Transfer Learning for Construction Robotics: Leveraging Road Construction Data to Enhance Building Site Segmentation

Obiora Odugu[1], Muhammad Tahir Khan[1], Chao Wang[1] and Fellipe Fronchetti[2]

***Abstract*— Gathering data from the construction site environments is crucial for safe and autonomous robot navigation in the construction field. However, collecting large amounts of labeled data directly from construction sites can be difficult due to access restrictions and contractual limitations. This study investigates the use of transfer learning as an alternative to overcome the challenges of model robustness in data scarce construction domain. By training semantic segmentation models using accessible road construction data, we evaluated how road construction knowledge may support image segmentation for building construction. Specifically, we trained and fine-tuned a SegFormer model to identify four key classes for robot navigation on construction sites: equipment, workers, walkable terrain, and risky terrain. Our results indicate that models trained on road construction data consistently outperform those trained on general city data by at least +19% mIoU, particularly in accurately identifying walkable, risky terrains and workers. Overall, the findings demonstrate the potential of transfer learning from road construction to enhance robots comprehension of building construction environments.**

I. INTRODUCTION

For robots to operate safely on construction sites, they must accurately recognize various elements in their environment, such as equipment, workers, safe walking areas, and hazardous terrain [1]. Semantic segmentation provides visual understanding at a pixel-level, enabling robots to make informed decisions about obstacle avoidance, path planning, and safety monitoring [2]. However, good segmentation models typically require extensive and diverse datasets. Gathering data on active construction sites is challenging due to privacy issues, contractual limitations, and safety regulations [3][4]. Consequently, construction researchers often have limited site-specific data, restricting the training, performance, and generalization of segmentation models.

Data scarcity has been shown to slow research progress [5]. In construction, researchers have used Building Information Modeling (BIM) and other 3D tools as a simulation environment for data gathering [6], [7], [8]. However, these environments are majorly indoor scenes and does not account for navigation challenges present in early- and mid-stage construction phases. In this study, we address the issues with model training in data restricted field like building construction. We leverage publicly available road construction data, which share visual features with construction sites, such as equipment, workers, and varied terrain conditions (see Fig. 1). Inspired by successful applications of transfer learning in fields like medical imaging [9], [10] and construction [11], we propose that models trained on road construction images can transfer their learned features to building construction tasks. Specifically, we retrain SegFormer B0 models on road construction data and fine-tune them on smaller subsets of labeled site data. SegFormer was selected as it has the potential for real-time inference in vision systems [12].

Fig. 1. Visualizing the overlap: Shared characteristics between road and site constructions that enable effective transfer learning.

This paper aims to determine if knowledge transferred from road construction can enhance segmentation accuracy on construction sites, thereby providing researchers with a new source domain for building construction scene understanding. Additionally, we perform statistical analysis to validate the difference in the models [13]. The main novelty of this work is to identify and validate road construction as a more effective pretraining source than generic domains for data-scarce building construction tasks.

II. RELATED WORKS

Visual navigation on construction sites enables robots to move safely and autonomously in complex environments. Several recent studies have successfully demonstrated robotic navigation indoors, as well as in structured outdoor areas like city roads [14], [15], [16]. However, navigation during the early and mid-stages of construction remains a significant challenge due to constantly changing layouts, uneven surfaces, and dynamic obstacles such as moving equipment and personnel [1].

Reliable robot navigation in construction environments require an accurate understanding of the surroundings at

This material is based upon work supported by the National Science Foundation under Grant No. 2222881.

[1]Obiora Odugu and Muhammad Khan are Ph.D. Students, Dr. Chao Wang is an Associate Professor at the Bert S. Turner Department of Construction Management, Louisiana State University, Baton Rouge, LA, USA. {oodugu1, mkhan49, chaowang}@lsu.edu

[2]Dr. Fellipe Fronchetti is an Assistant Professor at the Department of Computer Science, Louisiana State University, Baton Rouge, LA, USA. ffronchetti@lsu.edu

