Attention Based Coupled Framework for Road and Pothole Segmentation

Shaik Masihullah, Dr. Ritu Garg, Dr. Prerana Mukherjee and Dr. Anupama Ray

Email: masihullah.s18@iiits.in , prerana@jnu.ac.in, ritu.garg@intel.com, anupamar@in.ibm.com

Motivation

Automated inspection and assessment of functional and physical conditions of the road is important for the purpose of maintaining roads and to ensure vehicle safety. In recent years, automated driving systems (ADS) and advanced driver-assistance systems (ADAS) have gained popularity in the research community. Road segmentation i.e. demarcating drivable region is crucial task in any ADS/ADAS.

Challenges

1. Difficult to segment the drivable path in case of unstructured scenarios, where the roads are neither well maintained nor well-marked.
2. Monocular vision based road and pothole segmentation is challenging in the context of unstructured road environments because of illumination variations, occlusion due to varying object or background appearance, and poor traffic conditions.
3. Need of an end-to-end framework segmenting both roads and potholes in unstructured driving environments such as Indian roads.

Contribution

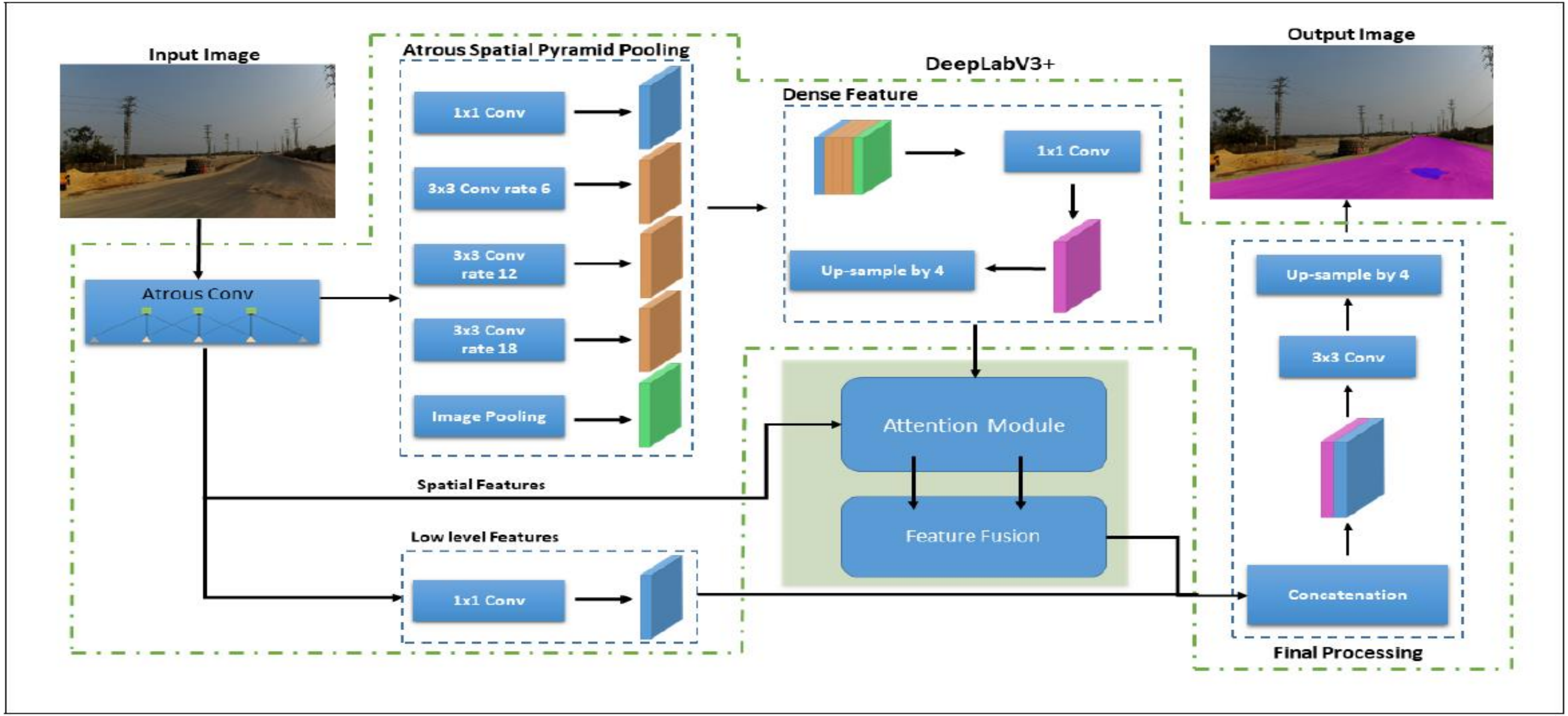
1) Propose coupled framework for road and pothole segmentation, particularly in case of unstructured environments.
2) Extend DeepLabv3+ architecture by incorporating attention based refinement with feature fusion to improve the segmentation.
3) Explored few-shot learning approach for pothole detection to leverage accuracy improvement with fewer training samples on IDD [1] dataset.

