Cross-scale Urban Land Cover Mapping: Empowering Classification through Transfer Learning and Deep Learning Integration

Zhe Wang
Deparment of Computer Science
University of Idaho
Moscow, United States
zwang@uidaho.edu

Chao Fan
Department of Geography
New Mexico State University
Las Cruces, United States
cfan@nmsu.edu

Xian Min
Department of Computer Science
University of Idaho
Idaho Falls, United States
mxian@uidaho.edu

Shoukun Sun
Department of Computer Science
University of Idaho
Idaho Falls, United States
ssun@uidaho.edu

Xiaogang Ma
Deparment of Computer Science
University of Idaho
Moscow, United States
max@uidaho.edu

Xiang Que
Deparment of Computer Science
University of Idaho
Moscow, United States
xiangq@uidaho.edu

***Abstract*— Urban land cover mapping is essential for effective urban planning and resource management. Thanks to its ability to extract intricate features from urban datasets, deep learning has emerged as a powerful technique for urban classification. The U-net architecture has achieved state-of-the-art land cover classification performance, highlighting its potential for mapping urban trees at different spatial scales. However, deep learning approaches often require large, labeled datasets, which are challenging to acquire for specific urban contexts. Transfer learning addresses this limitation by leveraging pre-trained deep learning models on extensive datasets and adapting them to smaller urban datasets with limited labeled samples. Transfer learning can enhance classification performance and generalization ability. In this study, we proposed a novel cross-scale framework that integrates transfer learning and deep learning for urban land cover mapping. The framework utilizes pre-trained deep learning models, trained on diverse urban datasets, as a foundation for classification. These models are then finetuned using transfer learning techniques on smaller urban datasets, tailoring them to the specific characteristics of the target urban context. To evaluate the effectiveness and feasibility of the proposed framework, extensive evaluations are conducted across different cities and years. Performance metrics such as accuracy and dice score are employed to assess the framework's classification capabilities. The results of this study contribute to advancing the field of urban classification by demonstrating the effectiveness and feasibility of the cross-scale framework. By combining transfer learning and deep learning, the framework improves classification accuracy, efficiency, and scalability in urban land cover mapping tasks. Leveraging the strengths of transfer learning and deep learning holds great promise for accurate and efficient urban land cover mapping, providing valuable insights for urban planning and resource management decision-making.)**

Keywords—Remote sensing, Deep Learning, Transfer Learning, U-net, Land Cover Classification Introduction

1. Introduction

Advancements in remote sensing technology have improved data quality, spatial resolution, revisit periods, and coverage area [1]. This abundance of data presents challenges in managing image collections. Image classification is a fundamental task in remote sensing for applications such as urban planning and land cover mapping [2]. Traditional methods often struggle to distinguish complex land structures using limited rules that rely on low-level spectral and spatial features [3]. Deep learning, particularly the U-net architecture [4], has demonstrated superior land cover mapping performance due to its ability to extract multiscale and multilevel features [5].

Deep learning models, built upon the Convolutional Neural Network (CNN) architecture have shown remarkable abilities in expressing and processing data with high accuracy [6]. The Fully Convolutional Network (FCN) extends the CNN framework to perform dense pixel-level prediction, enabling multi-class classification at the pixel level [7]. Derived from the FCN, the U-net further refines boundary delineation and has been successfully applied in biomedical segmentation, medical image reconstruction, and speech enhancement [8]–[10]. Despite its success in these fields, the utilization of U-net in land cover mapping remains limited.

In my previous study, I evaluated the feasibility and effectiveness of the U-net in tree canopy extraction, demonstrating its state-of-the-art performance [5]. The paper highlights the potential of applying U-net in urban tree mapping at different spatial scales, accurately identifying and delineating tree canopies and various land cover elements. However, challenges persist, such as the availability of freely accessible high-resolution imagery and the scarcity of publicly available training datasets. One potential solution lies in incorporating transfer learning techniques.

Transfer learning is a widely adopted approach in deep learning, encompassing the transfer of learned skills and knowledge from one learning situation to another. In remote sensing, transfer learning proves beneficial in addressing the limited availability of training data, a common challenge due to the cost and effort required for data collection and annotation.

By leveraging pre-trained models trained on large-scale image datasets, transfer learning enables the finetuning of these models on smaller target datasets, improving their performance. Notably, pre-trained models like Very Deep Convolutional Networks (VGGNet) [11], Residual Neural Network (ResNet) [12], and U-net offer transfer learning capabilities for remote sensing applications.

