AutoNUE Challenge: Enhancing Semantic Segmentation by Enabling Expertise between Confusing Classes

Aditya Ganeshan, Gaurav Singh Rajput, Rahul MV, Vivek BS, R. Venkatesh Babu
Video Analytics Lab, Indian Institute of Science, Bangalore, India

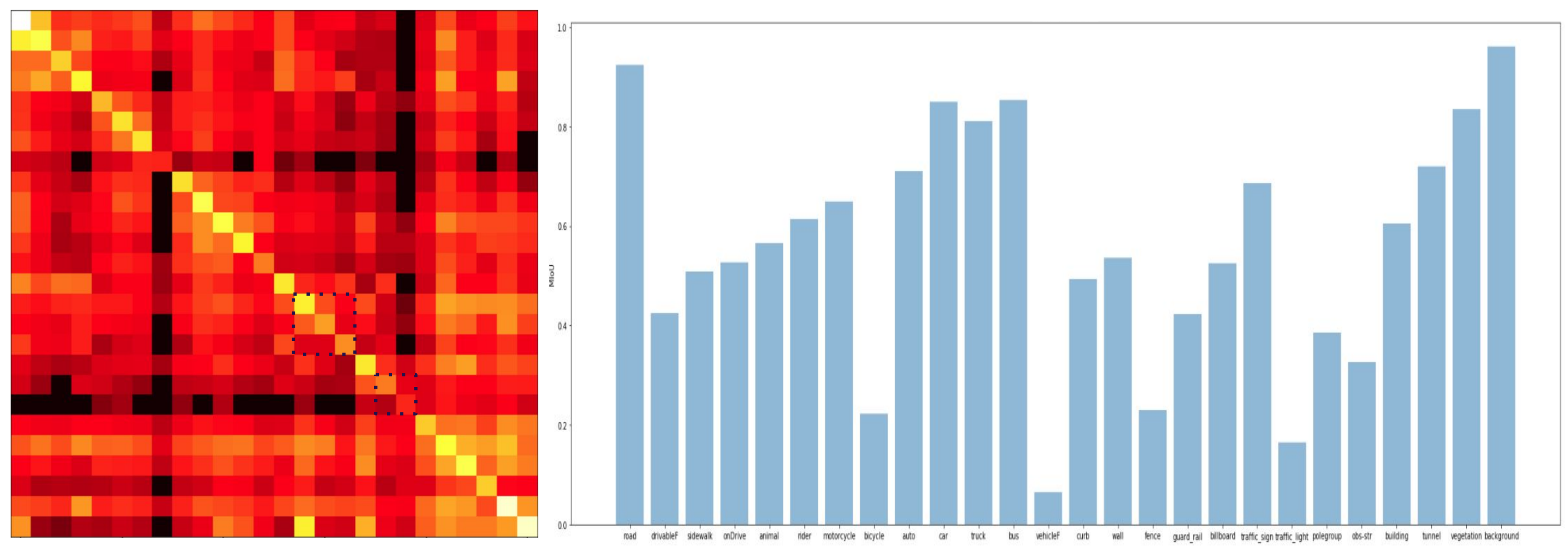
TL;DR: A simple approach to enhance the performance of Semantic Segmentation model.

Overview

Introduction

- Semantic Segmentation is much more challenging in the presence of multiple similar classes, and high intra-class variations.
- Datasets such as AutoNUE model real-life scenarios, and feature:
  - Large intra-class appearance variations,
  - Presence of low-shot or novel classes.
- In such scenarios, simple deep-learning approaches can have high confusion among similar classes, and hence perform poorly.

Top: AutoNUE dataset. Bottom: Cityscapes dataset.



With simple training, the network gets highly confused between similar classes (confusion matrix on left). This in turn results in poor IOU for such classes (barplot of IOU per class on right).