Proposed Methodology

A. Semantic Segmentation

We present the solution using DeepLabv3+ as the backbone framework. DeepLabv3+ framework captures rich spatial and contextual features. The network consists of two major components: Atrous convolution also called as dilated convolution and atrous spatial pyramid pooling(ASPP). We evaluated DeepLabv3+ using MobileNetV2 as the backbone network. This model is known to have reduced computational complexity which is an essential requirement for ADAS/ADS like use-case.

Proposed Framework: Architecture

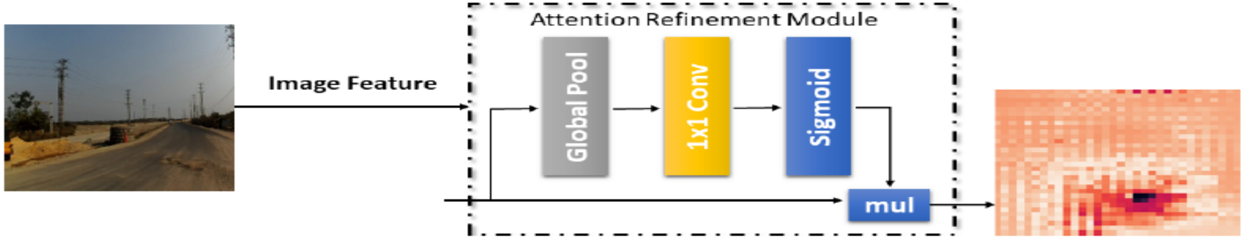


Proposed Methodology

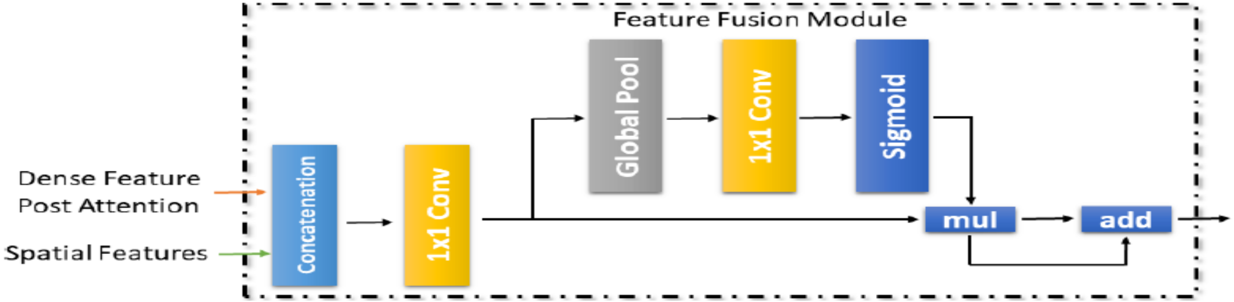
B. Attention Based Refinement with Feature Fusion (AFF)

In the context of semantic segmentation, multi-scale features have proved to be useful. In DeepLabv3+, the atrous convolution encodes rich spatial information, while ASPP provides global to local context aggregation feature learned at multiple scales. We propose to incorporate an attention guided feature learning module to refine the features at each stage. The attention module uses global average pooling to lift the global-context and spatial information to compute a attention vector resulting in a more powerful representation. Next, we utilize global pooling on the concatenated features to obtain a feature vector and compute weights. The weights learnt can assist in feature selection and combination.

Attention Based Refinement With Feature Fusion



(a) Attention Based Refinement Module, Right:corresponding Attention Map.



(b) Feature Fusion Module.

