

- Image Features
- