A Hybrid Deep Learning Approach for Accurate Urban Image Segmentation Integrating SegNet and DeepLabv3+

Yethiraj N G

Research Scholar, Rayalaseema University, Kurnool, A.P.
Associate Professor of Computer Science, Maharani's Science College for Women, Bangalore, Karnataka.

Dr. Siddappa M

Professor of Computer Science and Engineering, Sri Siddhartha Institute of Technology, Tumkur, Karnataka

Abstract

Semantic image segmentation is crucial for urban environment analysis, enabling extraction of meaningful categories such as roads, buildings, vegetation, and pedestrians. However, segmenting real-world urban street images, particularly Google Street View (GSV) data, poses challenges due to significant differences from publicly available labelled datasets and complexity of scenes. This paper presents a novel hybrid segmentation framework that combines the strengths of SegNet and DeepLabv3+ models. A pre-processing pipeline adapts the Cityscapes dataset to reflect GSV characteristics, followed by independent training of segmentation models. The hybrid prediction map is generated by intelligently merging class-wise results from both models using a weighted sum approach, accounting for model strengths on specific object classes. Extensive experiments confirm significant improvements in Intersection over Union (IOU) scores for various urban classes compared to individual models, along with superior qualitative results confirmed by user studies. This approach offers a practical and effective solution for urban greenness measurement and pedestrian environment analysis, with potential extension to other urban image segmentation applications.

Keywords

Urban image segmentation, Semantic segmentation, Deep learning, SegNet, DeepLabv3+, Google Street View, Cityscapes dataset, Hybrid model, Green coverage measurement.

1. Introduction

Semantic segmentation partitions an image into semantically meaningful regions, associating each pixel with a specific class label. It is particularly important in urban environment analysis, where accurate segmentation of roads, buildings, vegetation, pedestrians, and vehicles contributes to urban planning, pedestrian safety, environmental monitoring, and smart city development.

Deep learning methods, especially convolutional neural networks (CNNs), have significantly enhanced urban image segmentation. Models like SegNet and DeepLabv3+ are widely used due to their robustness and accuracy. However, publicly available urban datasets (e.g., Cityscapes) differ substantially from street-level images captured in various cities (e.g., Google Street View images), which present diverse perspectives, varying image quality, and unrelated scene compositions.

Training models exclusively on public datasets results in suboptimal performance on such real-world data. To bridge this gap, this study proposes a hybrid segmentation framework integrating SegNet and DeepLabv3+, leveraging their complementary strengths while applying data pre-processing to reduce domain differences between datasets.

Key contributions include:

- Dataset pre-processing using sliding window cropping and filtering to align Cityscapes dataset with GSV image domain.
- Independent training and optimization of SegNet and DeepLabv3+ on pre-processed data.
- Class-specific fusion of segmentation maps using a weighted sum strategy favoring model strengths.
- Quantitative evaluation illustrating improved segmentation accuracy (IoU) on Cityscapes test sets.
- Qualitative validation through user studies confirming superior segmentation results on GSV images.

2. Objectives

- To pre-process public urban datasets to emulate Google Street View image characteristics for domain alignment.
- To independently train SegNet and DeepLabv3+ segmentation models on the pre-processed dataset.
- To construct a hybrid segmentation model combining class-wise predictions of both models with weighting based on class accuracy.
- To evaluate segmentation performance using Intersection over Union (IOU) and conduct a user study on GSV images for qualitative validation.
- To demonstrate improved segmentation results applicable for urban greenery assessment and pedestrian environment analysis.

3. Background and Related Work

Semantic segmentation has seen major advances through fully convolutional networks [Long et al., 2015], SegNet [Badrinarayanan et al., 2017], and Deep Lab architectures [Chen et al., 2017]. SegNet employs encoder-decoder structure with pooling indices preserving spatial details, providing accurate object shape detection. DeepLabv3+ utilizes atrous convolutions and spatial pyramid pooling to capture multi-scale context, improving segmentation of large-scale urban structures.

Challenges remain when transferring models trained on benchmarks (Cityscapes) to real-world GSV data due to different perspectives, image resolutions, and scene layout. Prior attempts at domain adaptation and hybrid fusion methods have shown promise in bridging performance gaps [Kim et al., 2023].

4. Proposed Methodology

4.1 Dataset Pre-processing

- **Source Datasets:** Cityscapes dataset (5000 labelled images, 2048×1024 resolution) and Yongsan-gu GSV images (256×256, unlabelled).
- **Resizing & Cropping:** Cityscapes images resized to half and quarter resolution (1024×512, 512×256), cropped using sliding windows with stride 32 pixels to create patches comparable to GSV image size.
- **Filtering:** Patches dominated (>80%) by single classes like sky, road, or building removed to avoid meaningless data.
- Resulting in approximately 2 million patches split 80% training, 20% testing.

