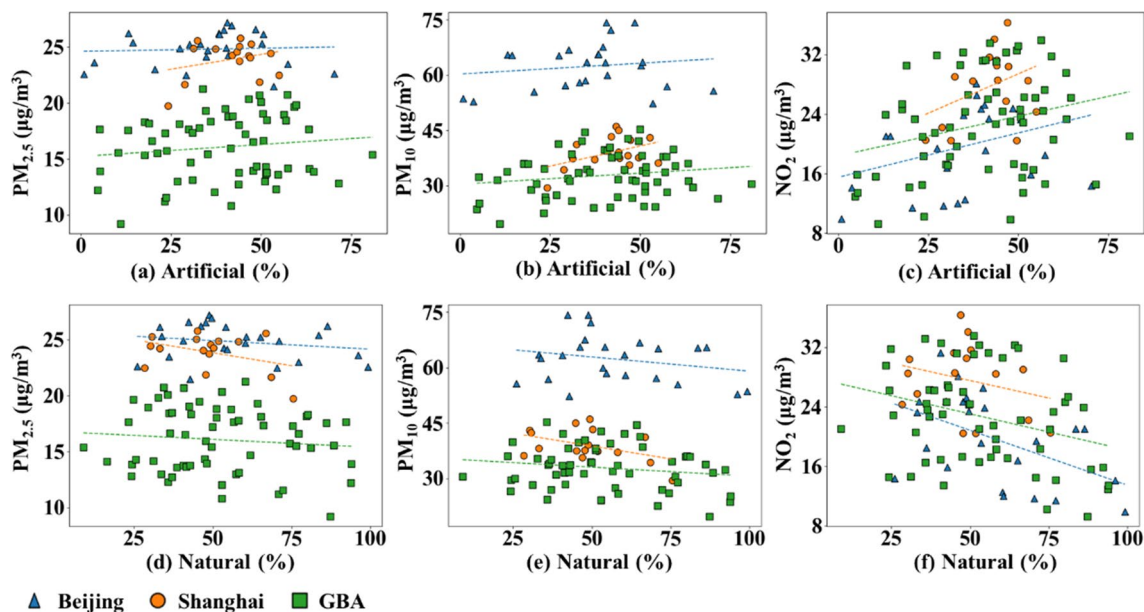

Fig. 13 Relationship between urban land cover indicators and air quality indicators (two-season average values, $PM_{2.5}$, PM_{10}, NO_2) in three regions
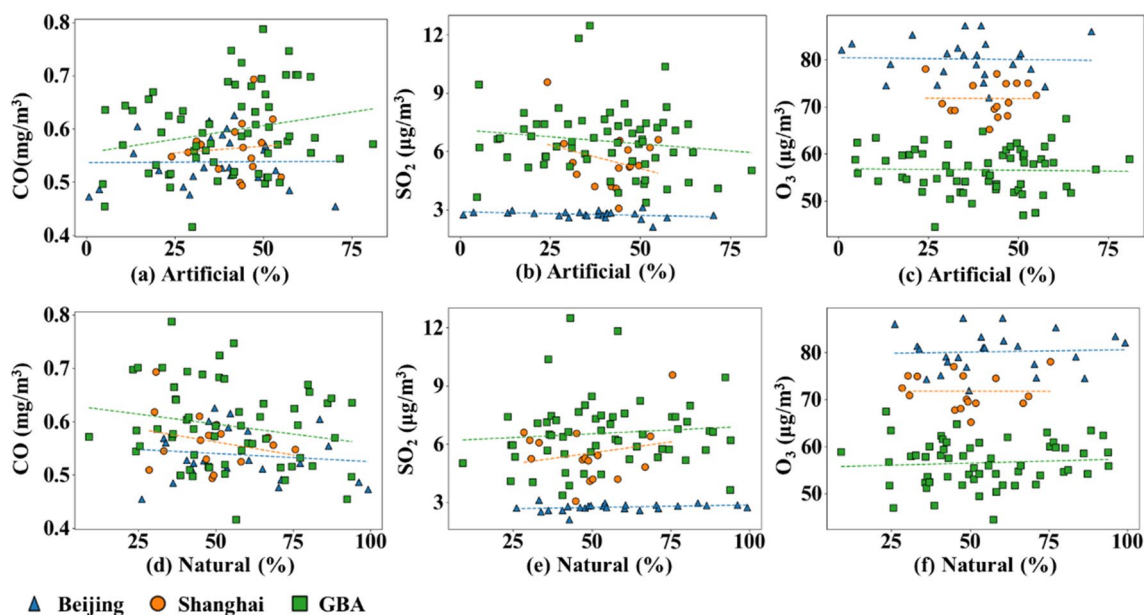
Fig. 14 Relationship between urban land cover indicators and air quality indicators (two-season average values, CO, SO$_2$, O$_3$) in three regions

Authors contributions Xiaotian Ding: Conceptualization, Methodology, Software, Formal analysis, Writing-Original Draft. Yifan Fan: Conceptualization, Resources, Writing-Review & Editing, Supervision. Yuguo Li: Writing-Review & Editing. Jian Ge: Resources, Supervision.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. There is no custom code produced during the data collection and model evaluation process. The models are mainly constructed with the semantic segmentation model DeepLab v3 (https://github.com/fregu856/deeplabv3), and CycleGAN (https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix).

Declarations

Ethical approval This manuscript does not contain any individual person's data in any form (including individual details, images or videos). There is no consent need for this manuscript.

Consent to participate This manuscript does not contain any individual person's data in any form. There is no consent need for this manuscript.

Consent for publication All of the authors have read and approved the paper for publication.

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

Aryal J, Neupane B (2023) Multi-Scale Feature Map Aggregation and Supervised Domain Adaptation of Fully Convolutional Networks for Urban Building Footprint Extraction. Remote Sens 15:488. https://doi.org/10.3390/rs15020488

Badrinarayanan V, Kendall A, Cipolla R (2017) SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Trans Pattern Anal Mach Intell 39:2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615

Chen W, Zhou Y, Wu Q et al (2020) Urban Building Type Mapping Using Geospatial Data: A Case Study of Beijing, China. Remote Sens 12:2805. https://doi.org/10.3390/rs12172805

Chen L-C, Papandreou G, Schroff F, Adam H (2017) Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv preprint arXiv:1706.05587. https://doi.org/10.48550/arXiv.1706.05587

Chen Y, Li W, Sakaridis C, et al (2018) Domain adaptive faster R-CNN for object detection in the Wild. Proceedings of the IEEE conference on computer vision and pattern recognition 3339–3348. https://doi.org/10.48550/arXiv.1803.03243

Diakogiannis FI, Waldner F, Caccetta P, Wu C (2020) ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS J Photogramm Remote Sens 162:94–114. https://doi.org/10.1016/j.isprsjprs.2020.01.013

Du H, Wang D, Wang Y et al (2016) Influences of land cover types, meteorological conditions, anthropogenic heat and urban area on surface urban heat island in the Yangtze River Delta Urban Agglomeration. Sci Total Environ 571:461–470. https://doi.org/10.1016/j.scitotenv.2016.07.012

Dugord PA, Lauf S, Schuster C, Kleinschmit B (2014) Land use patterns, temperature distribution, and potential heat stress risk - The case study Berlin, Germany. Comput Environ Urban Syst 48:86–98. https://doi.org/10.1016/j.compenvurbsys.2014.07.005

Fan Y, Wang Q, Yin S, Li Y (2019) Effect of city shape on urban wind patterns and convective heat transfer in calm and stable background conditions. Build Environ 162:106288. https://doi.org/10.1016/j.buildenv.2019.106288

Fan Y, Ding X, Hang J, Ge J (2020) Characteristics of urban air pollution in different regions of China between 2015 and 2019. Build Environ 180:107048. https://doi.org/10.1016/j.buildenv.2020.107048

Fan Y, Ding X, Wu J et al (2021) High spatial-resolution classification of urban surfaces using a deep learning method. Build Environ 200:107949. https://doi.org/10.1016/j.buildenv.2021.107949

Fu B, Li S, Lao Z et al (2023a) Multi-sensor and multi-platform retrieval of water chlorophyll a concentration in karst wetlands using transfer learning frameworks with ASD, UAV, and Planet CubeSate reflectance data. Sci Total Environ 901:165963. https://doi.org/10.1016/j.scitotenv.2023.165963

Fu B, Sun X, Li Y et al (2023b) Combination of super-resolution reconstruction and SGA-Net for marsh vegetation mapping using multi-resolution multispectral and hyperspectral images. Int J Digit Earth 16:2724–2761. https://doi.org/10.1080/17538947.2023.2234340

Gao L, Wen X, Guo Y et al (2014) Spatiotemporal Variability of Carbon Flux from Different Land Use and Land Cover Changes: A Case Study in Hubei Province, China. Energies 7:2298–2316. https://doi.org/10.3390/en7042298

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. IEEE conference on computer vision and pattern recognition (CVPR) 770–778. https://doi.org/10.48550/arXiv.1512.03385

Hoffman J, Wang D, Yu F, Darrell T (2016) FCNs in the wild: pixel-level adversarial and constraint-based adaptation. arXiv preprint arXiv:1612.02649. https://doi.org/10.48550/arXiv.1612.02649

Hoffman J, Tzeng E, Park T, et al (2018) CyCADA: Cycle-consistent adversarial domain adaptation. International conference on machine learning 1989–1998. https://doi.org/10.48550/arXiv.1711.03213

Hrebtov M, Hanjalić K (2019) River-Induced Anomalies in Seasonal Variation of Traffic-Emitted CO Distribution over the City of Krasnoyarsk. Atmosphere (basel) 10:407. https://doi.org/10.3390/atmos10070407

Hu M, Wang Y, Wang S et al (2021) Spatial-temporal heterogeneity of air pollution and its relationship with meteorological factors in the Pearl River Delta, China. Atmos Environ 254:118415. https://doi.org/10.1016/j.atmosenv.2021.118415

Ji S, Di W, Luo M (2021) Generative Adversarial Network-Based Full-Space Domain Adaptation for Land Cover Classification from Multiple-Source Remote Sensing Images. IEEE Trans Geosci Remote Sens 59:3816–3828. https://doi.org/10.1109/TGRS.2020.3020804

Jing L, Tian Y (2021) Self-Supervised Visual Feature Learning With Deep Neural Networks: A Survey. IEEE Trans Pattern Anal Mach Intell 43:4037–4058. https://doi.org/10.1109/TPAMI.2020.2992393

Klonus S, Ehlers M (2009) Performance of evaluation methods in image fusion. 2009 12th Int Conf Inf Fusion. FUSION 2009:1409–1416

Lee JH, Wu CF, Hoek G et al (2014) Land use regression models for estimating individual NOx and NO2 exposures in a metropolis with a high density of traffic roads and population. Sci Total Environ 472:1163–1171. https://doi.org/10.1016/j.scitotenv.2013.11.064

Li S, Feng K, Li M (2017) Identifying the main contributors of air pollution in Beijing. J Clean Prod 163:S359–S365. https://doi.org/10.1016/j.jclepro.2015.10.127

Li C, Wang Z, Li B et al (2019a) Investigating the relationship between air pollution variation and urban form. Build Environ 147:559–568. https://doi.org/10.1016/j.buildenv.2018.06.038

Li Y, Yuan L, Vasconcelos N (2019b) Bidirectional learning for domain adaptation of semantic segmentation. IEEE/CVF conference on computer vision and pattern recognition (CVPR) 6929–6938. https://doi.org/10.48550/arXiv.1904.10620

Li W, Li F, Luo Y, et al (2020) Deep domain adaptive object detection: a survey. IEEE Symposium series on computational intelligence (SSCI) 1808–1813. https://doi.org/10.48550/arXiv.2002.06797

Lin Y-C, Cheng M-T (2007) Evaluation of formation rates of NO2 to gaseous and particulate nitrate in the urban atmosphere. Atmos Environ 41:1903–1910. https://doi.org/10.1016/j.atmosenv.2006.10.065

Lin L, Chan TO, Ge E et al (2020) Effects of urban land expansion on decreasing atmospheric moisture in Guangdong, South China. Urban Clim 32:100626. https://doi.org/10.1016/j.uclim.2020.100626

Liu W, You H, Dou J (2009) Urban-rural humidity and temperature differences in the Beijing area. Theor Appl Climatol 96:201–207. https://doi.org/10.1007/s00704-008-0024-6

Liu XG, Li J, Qu Y et al (2013) Formation and evolution mechanism of regional haze: A case study in the megacity Beijing, China. Atmos Chem Phys 13:4501–4514. https://doi.org/10.5194/acp-13-4501-2013

Liu B, Wu J, Zhang J et al (2017) Characterization and source apportionment of PM2.5 based on error estimation from EPA PMF 5.0 model at a medium city in China. Environ Pollut 222:10–22. https://doi.org/10.1016/j.envpol.2017.01.005

Liu Y, Xu Y, Weng F et al (2021) Impacts of urban spatial layout and scale on local climate: A case study in Beijing. Sustain Cities Soc 68:102767. https://doi.org/10.1016/j.scs.2021.102767