Although transfer learning has been extensively used in remote sensing, its application with free-access remote sensing products is relatively limited. Free-access products often possess lower spatial resolution and limited spectral bands than commercial high-resolution data. This mismatch between pre-trained models and the unique characteristics of free-access products may hinder the effectiveness of transfer learning. Additionally, the requirement for large and diverse training datasets poses challenges in finetuning pre-trained models for free-access products, leading to potential overfitting or poor generalization.

Based on the findings of my previous study, we proposed a cross-scale transfer learning framework to extract land cover features automatically. This framework addresses the challenges faced in remote sensing image classification while processing vast data. The insights and results obtained from this study contribute to the advancement of land cover mapping in remote sensing applications.

2. Methodology and Data

2.1 Methodology

2.1.1 U-net

U-net was developed and first used for biomedical image segmentation. Its architecture is an encoder network followed by a decoder neural network. The U-net architecture consists of a contracting path, which gradually reduces the spatial resolution of the input image, and an expanding path, which gradually increases the spatial resolution of the output segmentation map. The contracting path comprises convolutional and pooling layers, while the expanding path comprises deconvolutional layers, which upsample the feature maps to the original spatial resolution. Unlike classification tasks, where the end result of the deep network is all that matters, semantic segmentation involves not just discrimination at the pixel level but also a technique to project the discriminative features learned at different stages of the encoder onto the pixel space.

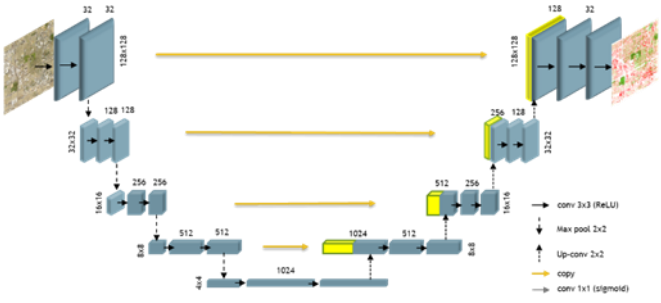

Figure 1. U-net architecture

2.1.2 Finetuning

Transfer learning is one of the most popular approaches in deep learning. We can say transfer learning is a machine learning method. The skills and knowledge learned in one situation can be applied to another learning situation [13]. **F**inetuning is a crucial step in transfer learning. In transfer learning, a pre-trained model, which has been trained on a large dataset, is used as a starting point for a new task or dataset. Finetuning involves further training the pre-trained model on the new task-specific or target dataset. Finetuning enables the transfer of knowledge from the pre-trained model, which has learned useful features from a large dataset to the new target task with a smaller dataset. By leveraging the pre-trained model's learned representations, finetuning can significantly improve the performance and convergence speed of the model on the target task.

In this study, we observe a model developed for one task being utilized as a starting point for a model on another task. These tasks typically share certain similarities. Although the two datasets employed exhibit some differences as they were acquired in distinct locations with varying urban structures, they also demonstrate significant similarities. Both datasets consist of high-resolution orthophotos with four identical spectral bands captured in urban areas with comparable features such as trees, buildings, roads, and grass. These shared characteristics provide opportunities to transfer knowledge through the trained models.

2.2 Data

We selected a small area in Phoenix as the target classification data. The target dataset used in this study was obtained from the National Agriculture Imagery Program (NAIP) [14]. Phoenix is a city located in the southwestern United States, in the state of Arizona. It is known for its hot, dry desert climate, hot summers and mild winters. The city is situated in a valley surrounded by mountains, which provide some relief from the intense heat. Figure 2 (a) shows the NAIP image from 2015, while Figure 2 (b) depicts the land cover and land use (LULC) map generated by the Central Arizona-Phoenix Long-Term Ecological Research (CAPLTER) project [15]. Figure 2 (a) represents one-quarter of Figure 2 (c), and the LULC data were clipped as labels for Figure 2 (c). Only one-quarter of the NAIP data and corresponding ground truth data were used for training the finetuning process. The proposed framework predicts the entire Figure 2 (a). The remaining LULC labels were also used to evaluate the performance of the proposed cross-scale framework.

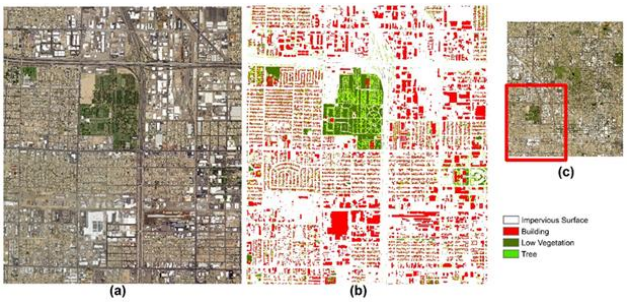

Figure 2. NAIP data in Phoenix in 2015 (a) NAIP data used for training and finetuning, (b) Corresponding ground truth, and (c) the location of the data used for training.

The International Society published the source Potsdam image dataset for Photogrammetry and Remote Sensing (ISPRS) [16]. The ISPRS Potsdam dataset is a widely used benchmark dataset for evaluating remote sensing image analysis methods. All images have the spatial resolution of 5 cm and consist of four spectral bands red (R), green (G), blue (B), and near-infrared (IR).

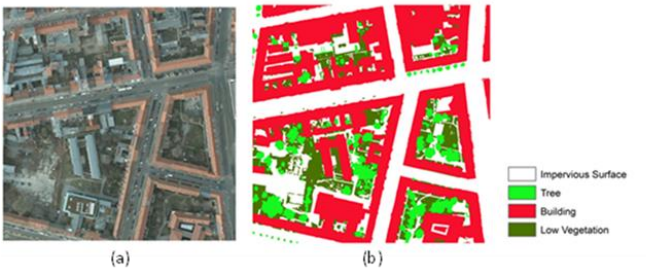

Figure 3. Example patches of the ISPRS Potsdam Dataset (a) orthophoto and (b) ground truth

3. Experiments and Results

3.1 Results of the Cross-Scale Framework Experiments

This study tests the effectiveness of U-net for urban land cover mapping and evaluates the transfer learning-based U-net through four experiments and two evaluations. Experiment 1 trains a U-net model on the Potsdam ISPRS dataset, to establish baseline hyperparameters. Experiment 2 evaluates the U-net's effectiveness in land cover mapping at different scales. Experiment 3 applies transfer learning using pre-trained models from Experiment 2 to classify NAIP images at different scales.

In Experiment 4, models are finetuned on a limited NAIP dataset and applied to classify the rest. Evaluation 1 compares the proposed cross-scale transfer learning framework with object-based image analysis (OBIA) and maximum likelihood (ML) methods. Evaluation 2 assesses the framework's robustness and feasibility by evaluating performance over time and in different geographic areas. Remote sensing data from the same study area in other years and different study areas in the same year are used. This evaluation investigates the framework's ability to detect land cover changes and its applicability to diverse areas.

Figure 2 shows the entire experiment framework. The experiment framework involves working with two datasets: a source dataset with a 5-cm resolution and labels, and a target dataset with a 100-cm spatial resolution from the NAIP dataset. By leveraging transfer learning and focusing on a quarter of the labels from the target image, the framework enables the prediction of the entire study area. Additionally, the framework maximizes the benefits of scale, which affects the receptive field during neural network training.

Table 1 shows the evaluation metrics from the finetuned cross-scale model on 2015 NAIP data. The model shows strong performance for predicting Impervious Surfaces and Building pixels, with accuracy above 83% and Dice scores around 0.82-0.827. However, it performs slightly lower for the Low Vegetation and Tree classes, achieving an accuracy of around 75% and Dice coefficients around 0.72-0.73. The model demonstrates reasonably good performance across classes, although specific task requirements should be considered for evaluating its adequacy.

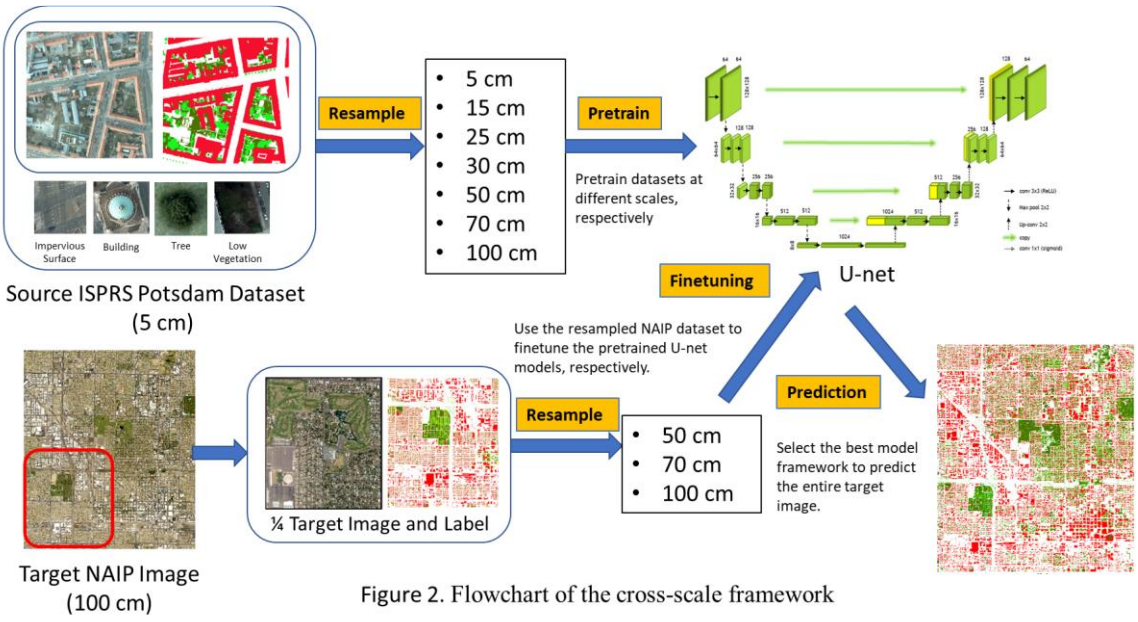

Figure 2. Flowchart of the cross-scale framework

Table 1. Evaluation metrics for each class on 2015 data.

	Accuracy	Precision	Recall	Dice
Impervious Surface	84.30%	87.20%	78.80%	0.83
Building	83.20%	83.70%	81.10%	0.82
Low Vegetation	75.40%	75.20%	70.20%	0.72
Tree	75.40%	70.20%	75.40%	0.73

3.2 Performance Evaluation: Year and City Variations

Figure 4 shows the classification result from the cross-scale framework in Figure 2. We also selected a small example to illustrate the detailed classification map (Figure 4(c) and 4(d)).

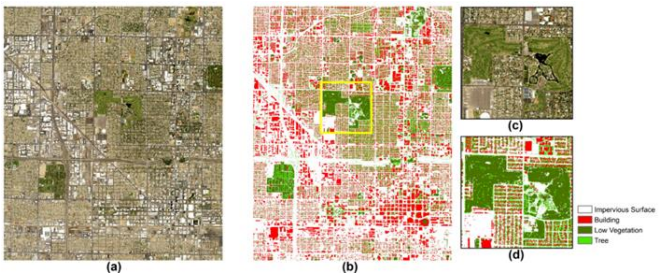

Figure 4. Land cover classification results of Phoenix in Figure 4. Land cover classification results of Phoenix in 2015: (a) NAIP data in Phoenix. (b) Corresponding predicted land cover map to (a). (c) NAIP data of a subset area. (d) Corresponding predicted land cover map to (c).

Additional data in the same area but from different years were downloaded and analyzed to explore further the effectiveness of the classification framework proposed in this study. Data in the study area were obtained for the years 2013 and 2019 (Figure 5 (a) and 6 (a)). The spatial resolution of the NAIP data was resampled to 50 cm, and the best model achieved in the previous steps was applied to the resampled data. The classification results were analyzed to determine whether the framework can achieve high accuracy and provide a reliable land cover classification for different years in the same area. Additional data provides a more comprehensive evaluation of the framework's effectiveness and improves its applicability for different periods.

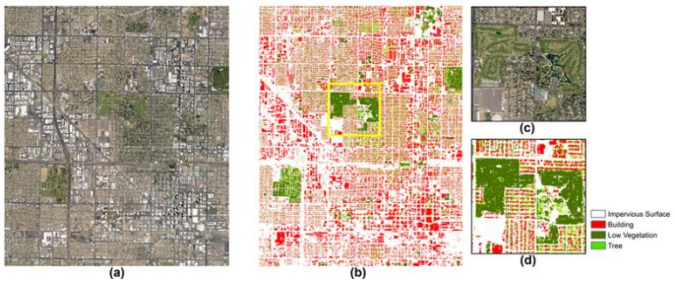

Figure 5. Land cover classification results of Phoenix in 2013: (a) NAIP data in Phoenix. (b) Corresponding predicted land cover map to (a). (c) NAIP data of a subset area. (d) Corresponding predicted land cover map to (c).