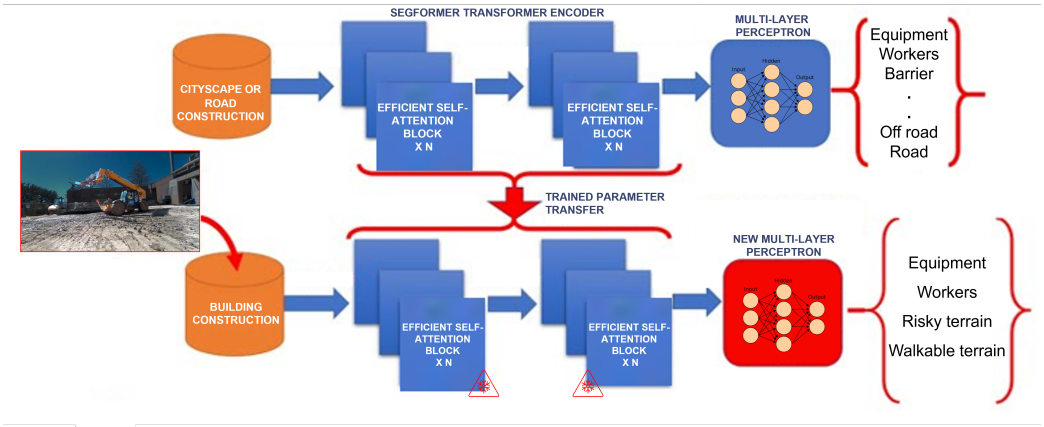

Fig. 2. Overview of transfer learning pipeline. SegFormer is first pretrained separately on source datasets (Cityscape or ROADWork), then fine-tuned on building construction images. Only encoder weights are transferred, while the MLP head is retrained to segment construction-specific classes.

a pixel-level granularity. To achieve this pixel-level understanding, many researchers utilize semantic segmentation — a technique that classifies every pixel of an image into predefined categories [11], [17]. Semantic segmentation has been shown to significantly improve the robot's ability to differentiate between navigable paths, obstacles, and hazardous areas, thereby directly enhancing safety and decision-making [23]. Although standard semantic segmentation methods (e.g., DeepLab, FCN, and SegFormer) have been extensively applied in autonomous driving and indoor robotics, their application to construction sites faces several hurdles. One of which is the limited availability of labeled site-specific data [22].

Data scarcity in construction research is mainly caused by strict privacy rules, contractual constraints, and safety concerns that make it difficult to collect large amounts of site data [3]. As a result, researchers often work with small datasets, which reduces the accuracy and generalizability of machine learning models. To address this, transfer learning has become a common solution. In this approach, models are first trained on large, general-purpose datasets like ImageNet[1] or Cityscapes[2] and then fine-tuned on target-specific data. This method helps improve the performance of segmentation models in the target settings by leveraging features learned from the source domain. Previous studies explored transfer learning in the medical domain and showed that using closely related source domains significantly boosts the performance of pre-trained models [9], [10]. Transfer learning has also been shown to result in negative transfer [24]. For construction sites, few studies [18], [19], [20], [21] have explored transfer learning across related domains. Wang et al. trained models on different datasets (Cityscapes and ImageNet) and fine-tuned them on more than 800 construction site images [11]. Their results showed that Cityscapes pre-training led to better Mean Intersection over Union (mIoU) than ImageNet, with a maximum mIoU of 0.65. However, the study did not explore the potential for more closely related domains such as road construction. Our work addresses this gap by evaluating whether semantic segmentation knowledge from publicly available road construction image data can be effectively transferred to data-limited building construction environments. We aim to determine if this approach can improve segmentation accuracy and enhance robotic navigation in real-world construction site settings.

III. FROM ROAD TO SITE

A. Target Dataset

We collected 6,516 images from an active mid-stage commercial building construction project using built-in RGB cameras of a Boston Dynamics Spot Robot[3]. The dataset includes indoor and outdoor environments to ensure coverage of diverse site conditions. Each collected image was manually annotated into four classes relevant to safe robot navigation: `risky terrain` (i.e., areas hazardous for robot navigation), `walkable terrain` (i.e., safe paths suitable for robot navigation), `workers` (i.e., personnel present on the site), `equipment` (i.e., machinery and tools). These four classes provide pixel-level information for safe and efficient robotic navigation. Annotations were performed using the Roboflow[4] annotation tool. To simulate typical data constraints encountered in construction research, we randomly sampled subsets of varying sizes — 20, 60, 100, 140, 180, 220, 260, 300, 340, 380, 420, 460, and 500 images — from the collected dataset. Each subset size represents realistic dataset sizes that are typically available in construction research [22]. In addition, we curated a fixed test set of approximately 500 images. This test set remained constant throughout all experiments to allow fair and consistent evaluation under different training conditions.