4.2 Individual Model Training

- **SegNet:** Symmetric encoder-decoder architecture initialized with VGG16 weights. Uses pooling indices to preserve spatial features. Mini-batch size 128, 300 epochs, learning rate decay every 30 epochs. Weighted cross-entropy loss for class imbalance.
- **DeepLabv3+:** U-Net-like architecture with Inception ResNetv2 backbone. Atrous convolutions and ASPP module capture multi-scale contextual info. Trained similarly on pre-processed patches.

4.3 Hybrid Model Fusion

- **Class-Specific Assignment:**
    - SegNet predictions used for classes with better accuracy, e.g., people and cars.
    - Deep Lab predictions used for classes like road, building, sky, and sidewalk.
- **Conflict Resolution:** Weighted sum of class probabilities from both models used, weights computed based on class IoU performance in validation set using equations:

$$w_{seg}(c) = \alpha_{11} \cdot output_{seg}(c) + \alpha_{12} \cdot output_{dl}(c)$$

$$w_{dl}(c) = \alpha_{21} \cdot output_{seg}(c) + \alpha_{22} \cdot output_{dl}(c)$$

Where α weights normalize contributions relative to model class strengths.

- Plant class combined by union of binary predictions from both models.

5. Algorithm Details

1. Pre-process input images to match domain characteristics.
2. Train SegNet and DeepLabv3+ separately on patches.
3. For each test image, generate segmentation masks from each model.
4. Fuse masks class-wise using weighted sum based on validation IoU scores.
5. Resolve conflicts pixel-wise picking maximum weighted class probability.
6. Output fused hybrid mask.

6. Experiments and Evaluation

6.1 Quantitative Evaluation on Cityscapes Test Set

Class	SegNet IoU	DeepLabv3+ IoU	Hybrid IoU
People	0.699	0.657	0.706
Car	0.681	0.658	0.717
Plant	0.702	0.706	0.715
Sky	0.705	0.792	0.813
Building	0.691	0.813	0.815
Road	0.705	0.796	0.801
Sidewalk	0.721	0.809	0.812
Background	0.760	0.814	0.808
Average	0.708	[illegible]	0.773

6.2 User Study on Google Street View Images

- Participants: 20 graduate-level users.
- Dataset: 30 unlabelled GSV images.
- Task: Rank segmentation quality for SegNet, DeepLabv3+, and Hybrid outputs.
- Result: Hybrid model selected as best in 77% cases, second-best in 20%, worst in 3%.

6.3 Visual Comparison

Figure 4 showcases segmentation masks overlaid on original images, highlighting hybrid model's superior boundary delineation and class accuracy.

7. Discussion

The hybrid model effectively merges complementary strengths of SegNet and DeepLabv3+, overcoming individual model limitations, especially in complex urban street scenarios. Pre-processing aligns external datasets with target domain images, enhancing training effectiveness. Weighted fusion improves class-level performance, balancing recall and precision. User studies substantiate quantitative findings, confirming hybrid approach's superiority.

8. Conclusion and Future Work

This research presents an effective hybrid image segmentation method for urban environment analysis, specifically for unlabelled GSV image segmentation and green ratio measurement. By integrating SegNet and DeepLabv3+ models trained on a pre-processed dataset mimicking GSV characteristics, the hybrid approach improves segmentation IOU and qualitative accuracy, supported by user evaluations. Future work will involve extending model fusion with additional architectures and improving semantic granularity for urban feature analysis, as well as deploying the method in real-world urban planning applications.

9. References

1. Long, J., Shelhamer, E., Darrell, T., "Fully convolutional networks for semantic segmentation," CVPR 2015.
2. Badrinarayanan, V., Kendall, A., Cipolla, R., "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," IEEE TPAMI, 2017.
3. Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A., "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," IEEE TPAMI, 2017.
4. Cordts, M., Omran, M., Ramos, S., et al., "The Cityscapes dataset for semantic urban scene understanding," CVPR 2016.
5. Kim, H., Lee, J.H., Lee, S., "A Hybrid Image Segmentation Method for Accurate Measurement of Urban Environments," Electronics, 2023.
6. Zarrin, I., "Leaf based trees identification using convolutional neural network," IEEE I2CT 2019.
7. Sun, Y., Wang, X., et al., "Using machine learning to examine street green space types at a high spatial resolution: Application in Los Angeles County," Sci. Total Environ., 2021.