Liu Q, Peng J, Ning Y et al (2023) Refined Prototypical Contrastive Learning for Few-Shot Hyperspectral Image Classification. IEEE Trans Geosci Remote Sens 61:1–14. https://doi.org/10.1109/TGRS.2023.3257341

Liu C, Zeng D, Wu H, et al (2020) Urban land cover classification of high-resolution aerial imagery using a relation-enhanced multi-scale convolutional network. Remote Sens 12. https://doi.org/10.3390/rs12020311

Lokoshchenko MA (2017) Urban Heat Island and Urban Dry Island in Moscow and Their Centennial Changes. J Appl Meteorol Climatol 56:2729–2745. https://doi.org/10.1175/JAMC-D-16-0383.1

Luo M, Ji S (2022) Cross-spatiotemporal land-cover classification from VHR remote sensing images with deep learning based domain adaptation. ISPRS J Photogramm Remote Sens 191:105–128. https://doi.org/10.1016/j.isprsjprs.2022.07.011

Mazza D, Pagani M (2020) Automatic Differentiation in PCF. Proc ACM Program Lang 5:1–27. https://doi.org/10.1145/3434309

National Bureau of Statistics of China (2015) China Statistical Yearbook. China Stat. Press, Beijing

Oke TR (1982) The energetic basis of the urban heat island. Q J R Meteorol Soc 108:1–24. https://doi.org/10.1002/qj.49710845502

Oke TR, Mills G, Christen A, Voogt JA (2017) Urban Climates. Cambridge University Press, Cambridge

Pan SJ, Yang Q (2010) A survey on transfer learning. IEEE Trans Knowl Data Eng 22:1345–1359. https://doi.org/10.1109/TKDE.2009.191

Peng J, Huang Y, Sun W et al (2022) Domain Adaptation in Remote Sensing Image Classification: A Survey. IEEE J Sel Top Appl Earth Obs Remote Sens 15:9842–9859. https://doi.org/10.1109/JSTARS.2022.3220875

Ren J, Hao Y, Simayi M et al (2021) Spatiotemporal variation of surface ozone and its causes in Beijing, China since 2014. Atmos

Environ 260:118556. https://doi.org/10.1016/j.atmosenv.2021.118556

Ronneberger O, Fischer P, Brox T (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 9351:12–20. https://doi.org/10.1007/978-3-319-24574-4_2

Scott GJ, Marcum RA, Davis CH, Nivin TW (2017) Training Deep Convolutional Neural Networks for Land-Cover Classification of High-Resolution Imagery. IEEE Geosci Remote Sens Lett 14:1638–1642. https://doi.org/10.1109/LGRS.2017.2722988

Sha C, Wang X, Lin Y et al (2018) The impact of urban open space and 'lift-up' building design on building intake fraction and daily pollutant exposure in idealized urban models. Sci Total Environ 633:1314–1328. https://doi.org/10.1016/j.scitotenv.2018.03.194

Sokhi RS, Singh V, Querol X, et al (2021) A global observational analysis to understand changes in air quality during exceptionally low anthropogenic emission conditions. Environ Int 157. https://doi.org/10.1016/j.envint.2021.106818

Stewart ID, Oke TR (2012) Local climate zones for urban temperature studies. Bull Am Meteorol Soc 93:1879–1900. https://doi.org/10.1175/BAMS-D-11-00019.1

Tao J, Zhang L, Cao J et al (2017) Source apportionment of PM2.5 at urban and suburban areas of the Pearl River Delta region, south China - With emphasis on ship emissions. Sci Total Environ 574:1559–1570. https://doi.org/10.1016/j.scitotenv.2016.08.175

Tao H, Xing J, Zhou H et al (2018) Impacts of land use and land cover change on regional meteorology and air quality over the Beijing-Tianjin-Hebei region, China. Atmos Environ 189:9–21. https://doi.org/10.1016/j.atmosenv.2018.06.033

Toldo M, Maracani A, Michieli U, Zanuttigh P (2020) Unsupervised Domain Adaptation in Semantic Segmentation: A Review. Technologies 8:35. https://doi.org/10.3390/technologies8020035

van Engelen JE, Hoos HH (2020) A survey on semi-supervised learning. Mach Learn 109:373–440. https://doi.org/10.1007/s10994-019-05855-6

Volpi M, Tuia D (2017) Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. IEEE Trans Geosci Remote Sens 55:881–893. https://doi.org/10.1109/TGRS.2016.2616585

