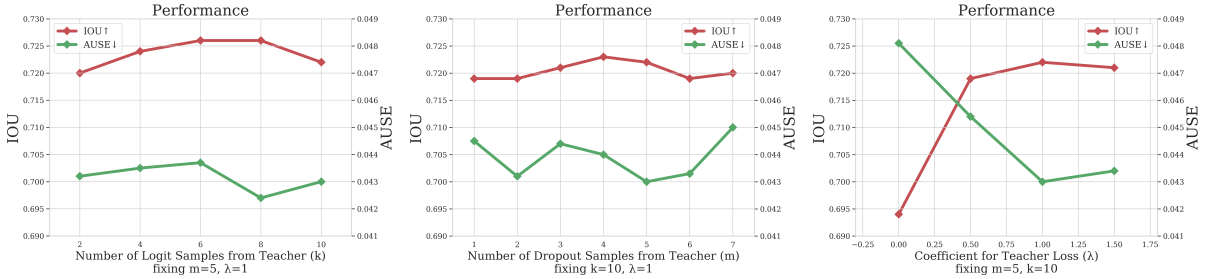
# 6. Appendix

| CamVid | | | | |
|---|---|---|---|---|
| | ACC | IOU | ECE $_{(\times 10^{-3})}$ | AUSE $_{(\times 10^{-2})}$ |
| Best Single | 0.900 | 0.641 | 8.27 | 4.46 |
| Teacher | 0.904 | 0.650 | 5.42 | 3.02 |
| Student | **0.909** | **0.653** | **2.96** | **1.91** |
| **NYU** | | | | |
| | RMSE | REL | ECE $_{(\times 10^{-3})}$ | AUSE $_{(\times 10^{-2})}$ |
| Best Single | 0.543 | 0.149 | 70.8 | 6.11 |
| Teacher | **0.510** | **0.140** | 56.4 | **5.58** |
| Student | 0.530 | 0.144 | **56.3** | 5.93 |

Table 4: Performance of teacher and student model when a Deep Ensemble is used as the teacher. "Best Single" represents the best NN among all in the ensemble in terms of IOU/RMSE. For "Best Single", only the aleatoric uncertainty is used to compute uncertainty metrics.




(a) (b) (c)

Figure 9: Ablation study conducted using VOC2012 dataset. (a-b) Performance of the student model when the number of samples from the teacher model are varied at each mini-batch. As seen in the plots, the performance is generally insensitive to the choice of sample size. Using larger number of samples only brings slight improvement in performance up to a point. (c) Performance of student model against $\lambda$, the weight put on the teacher loss (See Eqn. 11). As seen clearly, introducing the teacher loss improves the performance of the student and the student performs the best when $\lambda = 1$.

