SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

DWT-Based ResNet34 Framework for Road Segmentation

Champaka M.D[1*], K B Raja[2], Dr. Mamatha B[3]

[1*]University Visvesvaraya College of Engineering, Bangalore University, Bengaluru, India, champamd88@gmail.com
[2]University Visvesvaraya College of Engineering, Bangalore University, Bengaluru, India, rajakb429@gmail.com
[3]Associate Professor & HOD, Artificial Intelligence and Machine Learning, Machine Learning, Rajarajeshwari College of Engineering, Bangalore, Karnataka, 560074, rcid.org/0009-0001-4168-2146, mamatha.789@gmail.com

ABSTRACT
Road segmentation is a key task in intelligent vehicle systems, autonomous driving, and urban scene understanding. In this work, we propose a DWT-based ResNet-34 framework for road segmentation that uses the Discrete Wavelet Transform (DWT) and ResNet-34 to extract road regions from the Cityscapes database. First, resize and normalize the original Cityscapes images, then apply a one-level 2D-DWT using the Haar wavelet to decompose the image into four sub-bands: LL, LH, HL, and HH. The LL-band is selected as input to the ResNet-34-based segmentation model because it retains important low-frequency structural information while reducing unnecessary high-frequency noise. The ResNet-34 encoder extracts hierarchical spatial features from the LL-band image through convolution, max pooling, and residual blocks. The final encoder feature map is passed to a decoder/upsampling block to generate the final road segmentation mask. The proposed DWT with the ResNet-34 method is evaluated using typical performance metrics, including accuracy, precision, recall, F1 Score, and Intersection over Union (IoU). The model achieved 93.03% accuracy, 79.57% precision, 87.66% recall, an F1-score of 83.42%, and an IoU of 71.55%. The outcomes show that using the LL-band from DWT improves road segmentation performance by enhancing salient structural information and suppressing irrelevant image details. Therefore, the proposed DWT-based ResNet-34 model provides an effective approach for road segmentation in urban scene images.

Keywords: Cityscapes Dataset, Discrete Wavelet Transform, ResNet-34, Road Segmentation, Semantic Segmentation.

Introduction

Image segmentation is a crucial procedure in computer vision and image analysis that divides an image into discrete segments or regions with shared visual attributes, such as color, intensity, or texture. This method converts a complex image into a reduced, organized format that facilitates analysis and interpretation. Each divided zone typically represents a distinct object or component, enabling object recognition, classification, and scene comprehension. Conventional segmentation methods include thresholding [1], edge detection, and region-growing techniques [2]. Accurate road information is essential for both intelligent navigation and urban development. For extracting roads from remote sensing photographs, deep learning-based methods have outperformed traditional road detection systems. However, roads often resemble neighboring objects and face barriers such as plant and structural occlusion, leading to partial road extraction. Road segmentation is an essential computer vision problem, especially for applications like geographic information systems, autonomous driving, and traffic monitoring. This process allows a machine to recognize drivable areas in complex urban and rural environments by classifying individual image pixels as either road or non-road. Early methods used outdated image processing methods, including color thresholding, edge detection, and region growth, but they were not resilient to changes in weather and illumination. Deep Learning (DL) has completely changed road segmentation.

In recent years, DL methods such as Fully Convolutional Networks [3], U-Net [4], and Mask R-CNN [5] have significantly improved segmentation performance, especially in domains like medical imaging and autonomous vehicles. These models enable pixel-level classification, leading to more precise object localization and analysis. As a foundational component in many vision-based systems, image segmentation plays a vibrant character in enabling intelligent decision-making from visual data. Researchers generally evaluate road segmentation models on large annotated datasets, including BDD100K [6], Cityscapes [7], and KITTI [8]. More recently, researchers have investigated hybrid architectures and transformer-based models to improve segmentation quality and capture global context. Intelligent transportation schemes help with accurate road segmentation for safety, decision-making, and environmental perception.

Road segmentation offers pixel-level ground truth for various mobility and mapping applications: autonomous vehicles and ADAS systems integrate segmented road masks with Light Detection and Ranging (LiDAR) an active remote sensing system that is used for vegetation height across wide areas, and radar to formulate safe

trajectories in the absence of lane markings [9]; smart-city platforms utilize these masks with CCTV or aerial imagery to generate congestion heat maps, identify potholes, and optimize signal timing; remote-sensing frameworks transform segmented satellite images into vector road networks for disaster response and map updates [10]; civil engineering teams assess construction progress and surface deterioration on newly installed asphalt [11]; ground robots and drones differentiate navigable paths from agricultural fields or rugged terrain for last-mile delivery and emergency landings [12]; and AR navigation applications accurately position arrows and alerts on the pavement to enhance user engagement [13]. In these areas, precise per-pixel identification of drivable surfaces supports safer planning, improved analytics, and faster decision-making in modern transportation systems.