Evaluation Metrics

- Mean IoU is defined as the average IoU or Jaccard Index over all classes. It is defined as the area of intersection between the predicted segmentation regions and the ground-truth, divided by the area of union between the predicted and the ground truth:

$$mIoU = mean(\frac{A \cap B}{A \cup B})$$

- Precision / Recall / F1 score are popular metrics for reporting the accuracy of semantic image segmentation models. Recall and Precision can be defined for each class, as follows:

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

F1-score indicates the harmonic mean of precision and recall: $F1score = \frac{2 \times Precision \times Recall}{Precision + Recall}$

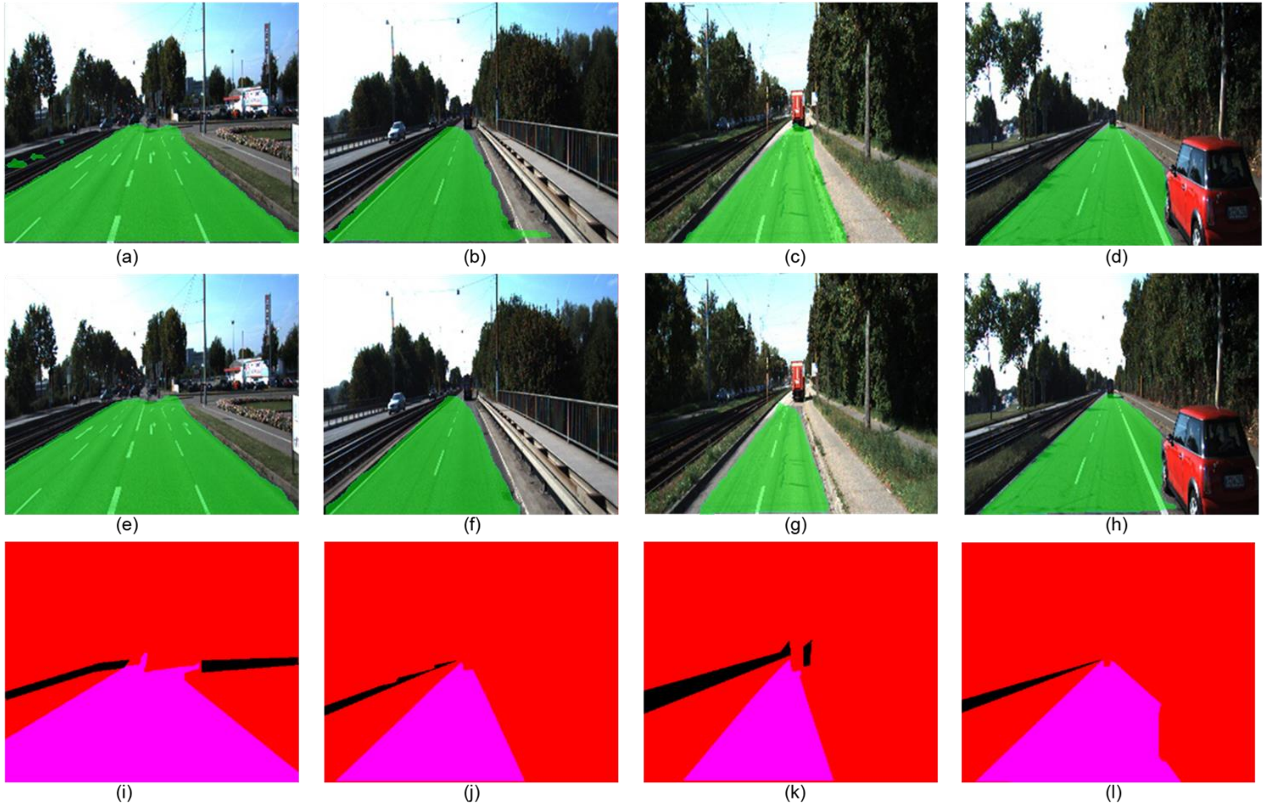
Datasets

- Road segmentation
  - KITTI Visual Benchmark Suite consists of 289 training and 290 test images.
  - Indian Driving Dataset (IDD) consists of 6993 training and 981 test images.
- Pothole segmentation
  - IDD has images depicting poor road conditions in the form of potholes, but no annotations are available.
  - So, for experiments on pothole detection, we have annotated 237 images from IDD dataset with varying pothole sizes using LabelMe toolbox.

