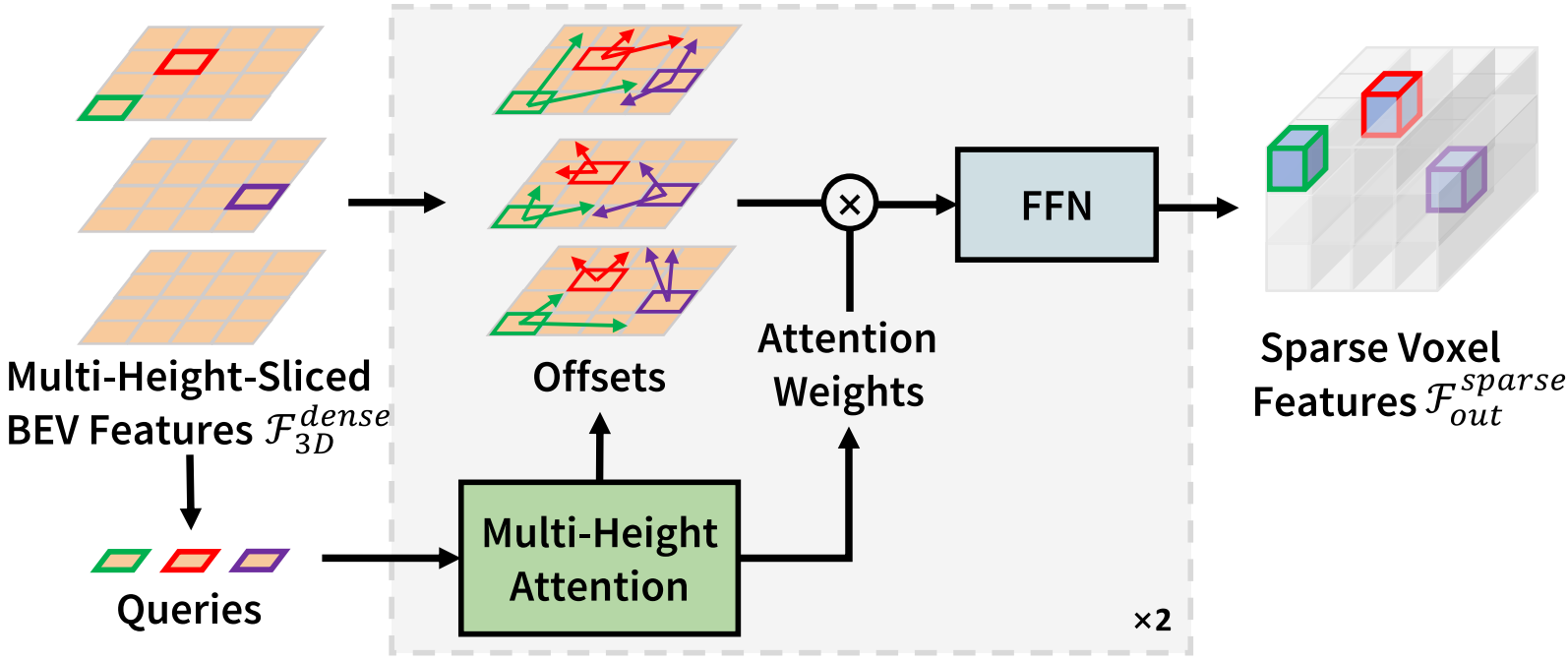
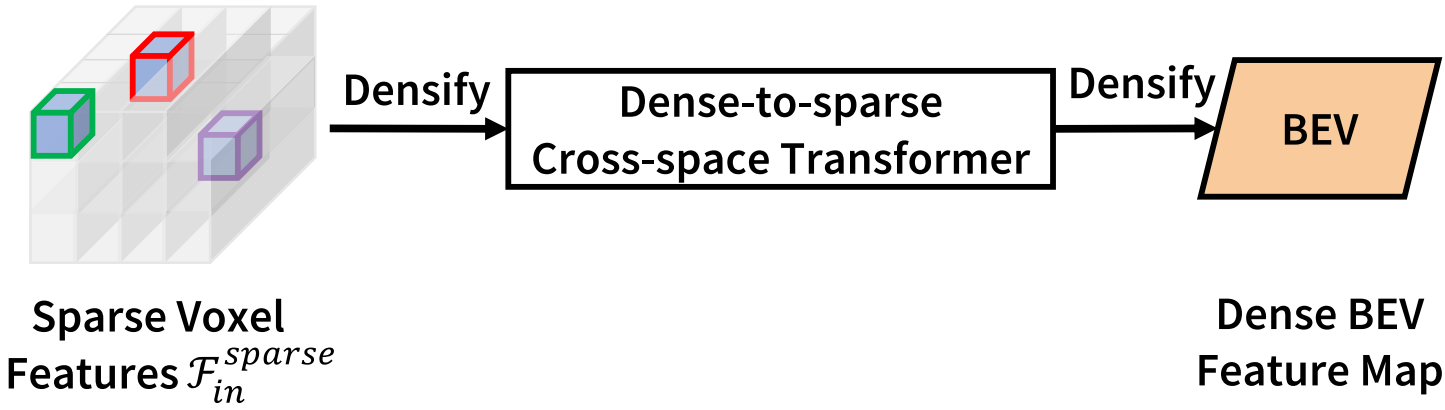


(a) Dense-to-sparse Cross-space Transformer



(b) Sparse-to-dense Cross-space Transformer

Figure 3: **Illustration of the Cross-space Transformer (XSF) module.** XSF consists of two parts: a multi-height deformable self-attention, and a feed-forward network. (a) convert dense BEV features to sparse voxel features, (b) convert sparse voxel features to dense BEV features with two more densify operations.