

**Attribute-Guided Transformer (AGT)** 

**Visual-Semantic Embedding Network (VSEN)**