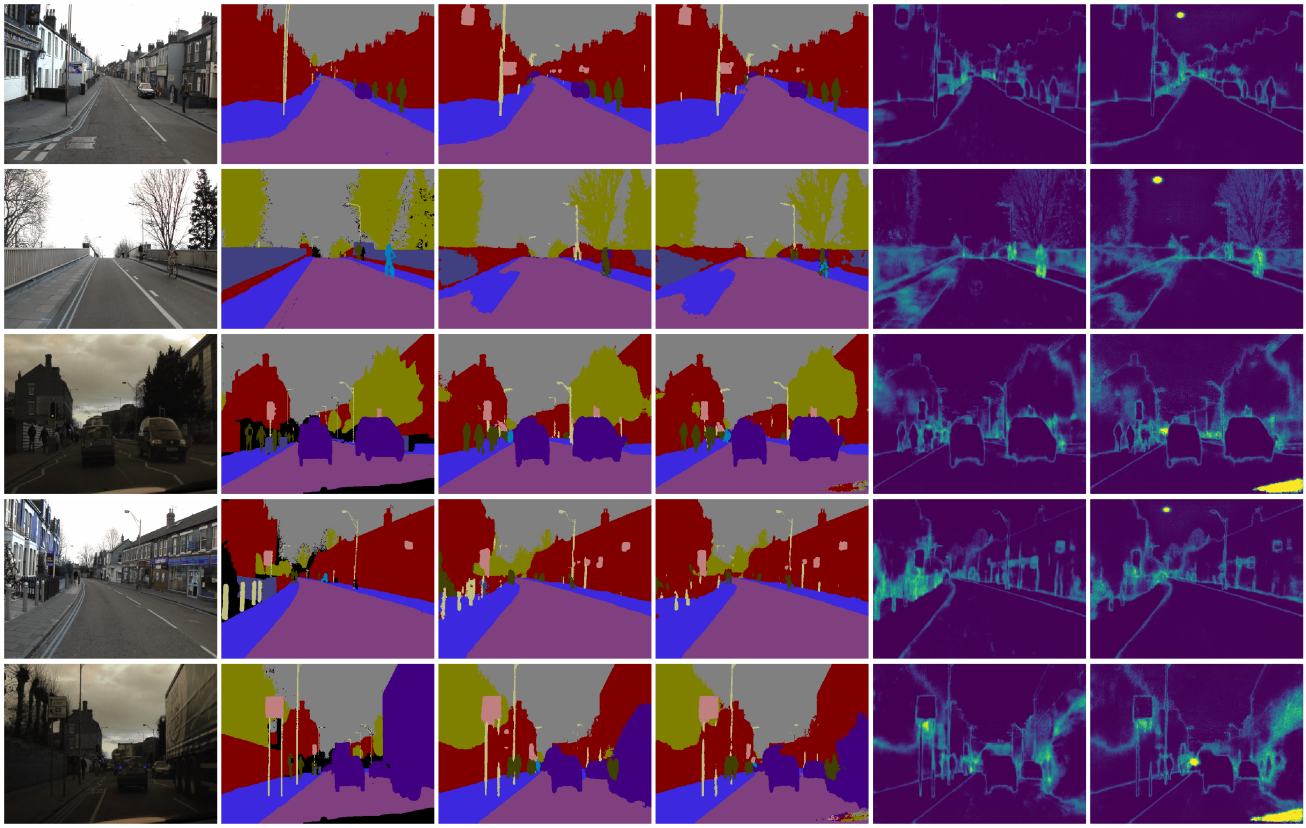



Figure 10: Additional example predictions on CamVid.

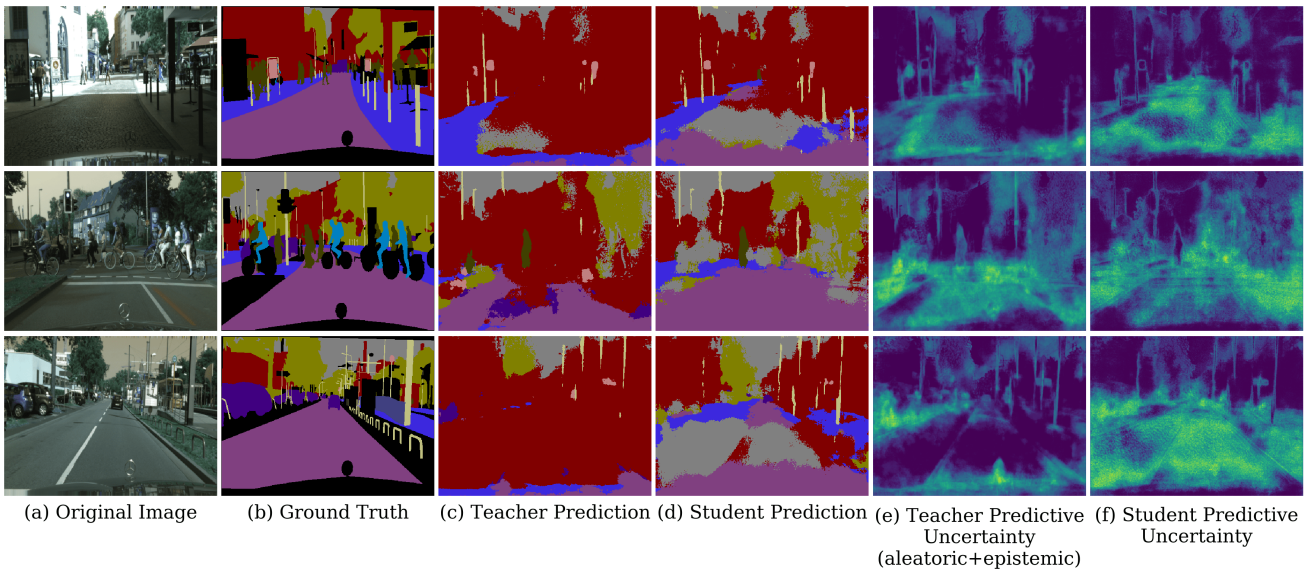



Figure 11: Example predictions on the Cityscapes dataset under distribution shift using models trained with CamVid.

