1. Ablation Study

Ablation Study of Co-Occurrence Learning Module. We test the effects of different weights used in combining Grad-CAM map and co-occurrence map. The experiments are performed on the dataset with SNR 0.03. We report the mIoU evaluation results in Figure 1 with Grad-CAM map weight from 0 to 1 in steps of 0.1. Assuming the weight of Grad-CAM is $w_1$, then the weight of the co-occurrence map is $1 - w_1$. We observe that the model gets significant performance improvements from combining Grad-CAM map with the co-occurrence map. The model gets the best performance when the Grad-CAM weight equals 0.3 and the co-occurrence weight equals 0.7.

![Figure 1: Ablation Study of Co-Occurrence Learning.](image)

2. Qualitative Analysis

We visualize the semantic segmentation results produced by different baselines and CIVA-Net in Figure 2. Grad-CAM, Respond-CAM and our CIVA-Net are trained with image-level class labels. VoxResNet is trained with full segmentation supervision. Our proposed network can generate more accurate segmentation results compared to image-level baselines, and produce comparable results to the model trained with stronger supervision.

*Corresponding Author
Figure 2: Visualization of different baselines and *CIVA-Net*. 