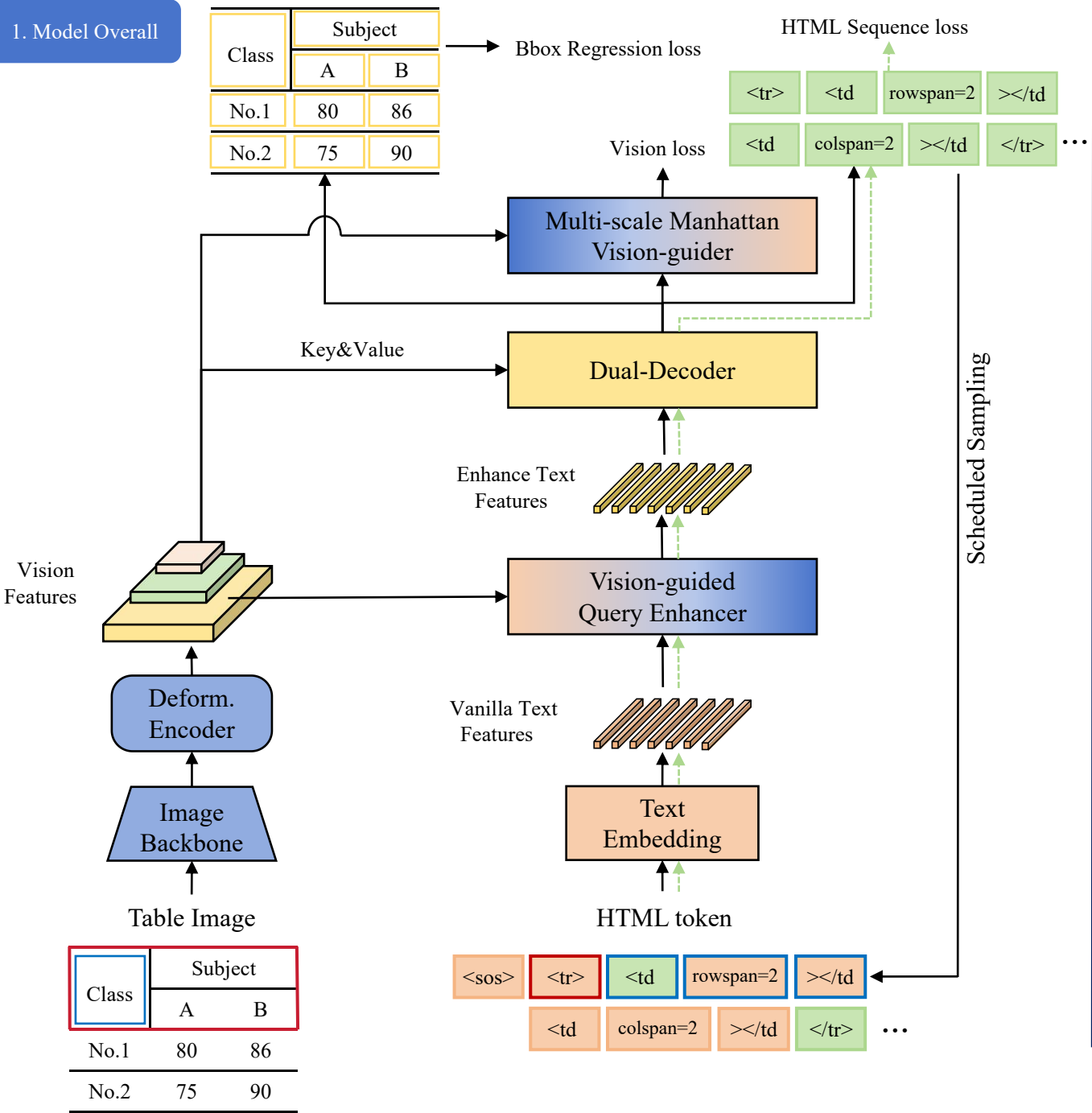
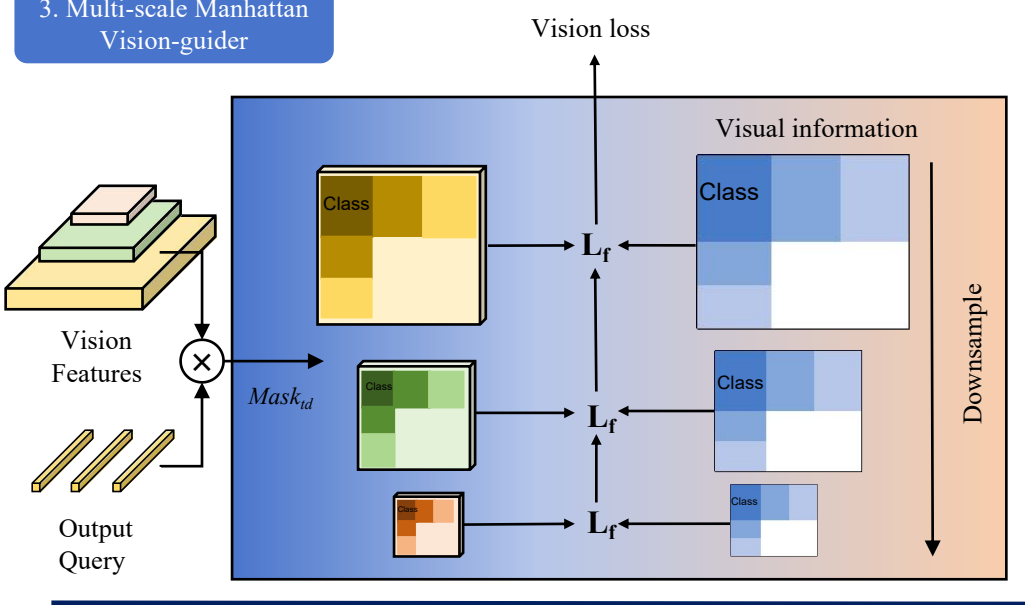


# 1. Model Overall



# 3. Multi-scale Manhattan Vision-guider



# 2. A Layer of Vision-guided Query Enhancer