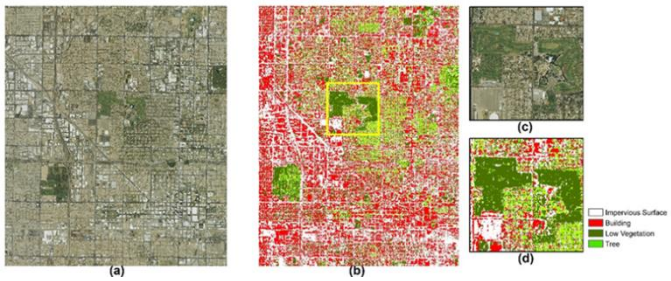

Figure 6. Land cover classification results of Phoenix in 2019: (a) NAIP data in Phoenix. (b) Corresponding predicted land cover map to (a). (c) NAIP data of a subset area. (d) Corresponding predicted land cover map to (c).

To test the robustness in a different city, I chose the city of Las Vegas in the same year, 2015. Las Vegas is also a city on the floor of the Mojave Desert. It has a similar landscape and climate environment to Phoenix. Figure 7 shows the orthophoto and corresponding prediction results.

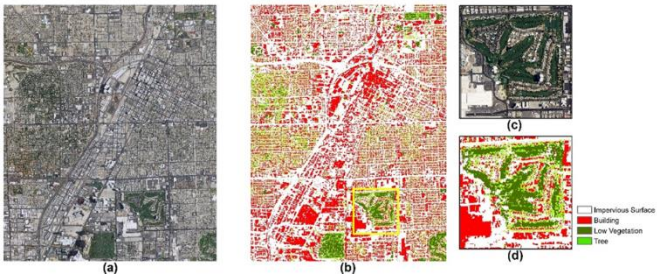

Figure 7. Land cover classification results of Las Vegas in 2015: (a) NAIP data in Phoenix. (b) Corresponding predicted land cover map to (a). (c) NAIP data of a subset area. (d) Corresponding predicted land cover map to (c).

Since ground truth data for these three images were unavailable, a sample of 1000 points within the study area was randomly selected, and each point was manually labeled to create a ground truth dataset. This approach ensured that the accuracy of the classification framework could be evaluated using a reliable and representative ground truth dataset.

The overall accuracy of the classification model on Phoenix data improved from 71.3% in 2013 to 85% in 2019. To test the robustness in a different city, I chose the city of Las Vegas in the same year, 2015. Las Vegas is also a city on the floor of the Mojave Desert. It has a similar landscape and climate environment to Phoenix. The overall accuracy (OA) of the classification model is moderate at 75.92%.

Table 2. Evaluation metrics for each class on different data

	Predicted Class	Accuracy
Phoenix 2013 OA: 71.3%	Impervious Surface	76.70%
	Building	80.60%
	Low Vegetation	89.50%
	Tree	0.958

Phoenix 2019 OA: 87.2%	Impervious Surface	89.00%
	Building	89.40%
	Low Vegetation	97.50%
	Tree	94.10%
Las Vegas 2015 OA: 85%	Impervious Surface	0.818
	Building	83.60%
	Low Vegetation	94.70%
	Tree	91.70%

3.3 Comparison with OBIA and ML

To provide a comparison of the classification framework's performance, the results were compared with two commonly used remote sensing classification methods, object-based image analysis (OBIA) and maximum likelihood (ML). OBIA and ML methods require extensive manual labor to achieve accurate and reliable results. Figure 8 presents the predicted land cover maps using the cross-scale framework proposed in this study and two commonly used remote sensing classification methods, OBIA and ML. Figure 8 (b) shows the predicted map from the cross-scale framework, while Figures 8 (c) and (d) show the predicted maps from OBIA and ML, respectively. From the maps, it is clear that the segmentation of land cover features is evident. The cross-scale framework was able to classify more impervious surfaces and better recognize open land areas as impervious surfaces, which is essential for urban land management and planning. ML, on the other hand, struggled to delineate building boundaries, resulting in irregularly shaped buildings in Figure 8 (d). OBIA also had some classification errors, wrongly classifying some roads as buildings and some low vegetation as trees, which can be problematic for urban planning and management.