B. Source Datasets

ROADWork[5] and Cityscape were each used to train distinct SegFormer models. ROADWork is an open-source

[1] `https://www.image-net.org/`
[2] `https://www.Cityscapes-dataset.com/`
[3] `https://bostondynamics.com/products/spot/`
[4] `https://roboflow.com/annotate`
[5] `https://www.cs.cmu.edu/~ROADWork/`

dataset that contains image samples collected from road construction scenarios. The dataset consists of 7,416 images captured from urban, suburban, and rural areas across 18 cities in the United States. Images include diverse road construction scenes with the manual semantic mask of essential classes such as roads, sidewalks, off road, cones, barriers, workers, construction vehicles, and equipment. Cityscapes Dataset is a large-scale collection of stereo video sequences captured in urban street environments across 50 cities. It includes high-quality, pixel-level annotations for 5,000 frames. Both dataset contain visual similarities that make them suitable for training and evaluating semantic segmentation models on construction sites.

C. Transfer Learning and Evaluation Method

SegFormer is a transformer-based model known for its strong real time performance in semantic segmentation tasks. In this work, we utilized two versions of the SegFormer-B0 model, one pretrained on Cityscapes[6] and another pretrained on ROADWork datasets. These datasets were selected for their semantic and visual relevance to building construction scenes. As shown in Fig. 2, we then fine-tuned the two models on subsets of labeled building construction images captured with our robot. Given the small data sizes involved, we freeze the pre-trained model backbone to avoid overfitting during fine-tuning and replace the original segmentation head with a new, randomly initialized head tailored for our four construction classes. To ensure our results were reliable, we ran each training setup five times using five random seeds. These seeds control random processes like weight initialization, data shuffling, and dropout. By repeating the experiments with different seeds, we were able to measure how stable and consistent the model's performance was, and improve robustness. Mean Intersection over Union (mIoU) was used to evaluate segmentation performance, measuring

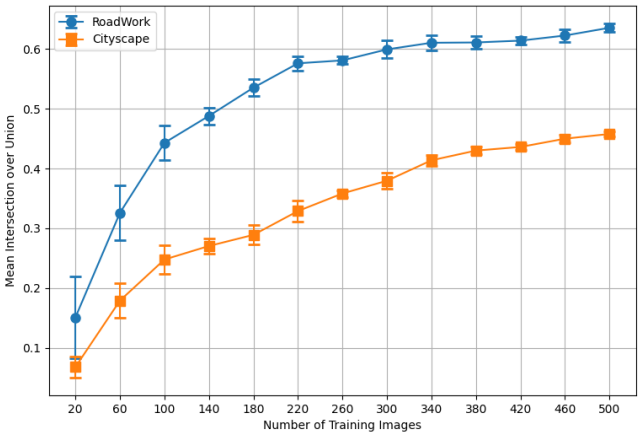

Fig. 3. Mean Intersection over Union (mIoU) comparison across different training sizes for models pretrained on ROADWork and Cityscape. Error bars represent standard deviation across random seeds, capturing variation due to different training initializations.

[6]https://huggingface.co/docs/transformers/en/model_doc/segformer

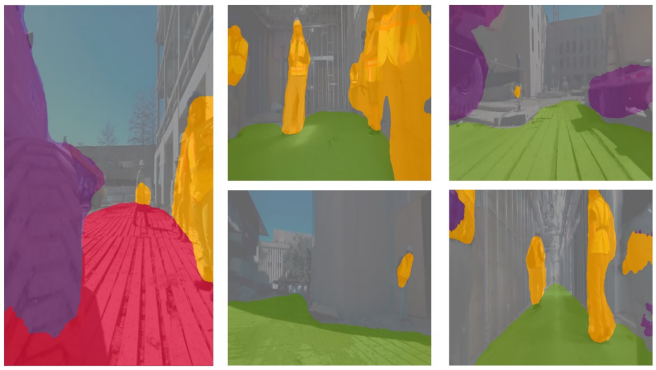

Fig. 4. Semantic segmentation results overlaid on construction site images using the ROADWork model (overall mIoU = 0.6450). Class-wise IoUs: Equipment = 0.3701, Risky terrain = 0.6882, Walkable terrain = 0.6442, Workers = 0.6478. Classes are color-coded as follows: Equipment (purple), Workers (orange), Walkable terrain (green), and Risky terrain (red).

the average overlap between predicted and ground truth masks across all classes. Higher mIoU values indicate better segmentation accuracy. For each dataset size and random seed, the model mIoU scores were recorded. We finetuned each models for 20 epochs with learning rate 3e-5, batch size 2, and five random seeds.

To determine if there is a significant difference between both models and validate performance gains, we concatenated all mIoU scores across dataset sizes and seeds into a single DataFrame and applied a mixed-effects regression model. We do this for both models (ROADWork vs. Cityscape), and set both approaches as a fixed effect, while training size and seed were modeled as random effects. We chose a mixed-effects regression model as it captures both performance trends and variability introduced by experimental conditions.