Quantitative results on road segmentation (KITTI dataset)

Method	mIoU	MaxF1	Precision	Recall	Accuracy	FPR	FNR	Runtime
Results on complete dataset								
SOTA results	-	97.05%	97.18%	96.92%	-	1.28	3.08	0.16s
SOTA results	-	95.78%	94.92%	96.66%	-	2.36	3.34	0.16s
Proposed Model (w/o Attention Module)	93.2%	94.63%	94%	95.17%	98.0%	1.29	4.05	0.017s
Proposed Model (With Attention Module)	93.6%	95.21%	94.02%	95.95%	98.1%	1.29	4.84	0.030s
Results on UM, UMM UU from KITTI dataset								
UM: Proposed Model (w/o Attention Module)	89.84%	96.52%	97.22%	95.53%	-	0.47	4.48	-
UM: Proposed Model (with Attention Module)	89.95%	95.75%	92.91%	98.45%	-	1.40	1.55	-
UMM: Proposed Model (w/o Attention Module)	90.63%	97.32%	97.73%	96.64%	-	0.61	3.37	-
UMM: Proposed Model (with Attention Module)	91.04%	96.59%	94.38%	98.62%	-	1.735	1.38	-
UU: Proposed Model (w/o Attention Module)	93.0%	95.46%	96.87%	93.80%	-	0.423	6.22	-
UU: Proposed Model (with Attention Module)	92.80%	95.77%	93.92%	97.37%	-	0.93	2.63	-

Road Segmentation results on KITTI



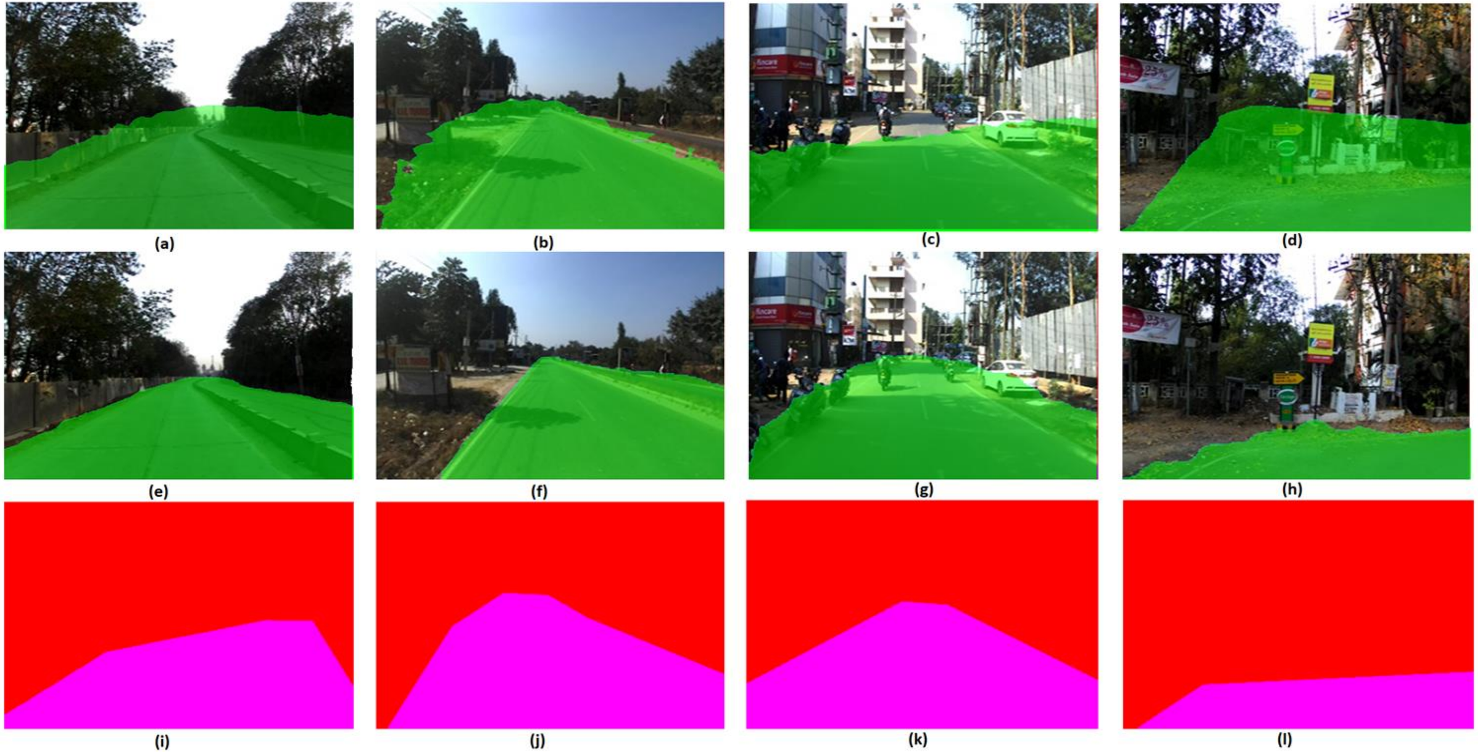
(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l)