Wang Y, Li L, Chen C et al (2014) Source apportionment of fine particulate matter during autumn haze episodes in Shanghai, China. J Geophys Res Atmos 119:1903–1914. https://doi.org/10.1002/2013JD019630

Wilson G, Cook DJ (2020) A Survey of Unsupervised Deep Domain Adaptation. ACM Trans Intell Syst Technol 11:1–46. https://doi.org/10.1145/3400066

Wong SC, Gatt A, Stamatescu V, McDonnell MD (2016) Understanding Data Augmentation for Classification: When to Warp? 2016 Int Conf Digit Image Comput Tech Appl DICTA 2016. https://doi.org/10.1109/DICTA.2016.7797091

Xu D, Zhou D, Wang Y et al (2019) Field measurement study on the impacts of urban spatial indicators on urban climate in a Chinese basin and static-wind city. Build Environ 147:482–494. https://doi.org/10.1016/j.buildenv.2018.10.042

Xu H, Chen H, Zhou X et al (2020) Research on the relationship between urban morphology and air temperature based on mobile measurement: A case study in Wuhan, China. Urban Clim 34:100671. https://doi.org/10.1016/j.uclim.2020.100671

Xu M, Wu M, Chen K et al (2022) The Eyes of the Gods: A Survey of Unsupervised Domain Adaptation Methods Based on Remote Sensing Data. Remote Sens 14:4380. https://doi.org/10.3390/rs14174380

Xue Y, Tian H, Yan J et al (2016) Temporal trends and spatial variation characteristics of primary air pollutants emissions from coal-fired industrial boilers in Beijing, China. Environ Pollut 213:717–726. https://doi.org/10.1016/j.envpol.2016.03.047

Xue R, Wang S, Li D et al (2020) Spatio-temporal variations in NO2 and SO2 over Shanghai and Chongming Eco-Island measured by Ozone Monitoring Instrument (OMI) during 2008–2017. J Clean Prod 258:120563. https://doi.org/10.1016/j.jclepro.2020.120563

Yan H, Dong L (2015) The impacts of land cover types on urban outdoor thermal environment: the case of Beijing, China. J Environ Heal Sci Eng 13:1–7. https://doi.org/10.1186/s40201-015-0195-x

Yang H, Chen T, Lin Y et al (2020) Integrated impacts of tree planting and street aspect ratios on CO dispersion and personal exposure in full-scale street canyons. Build Environ 169:106529. https://doi.org/10.1016/j.buildenv.2019.106529

Yang Y, Xu Z (2020) Rethinking the value of labels for improving class-imbalanced learning. Advances in neural information processing systems 33:19290–19301. https://doi.org/10.48550/arXiv.2006.07529

Yoo JM, Jeong MJ, Kim D et al (2015) Spatiotemporal variations of air pollutants (O3, NO2, SO2, CO, PM10, and VOCs) with land-use types. Atmos Chem Phys 15:10857–10885. https://doi.org/10.5194/acp-15-10857-2015

Zhang W, Huang B, Luo D (2014) Effects of land use and transportation on carbon sources and carbon sinks: A case study in Shenzhen, China. Landsc Urban Plan 122:175–185. https://doi.org/10.1016/j.landurbplan.2013.09.014

Zhang Q, Xu D, Zhou D et al (2020) Associations between urban thermal environment and physical indicators based on meteorological data in Foshan City. Sustain Cities Soc 60:102288. https://doi.org/10.1016/j.scs.2020.102288

Zhang J, Xu S, Sun J et al (2022) Unsupervised Adversarial Domain Adaptation for Agricultural Land Extraction of Remote Sensing Images. Remote Sens 14:1–20. https://doi.org/10.3390/rs14246298

Zhao J, Zhao X, Liang S et al (2020a) Assessing the thermal contributions of urban land cover types. Landsc Urban Plan 204:103927. https://doi.org/10.1016/j.landurbplan.2020.103927

Zhao Y, Chew LW, Kubilay A, Carmeliet J (2020b) Isothermal and non-isothermal flow in street canyons: A review from theoretical, experimental and numerical perspectives. Build Environ 184:107163. https://doi.org/10.1016/j.buildenv.2020.107163

Zheng S, Guldmann JM, Liu Z, Zhao L (2018) Influence of trees on the outdoor thermal environment in subtropical areas: An experimental study in Guangzhou, China. Sustain Cities Soc 42:482–497. https://doi.org/10.1016/j.scs.2018.07.025

Zhu J-Y, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. IEEE International conference on computer vision (ICCV) 2242–2251. https://doi.org/10.48550/arXiv.1703.10593