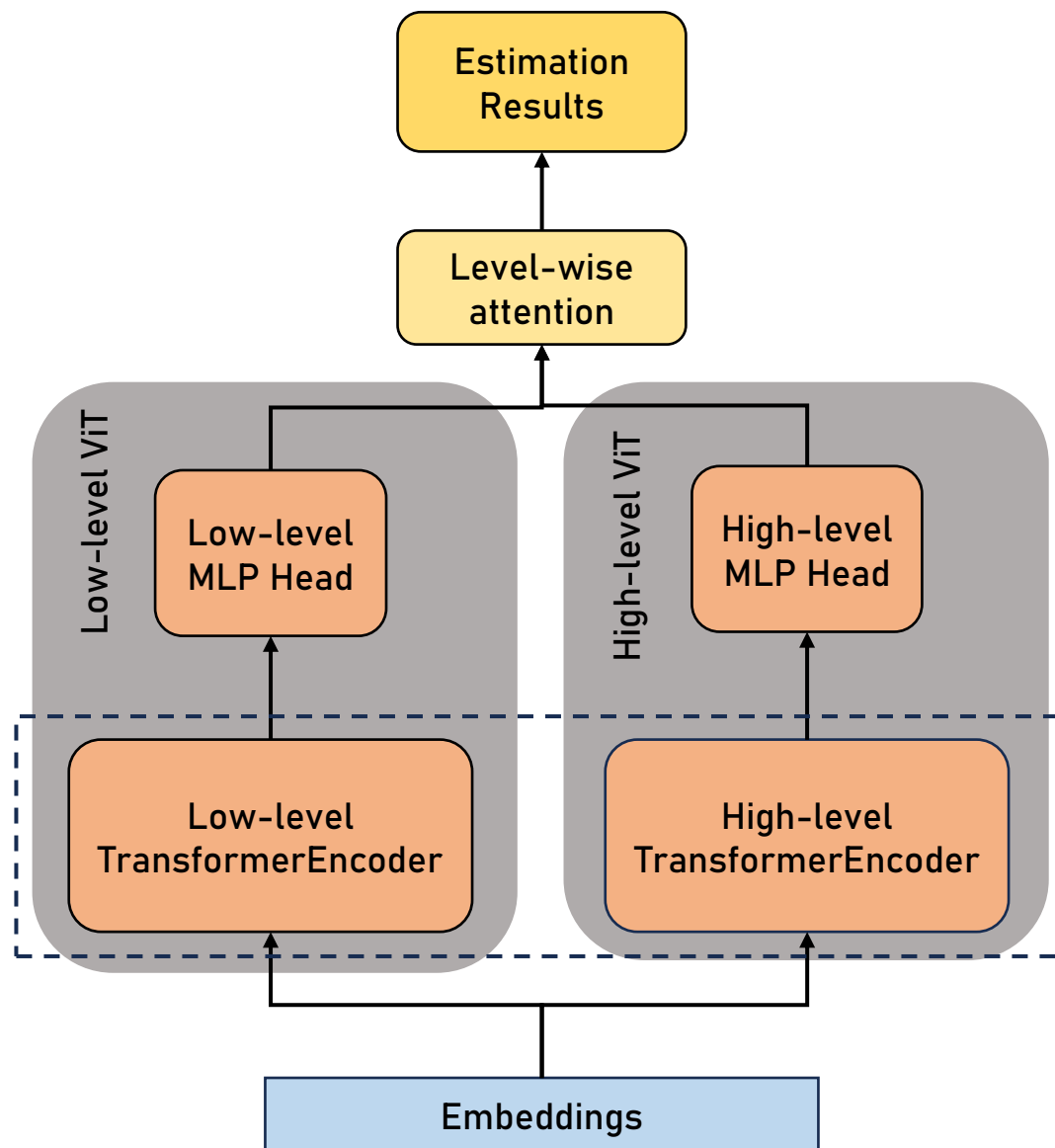


Level-wise Cross attention ViT



TransformerEncoder