(a)-(d) Road segmentation results of DeepLabv3+ without attention, (e)-(h) Road segmentation results of DeepLabv3+ with attention, (i)-(l) Ground-truth labeling for roads Vs background (pink: road, red : background).

Quantitative results on road segmentation(IDD dataset)

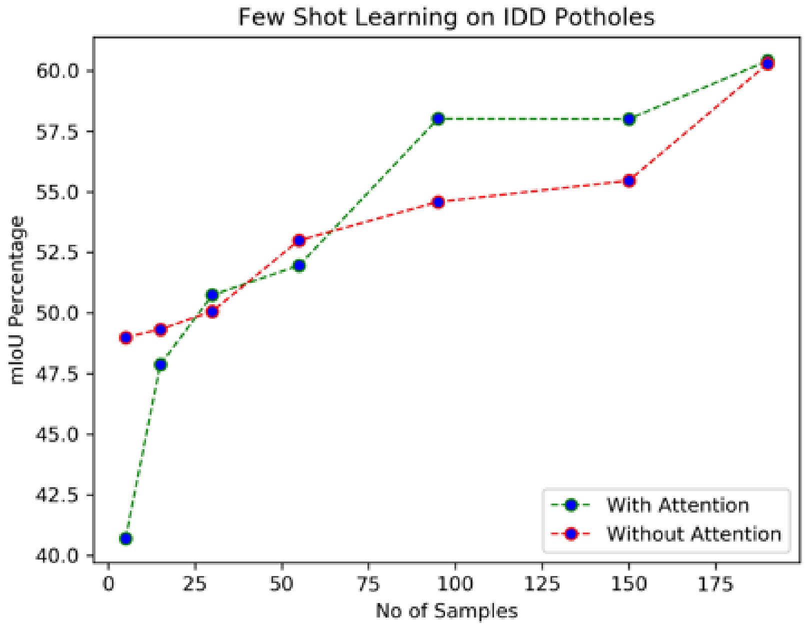
Method	mIoU%
Proposed	98.42
BiSeNet w/o CRF	96.54
BiSeNet with CRF	96.26
Baseline Results (Train: IDD; Test: IDD)	92
Baseline Results (Train: BD; Test: IDD)	83

Road Segmentation results on IDD



(a)-(d) Road segmentation results of DeepLabv3+ without attention, (e)-(h) Road segmentation results of DeepLabv3+ with attention, (i)-(l) Ground-truth labeling for roads Vs background, (pink: road, red : background)

Quantitative results on Pothole Detection (IDD dataset) with Few Shot Learning



Method	mIoU%
Proposed Model (Without Attention Module)	70.81
Proposed Model (With Attention Module)	73.83

Pothole Detection Results on IDD showing variability in sizes, depth, shadow

Conclusion

- The experiments show that the overall mIoU of the proposed attention framework is 98.42% for road segmentation and 73.83% for pothole segmentation on IDD dataset. The road segmentation on KITTI dataset is evaluated in terms of F1-score which is 95.21% with reduction in run-time by 81% in comparison.
- The future extension of this work is directed towards fine-tuning the model for tackling road-scene images under different weather (such as rain,fog, snow) and light (like road condition assessment during night) conditions.
- Another interesting area to explore would be semi-supervised learning technique to improve classification performance when there is a significant difference between the distributions of the unlabeled and the labeled data.

Thank You

Contact: masihulla17@gmail.com , prerana@jnu.ac.in
Code and model weights: https://github.com/Masihullah17/road-and-pothole-segmentation