Contributions: The proposed method applies a one-level 2D-DWT using the Haar wavelet to the resized input image and uses the LL band as input to the ResNet-34 segmentation model. This helps preserve important low-frequency road structure while reducing high-frequency noise and irrelevant details. The model combines wavelet-based feature representation with deep residual feature extraction, improving segmentation performance compared with conventional ResNet-34.

Organization: Section 2 reviews the literature on road segmentation methods currently in use; Section 3 discusses the proposed method; the Performance Evaluation is covered in Section 4, and finally, Section 5 contains the conclusion of the research.

2. Literature Survey

Haoming Bai et al. [14] developed RISENet, a DL model with three primary parts: a mixture of feature dilation-aware decoder, a multi-layer dynamic spatial channel union attention mechanism (MCSA), and a dual-branch combination encoder. To extract basic features and capture fine-grained information, the dual-branch encoder uses dual convolutions and multi-head deep convolutions. By combining local and global information, the feature fusion module improves the model's ability to represent features accurately. MCSA improves road discrimination by capturing long-range dependencies in remote sensing imagery. By continually expanding the receptive field, the dilation-aware decoder reduces fine-detail loss while preserving global characteristics. Bartas Lisauskas and Rytis Maskeliunas [15] proposed a Transformer-based backbone segmentation technique to ensure reliable feature extraction by the encoder module. To merge features efficiently, they also propose a unique decoder module with attention-based fusion methods. Unlike traditional approaches that typically use standard query-based decoders or simple projection layers, this decoder modification preserves fine spatial information and improves global context understanding. Wang et al., [16] presented an approach for extracting road surfaces, centerlines, edges, and intersections using a regularization technique and a multistage feature fusion attention network (MFFANet). There are three parts to MFFANet. The OpenStreetMap (OSM) points, vehicle trajectories, and remote sensing images are adaptively encoded by the complementary feature embedding module (CFEM) to capture specific modal properties. The multistage feature fusion module (MFFM) integrates multisource geospatial information to improve road extraction connectivity and completeness. The multilevel mask generation module (MMGM) improves road segmentation outcomes by using a weighting technique to predict local road segments, road sections, and road network masks simultaneously. We also present a new joint loss function to balance global and local optimization. Road segments are created during the regularization stage using fused hierarchical masks. Centerlines and widths are refined at the skeleton level, and then edges and junctions are extracted, and the segments are smoothly reconstructed. Yang et al., [17] suggested a Frequency Feature Compensation (FFC) technique to enhance segmentation performance in a road segmentation framework based on the Mamba architecture. In particular, they present a progressive FFC approach that divides features into high- and low-frequency components and captures fine-grained information via wavelet decomposition. Using this method, they decomposed multistage features extracted from the Mamba backbone and gradually combined them to capture the crucial information needed for precise route segmentation. Additionally, added a wavelet loss (WL) to improve the model's capability to represent subtle structural changes in the frequency domain. Additionally, a spatial perception Mamba block (SPMB) was created to improve the ability to capture spatial relationships.

Huang et al. [18] presented a DL known as the hybrid segmentation network (HSN-Net), which fluidly combines CNNs and transformers to take advantage of both architectures' benefits. They suggested the road continuity perception module (RCPM) to further improve road continuity. Experiments on the DeepGlobe and CHN6-CUG datasets validate the design decisions, showing that HSN-Net achieves state-of-the-art segmentation performance in road extraction tasks. Qu et al., [19] suggested a road-enhancement and context-aware road extraction network (CARENet). A bidirectional strip feature extraction module (BSFEM) with skip connections is designed to improve the continuity and integrity of extracted roads. This module introduces a new strip-feature extraction method that provides rich, precise road-detail features by preserving road edge information at every scale and passing it to the decoder. A dilated convolution-based Selective Scan module (DBSSM) is then developed to mitigate the adverse effects of complex backdrops and achieve linear attention. The DBSSM comprises multiple context-aware blocks that use 2-D Selective Scan (SS2D) to capture contextual linkages.

Hu et al., [20] introduced DualStrip-Net, a DL framework that can handle sparse annotations and restricted labeled data in road segmentation that is semi-supervised and weakly supervised. Through a dual-stream

architecture that combines a patch-level annotation strategy and strip-based feature learning, DualStrip-Net leverages the natural linear topology of road networks, unlike traditional CNN-based segmentation techniques that do not explicitly model road topology. The framework uses orthogonal strip processing in both horizontal and vertical orientations to collect road features. By using complementary viewpoints, the suggested DualStrip Learning process enables a reliable representation of road structures. Wu et al. [21] suggested using trajectory points to prompt SAM (TPP-SAM), a computerized road extraction system that removes the need for intricate manual prompts by using trajectory points as segmentation prompts for roads in HRSI. TPP-SAM migrates centroid information from building rooftops as a background constraint (BC) and adds road centreline constraint (RCC) and sampling constraint (SC) layers to choose trustworthy trajectory prompt points. This three-layer constraint system integrates the vision foundation model with real-world crowdsourced data. Additionally, the method accounts for prompt points' properties and relative geographic coordinates by employing map-matching and recoding procedures to manage their data structure. Road semantic masks are then created by embedding the analyzed data into the prompt encoder.

Wang et al., [22] introduced SwinURNet, a transformer-convolutional neural network (CNN) architecture for segmenting point clouds in unstructured environments in real time. The first step is to use spherical projection to project the point cloud onto a range image. A lightweight network based on ResNet34 is then used to encode abstract features. The Non-Square Swin Transformer (NSST) captures high-resolution transverse characteristics and decodes the data. It introduces a multidimensional information fusion module (MDIFM) to balance the semantic discrepancies between transformer attention maps and CNN feature maps. The network is trained with a multitask loss function (MTLF) introduced at the network end, comprising boundary, weighted cross-entropy, and Lovász–Softmax losses. Li and Zhou [23] presented the multiscale adaptive stream network (MASNet), a unique dual-modal semantic segmentation network for road environments that uses optical image features to enhance LiDAR point cloud features. An image information mask matrix is used to filter out effective 2-D point clouds and project 3-D point clouds onto the image coordinate system to achieve feature homogeneity. SalsaNext and ResNet34 encode features concurrently in the backbones of the LiDAR and image branches, respectively. Additionally, a module based on a point attention technique and residual mapping is created to unidirectionally transfer image features to the LiDAR branch across several encoding stages. Between the LiDAR branch's encoder and decoder, the multi-expansion-rate, multi-cascaded atrous spatial convolutional pooling pyramid (MmASPP) structure is designed. It consists of parallel bottleneck structures with varying dilation rates that efficiently capture multiscale contextual information, improve pixel correlations, and enable dense pixel feature aggregation. Furthermore, it interacts with and propagates projected label distributions at the network's output layer and provides an adaptive synergy-difference loss function based on cross-entropy that quantifies semantic discrepancies during cross-modal operations. To emphasize the perceptual benefits of particular modalities in application contexts, the modal computation weights in the loss function are allocated according to modal differences. Xiao et al. [24] introduced a road extraction network called MDC-Net (Multiscale Convolution Network), which extracts multiscale features while preserving information. This model uses an encoder-decoder network structure to extract information about road features. The network uses a pre-trained deep residual network, ResNet34, to shorten training convergence time. The road's multi-scale features are extracted by adding an ASPP structure based on dilated convolutions between the encoder and decoder.

3. Proposed Model

The proposed DWT-ResNet-34 road segmentation model shown in Figure 1 uses Cityscapes images. The original Cityscapes image (2048×1024×3) is first resized to 1024×1024×3. A one-level 2D DWT is then applied, producing four sub-bands: LL, LH, HL, and HH, each of size 512×512×3. We select the LL band as input to the ResNet-34 encoder because it contains important low-frequency structural information. The ResNet-34 encoder reduces the LL-band input from 512×512×3 to a high-level feature map of size 16×16×512. This feature map is then passed through a decoder/upsampling network to recover the spatial resolution and generate the final road segmentation mask of size 512×512×1. This output mask classifies each pixel as either road or non-road.