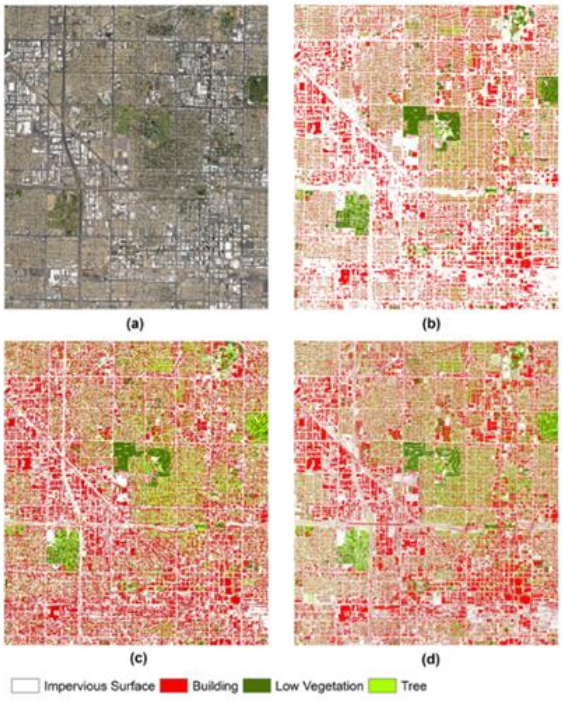

Figure 8. Comparisons of land cover classification performance between Cross-scale framework, OBIA, and ML (a) NAIP data of the selected area in Phoenix in 2019, (b) Prediction result using the proposed framework (cross-scale framework) in this study, (c) Prediction result using OBIA method, (d) Prediction result using ML method.

Table 3 shows the classification results for four land cover classes using three different classification methods. The cross-scale framework outperforms the other two methods, OBIA and ML, regarding overall accuracy (OA) for all classes except for the building class. For the "Impervious Surface" class, the cross-scale framework achieved an OA of 89%, higher than the OBIA's 75% and ML's 78%. Similarly, for the "Low Vegetation" class, the cross-scale framework achieved an OA of 97.5%, significantly higher than OBIA's 95.5% and ML's 94.4%. For the "Tree" class, the proposed method achieved an OA of 94.1%, which is higher than OBIA's 95.7% but lower than ML's 92.8%. The results suggest that the proposed cross-scale framework method outperforms the other two techniques.

Table 3. Evaluation metrics of different models

	Predicted Class	Accuracy
OBIA OA: 71.7%	Impervious Surface	74.80%
	Building	77.40%
	Low Vegetation	95.50%
	Tree	95.70%
ML OA: 72.8%	Impervious Surface	0.781
	Building	80.30%
	Low Vegetation	94.40%
	Tree	92.80%
Cross-scale Framework OA: 85%	Impervious Surface	89.00%
	Building	0.894
	Low Vegetation	97.50%
	Tree	94.10%

4. Conclusions and Discussions:

This study introduced a deep learning-based framework for land cover classification in urban areas using transfer learning and multiscale segmentation. The results demonstrated the effectiveness of the proposed framework in achieving high accuracy in classifying impervious surfaces, buildings, low vegetation, and trees, with an overall accuracy of 82.45%. The framework outperformed traditional classification methods like object-based image analysis (OBIA) and maximum likelihood (ML) in terms of overall accuracy metric.

The findings of the study highlight the potential of deep learning and transfer learning in remote sensing image classification for urban land cover mapping. Deep learning models, particularly the U-Net architecture, perform superiorly in capturing multiscale and multilevel features, enabling accurate land cover classification in complex urban areas.

Transfer learning proved valuable in addressing the limited availability of training data, allowing knowledge transfer from pre-trained models to improve performance on target datasets.

The experiments conducted in this study provided valuable insights. The importance of spectral information was demonstrated, showing that the inclusion of the infrared band improved classification results for low vegetation and trees, while RGB models performed better for buildings. Spatial resolution was also found to be critical, with the 15-cm model achieving the highest overall accuracy when evaluating land cover mapping at different scales.

However, the study also identified limitations that need to be addressed in future research. The limited sample size and reliance on manually labeled data may impact the generalizability and cost-effectiveness of the framework. To enhance the robustness and applicability of the framework, it is important to incorporate more diverse datasets from different regions and explore the use of semi-supervised or unsupervised learning methods to reduce the need for manual labeling.