(a) Original Image (b) Ground Truth (c) Teacher Prediction (d) Student Prediction (e) Teacher Predictive Uncertainty (aleatoric+epistemic) (f) Student Predictive Uncertainty

Figure 12: Additional example predictions on Pascal VOC2012.

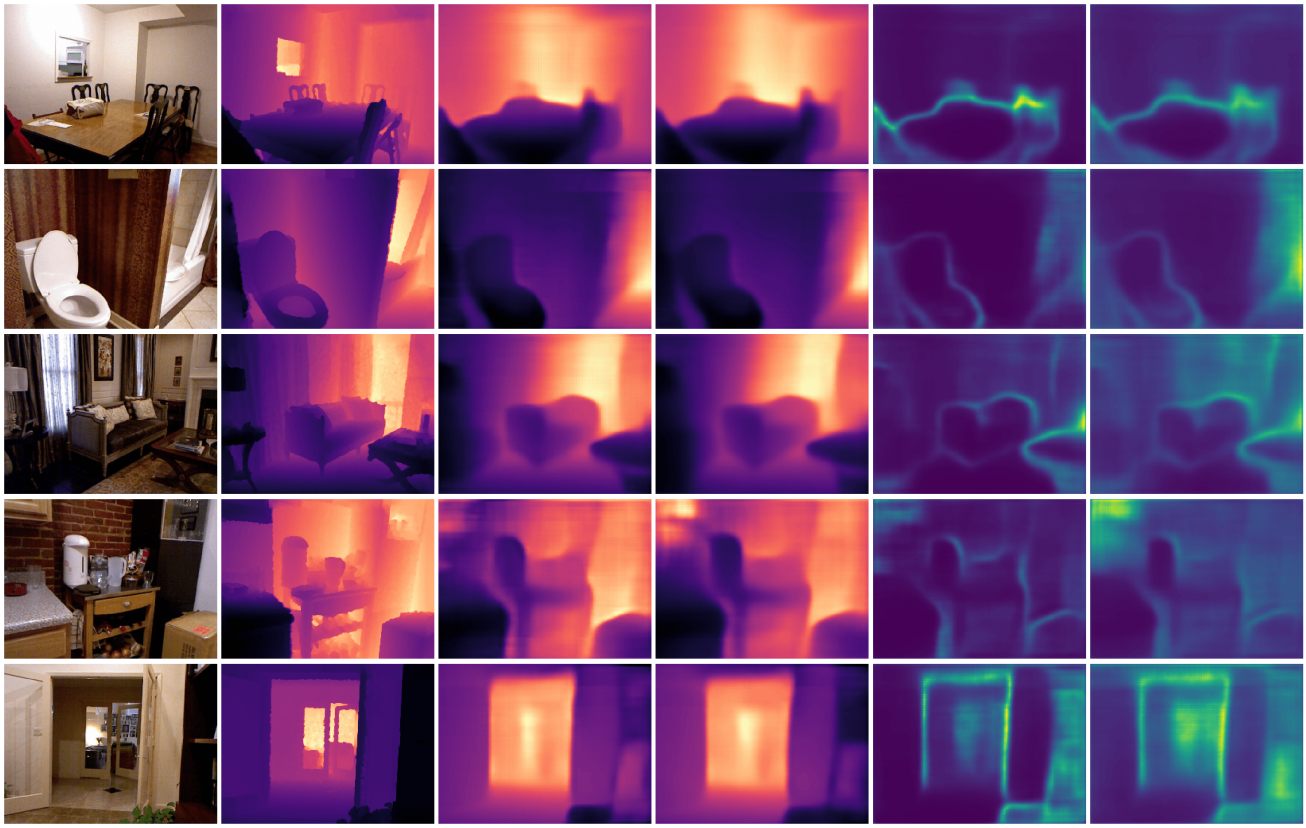



Figure 13: Additional example predictions on NYU.