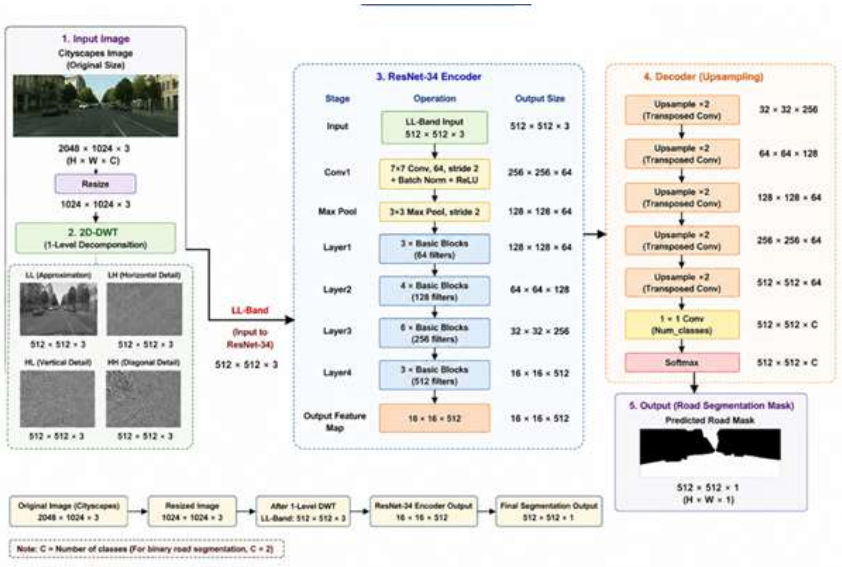

Fig 1. Proposed road segmentation method

3.1. Cityscapes dataset

The dataset consists of 5000 high-resolution images of urban environments, precisely labeled at the pixel level, taken from 50 locations across Germany. Each pixel in every image in the collection is characterized into one of 19 semantic classes, representing standard elements of the urban roadway environment, such as roadways, automobiles, pedestrians, walkways, and traffic signs [25, 26]. The original Cityscapes dataset images are uniformly sized, generally 2048×1024. Figure 2 shows a sample image. Although Cityscapes is not a satellite image dataset, it is highly suitable for urban road scene segmentation using vehicle-mounted camera images.

Fig 2. Selected photos from the Cityscapes collection that highlight the variety of urban environments seen in German cities [26].

3.2 Image preprocessing

The Cityscapes images are preprocessed to suit wavelet decomposition and ResNet-34-based road segmentation. The original Cityscapes RGB image (2048×1024×3) is first loaded from the dataset. It is then resized to 1024×1024×3 to obtain a uniform square input image. After resizing, normalization converts pixel values from the range 000–255 to 000–111. This helps reduce pixel-value variation and progresses the stability of DWT decomposition. The final preprocessed image, of size 1024×1024×3, is then used as input to a one-level 2D-DWT to produce the LL, LH, HL, and HH sub-bands.

3.3 Discrete Wavelet Transformation

The Haar wavelet (db1) decomposes the image into four frequency sub-bands: LL, LH, HL, and HH, each of size 512×512×3 as shown in Figure 3. The LL band represents the low-frequency approximation component and contains the image's main structural information, such as the road region and the boundaries of large objects. The LH, HL, and HH bands contain horizontal, vertical, and diagonal high-frequency details, respectively. In the proposed method, we select the LL band as input to ResNet-34 because it preserves important road structure while reducing noise and unnecessary texture details. This helps the segmentation model focus on meaningful road features and expands the quality of the predicted road mask.

