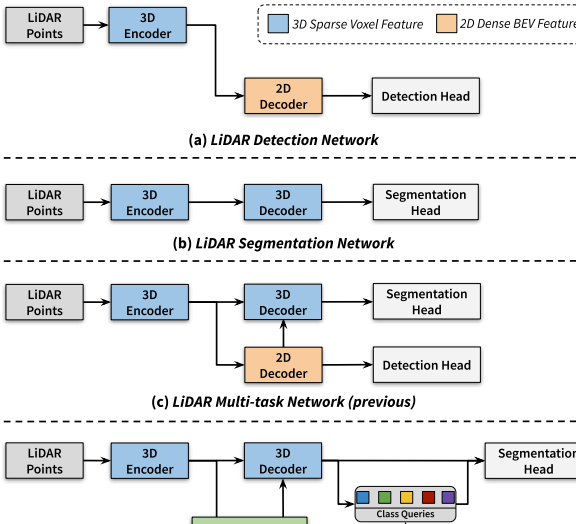

Figure 1: **LiDAR Perception Network Designs.** LiDAR detection (a) and segmentation (b) networks typically extract feature representations on distinct feature maps. While a recent multi-task network [66] (c) integrates these tasks into a single network, it often overlooks differences among feature maps and the higher-level connections between tasks. Our network (d) utilizes transformer attention to establish more effectively the transformations between 3D sparse and 2D dense features. Moreover, the cross-task information is further shared through class-level and object-level feature embeddings in the multi-task transformer decoder.