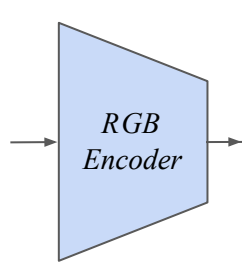


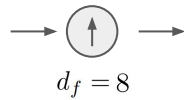
N camera images



Multiscale camera features

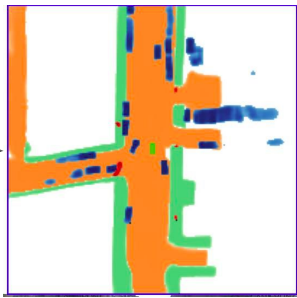
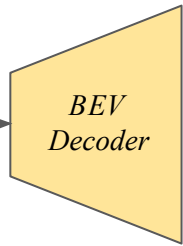


$d_f = 16$



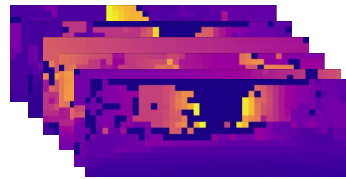
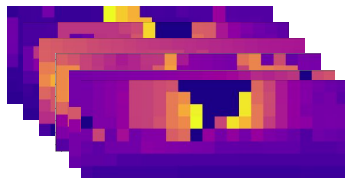
$d_f = 8$

Per-scale BEV map



Semantic Grids

Per-camera / per resolution depth images



 - *Projection to 3D + SumPooling*