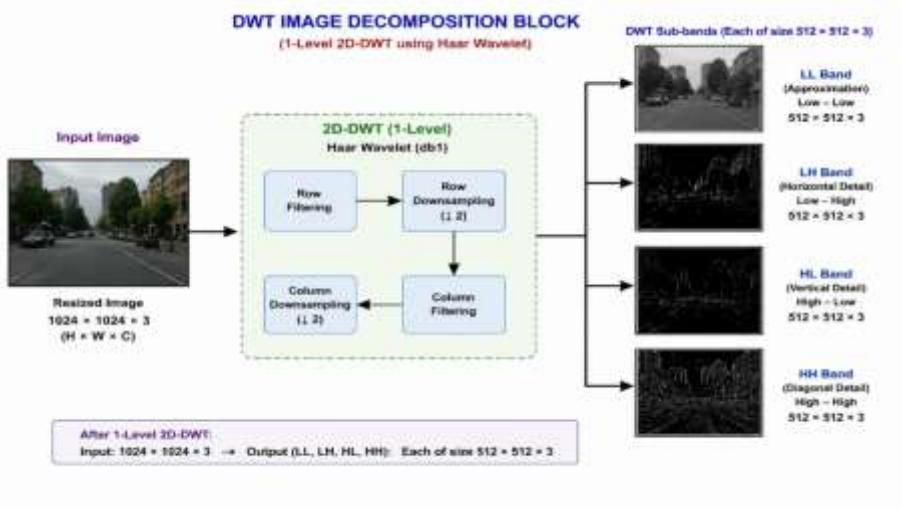

Fig 3. DWT decomposition

3.2 ResNet 34

DL is a subset of machine learning that uses models based on multi-layered neural networks (NNs). ResNet-34, a residual network, serves as a versatile and efficient backbone for many computer vision tasks, leveraging hierarchical feature extraction to transform raw images into rich semantic representations. ResNet-34 is a deep residual NN architecture [27, 28]. It is a 34-layer DCNN with residual connections, built primarily from 3x3 convolutional filters, and it overcomes vanishing gradient problems while maintaining computational efficiency, making it ideal for both research and production environments. The ResNet34 architecture is structured into three main chunks, viz., input, hidden, and output blocks, with the input image typically preprocessed to 512 × 512 pixels. Figure 4 shows the ResNet-34 Architecture.

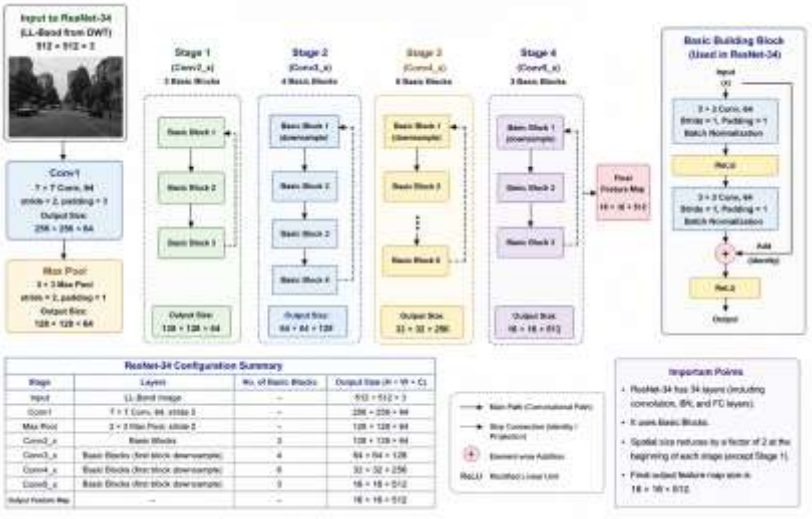

Fig 4: ResNet-34 Architecture [29]

In the projected DWT-based ResNet-34 road segmentation model, the LL-band image (512×512×3) from the DWT of the original image is given to the ResNet-34 encoder. The first convolution layer applies a 7×7 convolution with 64 filters and a stride of 2, reducing the feature map to 256×256×64. A max pooling layer further reduces it to 128×128×64. The feature map is then passed through four residual stages containing 3, 4, 6, and 3 basic blocks, respectively. These stages produce feature maps of size 128×128×64, 64×64×128, 32×32×256, and 16×16×512. The final encoder output contains high-level semantic information required for road segmentation. This output is given to a decoder/up-sampling block to produce the final road segmentation mask of size 512×512×1.

4. Performance Evaluation

4.1 Definitions

4.1.1 Accuracy

It is a performance measure that shows how many predictions a model makes correctly out of the total number of predictions. The model accuracy is the ratio of correctly classified pixels to the total number of pixels in the

image and is computed pixel-wise using Equation (1)

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{----------(1)}$$

Where:
TP = True Positives = Road pixels correctly predicted as road
TN = True Negatives = Non-road pixels correctly predicted as non-road
FP = False Positives = Non-road pixels wrongly predicted as road
FN = False Negatives = Road pixels wrongly predicted as non-road

4.1.2 Precision

The ratio of correctly predicted road pixels to the total number of pixels predicted as road, as given in Equation (2). In road segmentation, high precision means the model produces fewer false road pixels.