Motivation

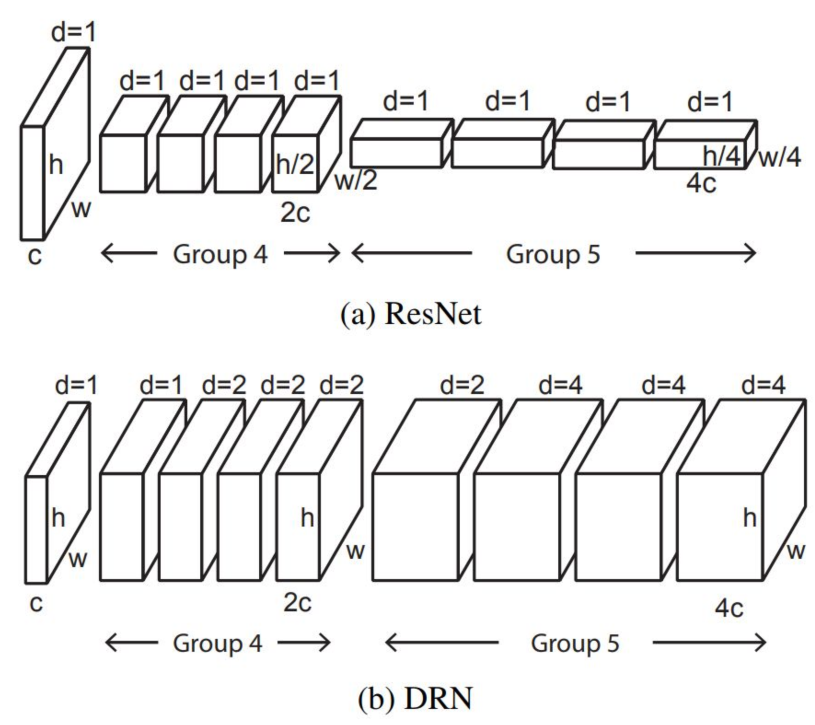
To improve performance in such a unconstrained dataset, it is important to clearly discern the differences between confusing classes.

Hence, in our approach, we propose a novel *Expertise-Layer* to enhance the learned model's discerning ability.

Approach

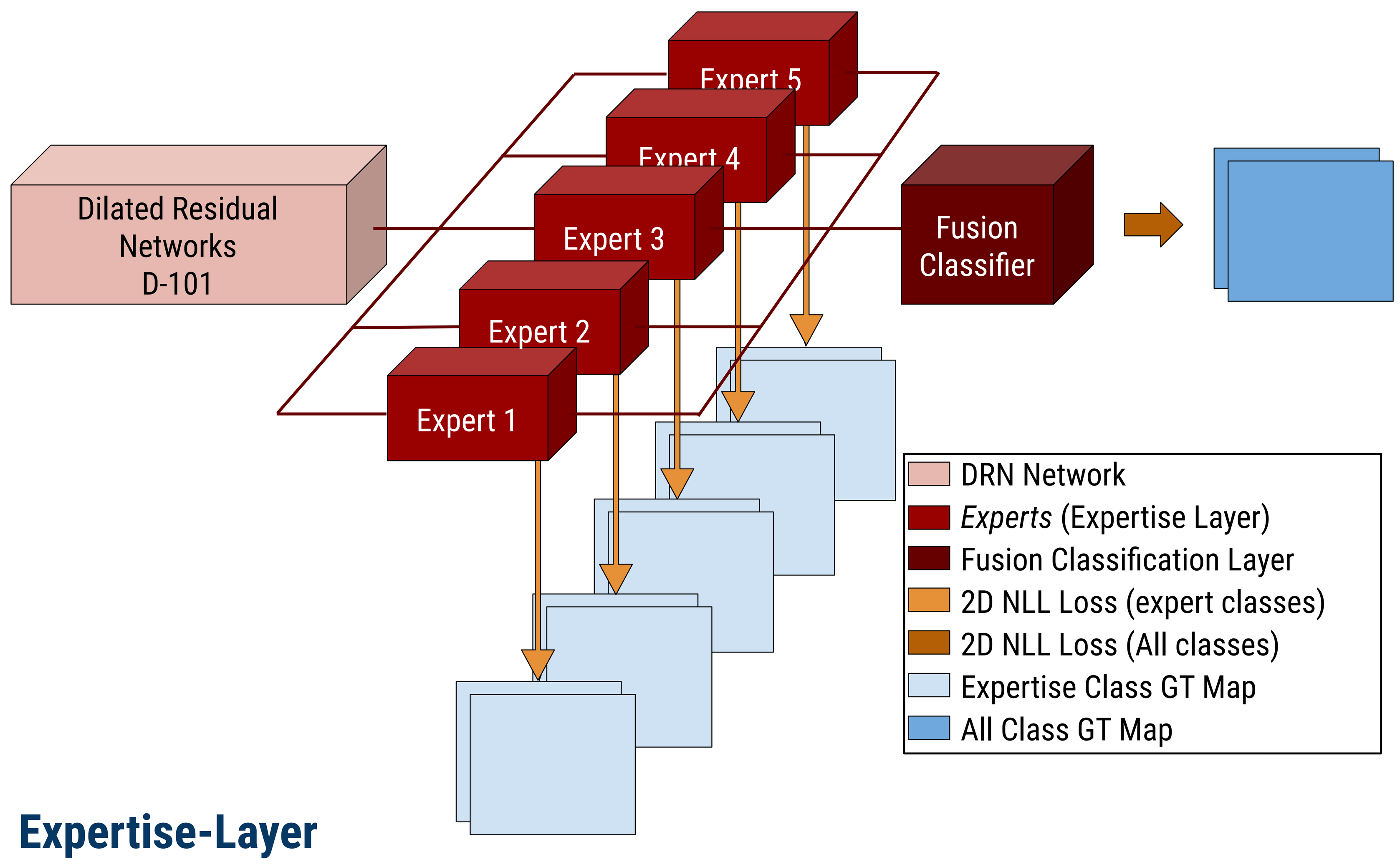
Base Model:

We build our *Expertise-Layer* on top of *Dilated Residual Networks* [Fisher Yu et al., CVPR 2017] (referred as DRN).



- (a) to (b) by removing striding in the convolution layers.
- *Gridding* artifacts removed by:
  - Remove Pooling
  - Adding Convolution layers
  - Removing residual connections

- Use dilation to reduce loss of spatial acuity.
- Architectural innovation to handle *Gridding* artifacts.



Expertise-Layer

- Consist of multiple separate per-pixel classification layers (*Experts*).
- Each *Expert* learns to classify only its own expertise classes.
- Finally, the *Fusion-classification layer* takes all the expert outcomes, and returns the probabilities for the all classes.

Results

- First, we compare simple training v/s training with Expertise Layer. Additionally, we also evaluate which initialization is better suited.

Variant	Mean IOU	PP Accuracy
DRN-D-101	65.76	88.81
DRN-D-101 + Expertise (Imagenet)	**67.21**	**91.06**
DRN-D-101 + Expertise (Cityscapes)	66.33	90.10

- Now, we evaluate whether training with Cityscapes data improves performance.

Variant	Mean IOU	PP Accuracy
AutoNUE data Only	**67.21**	**91.06**
AutoNUE data+ Cityscapes data	62.84	87.20

- Our Final Approach and results on the AutoNUE Test set.

Approach	Mean IOU (AutoNUE Test set)
DRN-D-101+ Expertise	66.73
DRN-D-101+ Expertise + Ensemble	**67.94**

Conclusions

- Our Approach yields an improvement over simple training procedure.
- With this approach, we were able to train a ensemble of 4 DRN-D-101 networks, which enabled us to get MeanIOU of **67.94** on AutoNUE test dataset (*highest from India*).
- This work lays the groundwork for the idea of an *Expertise-Layer*. In future work, we will explore better fusion of "Expertise" and also data-driven formulation of each Expert's classes.

ACKNOWLEDGEMENT: We would like to thank Intel for providing the travel grants and SERC, IISc for providing hardware support.