IV. RESULTS & DISCUSSION

We evaluated both models on a test set of 500 construction site images. The ROADWork-pretrained model consistently outperformed the Cityscape-pretrained model, achieving a mean Intersection-over-Union (mIoU) of 0.6450 at the largest training size (500 images). Fig. 3 shows how performance improved with increasing dataset size for both pretraining approaches. This consistent performance gain shows that visual and semantic similarities between road and building construction environments make ROADWork a more effective pretraining source for site segmentation tasks. Segmented image samples are shown in Fig. 4. Among the predicted classes, the model achieved good IoU scores on walkable terrain (0.6442), risky terrain (0.6882), and workers (0.6478), indicating strong and consistent performance in these categories. The equipment class had a low IoU of 0.3701, although the visual results were acceptable. More analysis is needed to understand why the model struggled, but one possible reason is the lack of different types of equipment in the training data. Additionally, Risky terrain predictions were sometimes confused with walkable areas.

TABLE I

SUMMARY OF MIXED-EFFECTS REGRESSION RESULTS COMPARING MIOU PERFORMANCE BETWEEN CITYSCAPE AND ROADWORK PRETRAINED MODELS.

Parameter	mIoU Estimate	z-value	p-value
Intercept	0.331	20.997	<0.001
approach[T.ROADWork]	0.191	27.963	<0.001

To statistically evaluate the performance difference, we applied a mixed-effects regression model using training size and random seed as random effects, and pretraining approach as a fixed effect. Table I summarizes the key results. In the mixed-effects regression model, the Cityscape model (**intercept**) was chosen as the baseline (mIoU = 0.331). The ROADWork model (**approach[T.ROADWork]**) showed a significant positive effect ($p < 0.001$), with an estimated 0.191 mIoU gain over the Cityscape baseline, reinforcing the observed performance advantage.

V. CONCLUSION

This study explored transfer learning from readily available road construction data to building construction sites to improve semantic segmentation in a data-constrained environment. Our results show that pretraining on the ROADWork dataset led to consistently better performance than Cityscape across all training sizes. The findings demonstrate that ROADWork, with its structural and contextual similarity to construction sites, is a more effective pretraining choice than general-purpose datasets such as Cityscape. Future work will focus on improving segmentation of lower-performing classes, such as equipment, through class balancing, inclusion of diverse construction equipment, expanding evaluation with more test seeds, and benchmarking other models on common datasets.

ACKNOWLEDGMENT

Special thanks to Caleb Taylor (ctay168@lsu.edu), Rohan Durgum (rdurgu@lsu.edu), and Nguyen Vu (nvu22@lsu.edu) for their contributions in labeling the building construction dataset.

REFERENCES

[1] L. Yang and H. Cai, "Enhanced visual SLAM for construction robots by efficient integration of dynamic object segmentation and scene semantics," *Advanced Engineering Informatics*, vol. 59, p. 102313, Jan. 2024.

[2] C. Yu, Z. Liu, X.-J. Liu, and F. Xie, "DS-SLAM: A semantic visual SLAM towards dynamic environments," in *Proc. 2018 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, Oct. 2018, pp. 1168–1174. doi:10.1109/IROS.2018.8593691.

[3] A. Y. Barrera-Animas and J. M. Davila Delgado, "Generating real-world-like labelled synthetic datasets for construction site applications," *Automation in Construction*, vol. 151, p. 104850, Jul. 2023. doi:10.1016/j.autcon.2023.104850.

[4] S. Arabi, A. Haghighat, and A. Sharma, "A deep learning based solution for construction equipment detection: from development to deployment," *arXiv preprint*, arXiv:1904.09021, 2019. doi:10.48550/arXiv.1904.09021.

[5] A. Ghosh, R. Tamburo, S. Zheng, J. R. Alvarez-Padilla, H. Zhu, M. Cardei, N. Dunn, C. Mertz, and S. G. Narasimhan, "ROADWork dataset: Learning to recognize, observe, analyze and drive through work zones," *arXiv preprint*, arXiv:2406.07661, 2024.

[6] E. Araya-Aliaga, E. Atencio, F. Lozano, and J. Lozano-Galant, "Automating dataset generation for object detection in the construction industry with AI and robotic process automation (RPA)," *Buildings*, vol. 15, no. 3, p. 410, 2025. doi:10.3390/buildings15030410.

[7] Y. Hong, S. Park, H. Kim, and H. Kim, "Synthetic data generation using building information models," *Automation in Construction*, vol. 130, p. 103871, 2021. doi:10.1016/j.autcon.2021.103871.