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{---------- (2)}$$

4.1.3 Recall

The ratio of correctly predicted road pixels to the total number of actual road pixels, as given in Equation (3). In road segmentation, high recall means the model misses fewer road pixels.

$$\text{Recall} \frac{TP}{TP+FN} \quad \text{----------(3)}$$

4.1.4 F1-Score

The harmonic mean of precision and recall. It yields a single performance value by combining the two metrics, as given in Equation (4).

$$F1 = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad \text{----------(4)}$$

4.1.5 Intersection over Union (IoU)

The ratio of correctly predicted road pixels to the union of predicted road pixels and actual road pixels, as given in Equation (5). In road segmentation, IoU is one of the most important metrics because it quantifies how well the predicted road region overlaps with the actual one.

$$\text{IoU} = \frac{TP}{TP+FP+FN} \quad \text{----------(5)}$$

4.2 Performance Parameter Calculations Using ResNet-34

The road segmentation model achieved 92.31% accuracy, 78.90% precision, 86.80% recall, 82.66% F1-score, and 70.42% IoU, as shown in Table 1. The high accuracy indicates that the model correctly classifies most road and non-road pixels. Recall is higher than precision, signifying that the model detects most road pixels but still produces some false road predictions. The F1-score of 82.66% indicates a good balance between precision and recall, while the IoU of 70.42% shows good overlap between the predicted road mask and the ground-truth road mask. Overall, the model achieves strong road segmentation performance after 100 epochs, but additional improvements are possible by reducing false positives and improving road boundary detection.

Table 1: Performance metrics using ResNet-34

Metric	Accuracy	Precision	Recall	F1-score	IoU
Result	92.31%	78.90%	86.80%	82.66%	70.42%

4.3 Performance Parameter Calculations Using DWT with ResNet-34

The ResNet-34 with DWT-based road segmentation model achieved an accuracy of 93.03%, precision of 79.57%, recall of 87.66%, F1-score of 83.42%, and IoU of 71.55% with 100 epochs is given in Table 2. The higher recall specifies that the model successfully detects most road pixels, while the lower precision shows that some false road detections still occur. The F1-score demonstrates a good balance between precision and recall, and the IoU value confirms a good overlap among the predicted road mask and the ground-truth mask. Overall, applying DWT before ResNet-34 helps extract useful low-frequency structural features from the LL band, thereby improving road segmentation performance.

Table 2: Performance metrics using DWT with ResNet-34

Metric	Accuracy	Precision	Recall	F1-score	IoU
Result	93.03%	79.57%	87.66%	83.42%	71.55%

4.4 Comparison of ResNet-34 and DWT with ResNet-34

The conventional ResNet-34 model and the ResNet-34 with DWT model are evaluated using accuracy,

precision, recall, F1-score, and IoU, as tabulated in Table 3. The ResNet-34 with DWT approach achieved better performance, with 93.03% accuracy, 79.57% precision, 87.66% recall, an F1-score of 83.42%, and an IoU of 71.55%. Compared with the conventional ResNet-34 model, the DWT-based method improved accuracy by 0.72%, precision by 0.67%, recall by 0.86%, F1-score by 0.76%, and IoU by 1.13%. This improvement indicates that the LL-band from 2D-DWT helps the model capture important structural information, thereby improving road segmentation performance.

Table 3: Comparison of ResNet-34 and DWT with ResNet-34

Sl.No.	Metric	ResNet-34	ResNet-34 with DWT	Improvement
1	Accuracy	92.31%	93.03%	+0.72%
2	Precision	78.90%	79.57%	+0.67%
3	Recall	86.80%	87.66%	+0.86%
4	F1-score	82.66%	83.42%	+0.76%
5	IoU	70.42%	71.55%	+1.13%

ResNet-34 with the DWT method performs slightly better than the standard ResNet-34. This is because the LL-band obtained from 2D-DWT contains important low-frequency structural information of the image. Roads are typically continuous structures, so the LL-band helps the model focus on major road regions while reducing unwanted noise.

4.5 Existing Techniques Comparison with Projected Method using Cityscapes dataset

Table 4 compares the performance of the projected DWT with ResNet-34 against existing Cityscapes-based semantic segmentation techniques. Wu et al. [30] achieved 64.8% mIoU using context-guided semantic segmentation, while Zhao et al., [31] obtained 69.5% mIoU using the Image Cascade Network. Yu et al., [32] achieved 68.4% mIoU using a spatial path and context path-based segmentation model, and Romera et al., [33] reported 69.7% IoU using an efficient residual factorized convolutional network. The proposed method achieved 71.55% IoU, which is higher than the listed existing techniques. This shows that integrating DWT with ResNet-34 improves road segmentation performance by preserving important structural information from the LL band and increasing the overlap between the predicted road mask and the ground-truth mask.