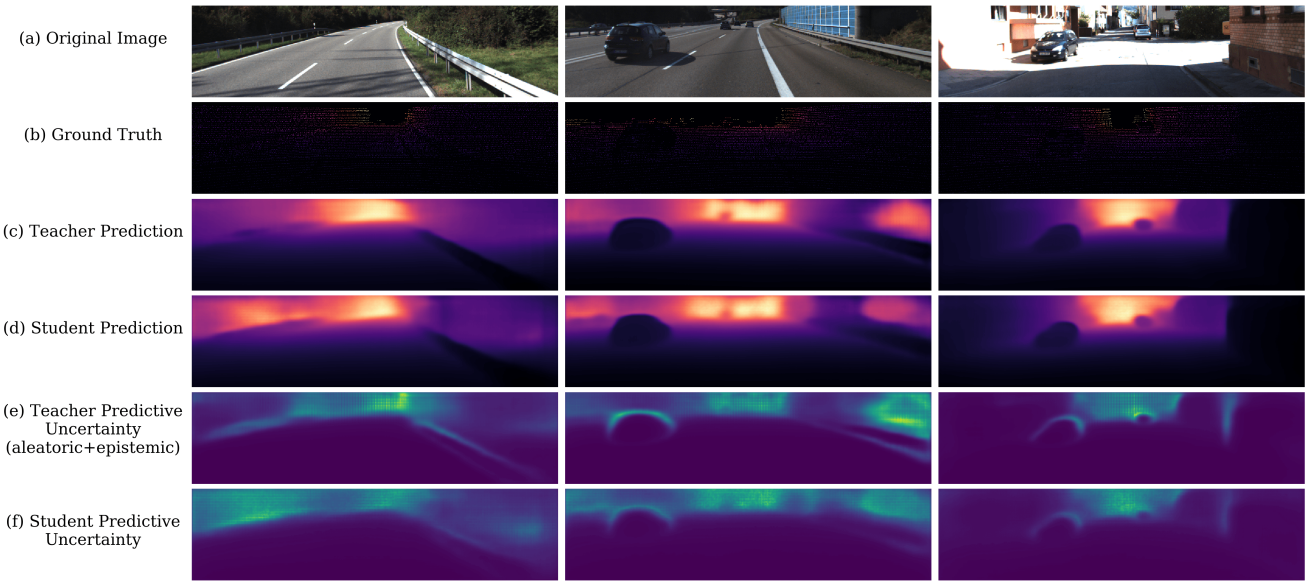
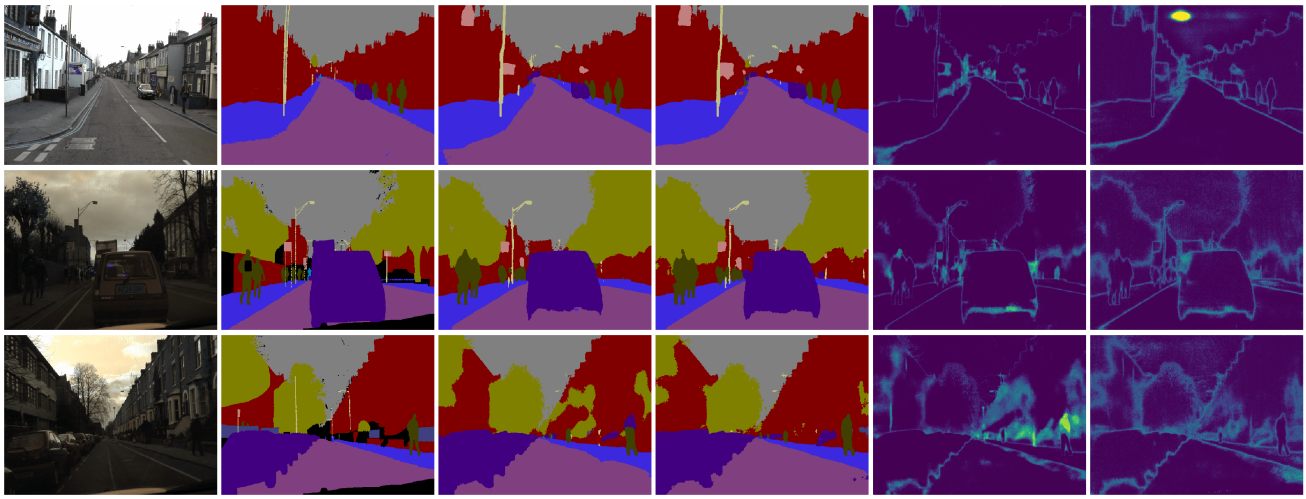



Figure 14: Example predictions on KITTI.



(a) Original Image (b) Ground Truth (c) Teacher Prediction (d) Student Prediction (e) Teacher Predictive Uncertainty (aleatoric+epistemic) (f) Student Predictive Uncertainty

Figure 15: Example predictions on CamVid when using deep ensemble as the teacher model.

(a) Original Image (b) Ground Truth (c) Teacher Prediction (d) Student Prediction (e) Teacher Predictive Uncertainty (aleatoric+epistemic) (f) Student Predictive Uncertainty

Figure 16: Example predictions on NYU when using deep ensemble as the teacher model.