Future studies could also consider expanding the range of land cover classes and exploring advanced deep learning architectures, such as attention-based models or graph convolutional networks, to improve classification performance. Integrating other data sources, such as socioeconomic and demographic data, would provide a more comprehensive understanding of the urban environment and enhance land cover classification accuracy.

Evaluating the framework on different datasets and over time demonstrated its adaptability and reliability. The framework exhibited consistent performance across additional years in the same area and achieved moderate accuracy when applied to another city. These results indicate the potential for the framework to be applied to various datasets and landscapes, making it a promising approach for land cover classification studies in different urban contexts.

The proposed cross-scale framework has practical implications for urban planning, environmental management, and disaster response. Accurate and efficient land cover classification is essential for making informed decisions in these areas. The proposed framework can further enhance its feasibility, accuracy, and applicability by addressing the limitations and considering future directions, such as dataset diversification and advanced architecture exploration.

In conclusion, this study advances remote sensing image classification for urban land cover mapping. The proposed deep learning-based framework, incorporating transfer learning and multiscale segmentation, demonstrates high accuracy in land cover classification. The findings underscore the significance of spectral information and spatial resolution and the benefits of transfer learning in achieving accurate and efficient land cover mapping. Further research and improvements in dataset diversity, labeling approaches, and model architectures will strengthen the framework's effectiveness and broaden its potential applications in urban planning, environmental management, and related fields.

References

[1] R. R. Navalgund, V. Jayaraman, and P. S. Roy, "Remote sensing applications: an overview," *current science*, pp. 1747–1766, 2007.

[2] L. M. G. Fonseca, L. M. Namikawa, and E. F. Castejon, "Digital image processing in remote sensing," in *2009 Tutorials of the XXII Brazilian Symposium on Computer Graphics and Image Processing*, IEEE, 2009, pp. 59–71.

[3] S. Zhao, X. Liu, C. Ding, S. Liu, C. Wu, and L. Wu, "Mapping rice paddies in complex landscapes with convolutional neural networks and phenological metrics," *GIScience & Remote Sensing*, vol. 57, no. 1, pp. 37–48, 2020.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[5] Z. Wang, C. Fan, and M. Xian, "Application and Evaluation of a Deep Learning Architecture to Urban Tree Canopy Mapping," *Remote Sensing*, vol. 13, no. 9, Art. no. 9, Jan. 2021, doi: 10.3390/rs13091749.

[6] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing," *ISPRS journal of photogrammetry and remote sensing*, vol. 173, pp. 24–49, 2021.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," p. 10.

[8] M. Amiri, R. Brooks, and H. Rivaz, "Fine Tuning U-Net for Ultrasound Image Segmentation: Which Layers?," in *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data*, Q. Wang, F. Milletari, H. V. Nguyen, S. Albarqouni, M. J. Cardoso, N. Rieke, Z. Xu, K. Kamnitsas, V. Patel, B. Roysam, S. Jiang, K. Zhou, K. Luu, and N. Le, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 235–242. doi: 10.1007/978-3-030-33391-1_27.

[9] T. Falk *et al.*, "U-Net: deep learning for cell counting, detection, and morphometry," *Nature methods*, vol. 16, no. 1, pp. 67–70, 2019.

[10] P. Esser, E. Sutter, and B. Ommer, "A variational u-net for conditional appearance and shape generation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8857–8866.

[11] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition." arXiv, Apr. 10, 2015. doi: 10.48550/arXiv.1409.1556.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition." arXiv, Dec. 10, 2015. doi: 10.48550/arXiv.1512.03385.

[13] M. Rostami, S. Kolouri, E. Eaton, and K. Kim, "Deep Transfer Learning for Few-Shot SAR Image Classification," *Remote Sensing*, vol. 11, no. 11, Art. no. 11, Jan. 2019, doi: 10.3390/rs11111374.

[14] "NAIP Imagery," *temp_FSA_02_Landing_InteriorPages*. https://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-programs/naip-imagery/ (accessed Jul. 05, 2020).

[15] "Central Arizona-Phoenix Long-Term Ecological Research," *Central Arizona–Phoenix Long-Term Ecological Research*. https://sustainability.asu.edu/caplter/ (accessed Apr. 14, 2018).

[16] "ISPRS Benchmark Test on Urban Object Detection and Reconstruction - ISPRS." http://www2.isprs.org/commissions/comm3/wg4/tests.html (accessed Apr. 30, 2020).