Table 4: Projected technique compared with current approaches

Sl No	Authors	Techniques	Result
1	Wu et al., [30]	context-guided semantic segmentation	mIoU 64.8%
2	Zhao et al., [31]	Image Cascade Network.	mIoU 69.5%
3	Yu et al., [32]	A spatial path to preserve details and a context path to capture high-level semantic information.	mIoU 68.4%
4	Romera et al., [33]	Efficient residual factorized convolutional network	IoU 69.7%
3	Proposed Method	DWT with ResNet-34	IoU 71.55%

5 Conclusion

In this work, we propose a DWT-based ResNet-34 framework for road segmentation to extract road regions from the Cityscapes database. The original Cityscapes images are resized and preprocessed before applying one-level 2D-DWT using the Haar wavelet. The DWT operation decomposes the image into four sub-bands: LL, LH, HL, and HH. We selected the LL-band as input to the ResNet-34-based segmentation model because it retains important low-frequency structural information while reducing unnecessary high-frequency details. The ResNet-34 encoder extracted hierarchical features from the LL-band image, and the decoder generated the final binary road segmentation mask. The results indicate that integrating DWT with ResNet-34 helps the model focus on important road structures, improves segmentation quality, and reduces the influence of noise and irrelevant image details. Therefore, the projected technique is effective for road segmentation in urban scene images. In future work, the proposed method can be improved by using not only the LL-band but also high-frequency sub-bands such as LH, HL, and HH to preserve road boundary and edge details. Multi-band fusion of wavelet features may further improve the detection of thin roads, lane boundaries, and complex road structures.

References

1. N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62-66, Jan. 1979, doi: 10.1109/TSMC.1979.4310076.
2. Gonzalez, R. C., & Woods, R. E. (2018). Digital Image Processing (4th ed.). Pearson.
3. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," IEEE

Conference on Computer Vision and Pattern Recognition (CVPR), 2015, Boston, MA, USA, 2015, pp. 3431-3440, doi: 10.1109/CVPR.2015.7298965.
4. Ronneberger O, Fischer P, and BroxT, "U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention," (MICCAI). Arxiv, 2015, https://arxiv.org/abs/1505.04597
5. K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," IEEE International Conference on Computer Vision (ICCV), Venice, Italy, pp. 2980-2988, 2017, doi: 10.1109/ICCV.2017.322.
6. Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell, "BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 2633-2642, doi: 10.1109/CVPR42600.2020.00271.
7. Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding," IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 3213-3223, doi: 10.1109/CVPR.2016.350.
8. A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 2012, pp. 3354-3361, 2012, doi: 10.1109/CVPR.2012.6248074.
9. L. -C. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 834-848, 1 April 2018, doi: 10.1109/TPAMI.2017.2699184.
10. F. Bastani et al., "RoadTracer: Automatic Extraction of Road Networks from Aerial Images," IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 4720-4728, 2018, doi: 10.1109/CVPR.2018.00496.
11. J. Zhang, S. Sun, W. Song, Y. Li, and Q. Teng, "Automated Pavement Distress Detection Based on Convolutional Neural Network," IEEE Access, vol. 12, pp. 105055-105068, 2024, doi: 10.1109/ACCESS.2024.3434569.
12. V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481-2495, 1 Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.
13. L. Chen, P. Wu, K. Chitta, B. Jaeger, A. Geiger, and H. Li, "End-to-End Autonomous Driving: Challenges and Frontiers," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 46, no. 12, pp. 10164-10183, Dec. 2024, doi: 10.1109/TPAMI.2024.3435937.
14. Haoming Bai, Chao Ren, Zhenzhong Huang, and Yao Gu, "A Dynamic Attention Mechanism for Road Extraction from High-Resolution Remote Sensing Imagery Using Feature Fusion," Scientific Reports, vol. 15, Art. no. 17556, May 2025, doi: 10.1038/s41598-025-02267-6.
15. Bartas Lisauskas and Rytis Maskeliunas, "Efficient Transformer-Based Road Scene Segmentation Approach with Attention-Guided Decoding for Memory-Constrained Systems," Journal of Machines, vol 13, issue 6, pp 1-21, 2025, doi.org/10.3390/machines13060466.
16. Z. Wang, L. Xiang, Z. Liu, and Z. Wang, "Automatic Extraction of Roads from Multisource Geospatial Data Using Fusion Attention Network and Regularization Algorithm," IEEE Transactions on Geoscience and Remote Sensing, vol. 63, pp. 1-17, 2025, Art no. 5603517, doi: 10.1109/TGRS.2024.3520610.
17. J. Yang, T. Liu, L. Zhang and Y. Zhang, "Effective Road Segmentation with Selective State-Space Model and Frequency Feature Compensation," IEEE Transactions on Geoscience and Remote Sensing, vol. 63, pp. 1-13, 2025, Art no. 5603813, doi: 10.1109/TGRS.2024.3521483.
18. B. Huang, Y. Lu, R. Yang, Y. Tao, S. Wang and Y. Shi, "HSN-Net: A Hybrid Segmentation Neural Network for High-Resolution Road Extraction," IEEE Geoscience and Remote Sensing Letters, vol. 22, pp. 1-5, 2025, Art no. 2502105, doi: 10.1109/LGRS.2025.3558511.
19. Z. Qu, M. Li, and Z. Chen, "CARENet: Satellite Imagery Road Extraction via Context-Aware and Road Enhancement," IEEE Geoscience and Remote Sensing Letters, vol. 22, pp. 1-5, 2025, Art no. 6005005, doi: 10.1109/LGRS.2025.3545881.
20. J. Hu, Q. Li and Q. Wang, "DualStrip-Net: A Strip-Based Unified Framework for Weakly- and Semi-Supervised Road Segmentation from Satellite Images," IEEE Transactions on Geoscience and Remote Sensing, vol. 63, pp. 1-14, 2025, Art no. 5617514, doi: 10.1109/TGRS.2025.3555211.
21. T. Wu, Y. Hu, J. Qin, X. Lin, and Y. Wan, "TPP-SAM: A Trajectory Point Prompting Segment Anything Model for Zero-Shot Road Extraction From High-Resolution Remote Sensing Imagery," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, pp. 8845-8864, 2025, doi: 10.1109/JSTARS.2025.3548688.
22. Z. Wang, Z. Liao, B. Zhou, G. Yu, and W. Luo, "SwinURNet: Hybrid Transformer-CNN Architecture for Real-Time Unstructured Road Segmentation," IEEE Transactions on Instrumentation and Measurement, vol. 73, pp. 1-16, 2024, Art no. 5035816, doi: 10.1109/TIM.2024.3470042.