[8] H. Ying, R. Sacks, and A. Degani, "Synthetic image data generation using BIM and computer graphics for building scene understanding," *Automation in Construction*, vol. 154, p. 105016, 2023. doi:10.1016/j.autcon.2023.105016.

[9] M. Romero, Y. Interian, T. Solberg, and G. Valdes, "Targeted transfer learning to improve performance in small medical physics datasets," *Medical Physics*, vol. 47, no. 12, pp. 6394–6403, 2020. doi:10.1002/mp.14507.

[10] L. Alzubaidi, M. Al-Amidie, A. Al-Asadi, A. J. Humaidi, O. Al-Shamma, M. A. Fadhel, J. Zhang, J. Santamaría, and Y. Duan, "Novel transfer learning approach for medical imaging with limited labeled data," *Cancers*, vol. 13, no. 7, p. 1590, 2021.

[11] Z. Wang, Y. Zhang, K. M. Mosalam, Y. Gao, and S.-L. Huang, "Deep semantic segmentation for visual understanding on construction sites," *Computer-Aided Civil and Infrastructure Engineering*, vol. 37, no. 11, pp. 1409–1426, 2022. doi:10.1111/mice.12701.

[12] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," *arXiv preprint*, arXiv:2105.15203, 2021. doi:10.48550/arXiv.2105.15203.

[13] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Scientific Reports*, vol. 14, article no. 6086, Mar. 2024.

[14] H. Bavle, J. L. Sanchez-Lopez, M. Shaheer, J. Civera, and H. Voos, "Situational graphs for robot navigation in structured indoor environments," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10010–10017, 2022.

[15] H. Cai, S. Yuan, X. Li, J. Guo, and J. Liu, "BEV-LIO(LC): BEV image assisted LiDAR-inertial odometry with loop closure," *arXiv preprint*, arXiv:2502.19242, 2025. doi:10.48550/arXiv.2502.19242.

[16] P. Pauwels, R. de Koning, B. Hendrikx, and E. Torta, "Live semantic data from building digital twins for robot navigation: Overview of data transfer methods," *Advanced Engineering Informatics*, vol. 56, p. 101959, Apr. 2023.

[17] K. Asadi, H. Ramshankar, H. Pullagurla, A. Bhandare, S. Shanbhag, P. Mehta, S. Kundu, K. Han, E. Lobaton, and T. Wu, "Vision-based integrated mobile robotic system for real-time applications in construction," *Automation in Construction*, vol. 96, pp. 470–482, Dec. 2018. doi:10.1016/j.autcon.2018.10.009.

[18] X. Yan, H. Zhang, Y. Wu, C. Lin, and S. Liu, "Construction Instance Segmentation (CIS) dataset for deep learning-based computer vision," *Automation in Construction*, vol. 156, p. 105083, Dec. 2023. doi:10.1016/j.autcon.2023.105083.

[19] L. Chen, Y. Wang, and M. F. F. Siu, "Detecting semantic regions of construction site images by transfer learning and saliency computation," *Automation in Construction*, vol. 114, p. 103185, Jun. 2020. doi:10.1016/j.autcon.2020.103185.

[20] E. Mengiste, K. R. Mannem, S. A. Prieto, and B. G. de Soto, "Transfer-learning and texture features for recognition of the conditions of construction materials with small data sets," *Journal of Computing in Civil Engineering*, vol. 38, no. 1, Sep. 2023. doi:10.1061/JCCEE5.CPENG-5478.

[21] N. D. Nath, T. Chaspari, and A. H. Behzadan, "Single- and multi-label classification of construction objects using deep transfer learning methods," *Journal of Information Technology in Construction (ITcon)*, vol. 24, pp. 511–528, Dec. 2019. doi:10.36680/j.itcon.2019.028.

[22] J. M. Davila Delgado and L. Oyedele, "Deep learning with small datasets: Using autoencoders to address limited datasets in construction management," *Applied Soft Computing*, vol. 112, p. 107836, Nov. 2021. doi:10.1016/j.asoc.2021.107836.

[23] T. Guan, D. Kothandaraman, R. Chandra, A. J. Sathyamoorthy, K. Weerakoon, and D. Manocha, "GA-Nav: Efficient terrain segmentation for robot navigation in unstructured outdoor environments," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 8138–8145, Jul. 2022. doi:10.1109/LRA.2022.3187278.

[24] Z. Wang, Z. Dai, B. Poczos, and J. Carbonell, "Characterizing and avoiding negative transfer," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 11293–11302.