23. X. Li and J. Zhou, "MASNet: Road Semantic Segmentation Based on Multiscale Modality Fusion Perception," IEEE Transactions on Instrumentation and Measurement, vol. 73, pp. 1-13, 2024, Art no. 2500713, doi: 10.1109/TIM.2023.3318691.
24. P. Xiao, D. Yang, Y. Li, and J. Zhao, "Road Information Extraction from Remote Sensing Images Based on Fully Convolutional Network," IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Dali, China, 2022, pp. 1389-1394, doi: 10.1109/ICCASIT55263.2022.9986694. https://www.cityscapes-dataset.com/
25. M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, doi: 10.1109/CVPR.2016.350.
26. Kaiming He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, 2019, doi: 10.1109/CVPR.2016.90.
27. V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," IEEE Transactions Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481–2495, December 2017, doi: 10.1109/TPAMI.2016.2644615.
28. Andy T sai and PaulKKleinman, "Machine Learning to Identify Distal Tibial Classic Metaphyseal Lesions of Infant Abuse: A Pilot Study," Springer Nature Link Pediatric Radiology Journal, vol 52, pp 1095-1103, 2022, doi.org/10.1007/s00247-022-05287-w.
29. T. Wu, S. Tang, R. Zhang, J. Cao, and Y. Zhang, "CGNet: A Light-Weight Context Guided Network for Semantic Segmentation," IEEE Transactions on Image Processing, vol. 30, pp. 1169-1179, 2021, doi: 10.1109/TIP.2020.3042065.
30. Zhao H, Qi X, Shen X, Shi J, and Jia J, "ICNet for Real-Time Semantic Segmentation on High-Resolution Images," In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018. Lecture Notes in Computer Science (), vol 11207. Springer, Cham. https://doi.org/10.1007/978-3-030-01219-9_25.
31. Yu C, Wang J, Peng C, Gao C, Yu G, and Sang N, "BiSeNet: Bilateral Segmentation Network for Real-Time Semantic Segmentation," In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science (), vol 11217. Springer, Cham. https://doi.org/10.1007/978-3-030-01261-8_20.
32. E. Romera, J. M. Álvarez, L. M. Bergasa and R. Arroyo, "ERFNet: Efficient Residual Factorized ConvNet for Real-Time Semantic Segmentation," IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 1, pp. 263-272, Jan. 2018, doi: 10.1109/TITS.2017.2750080.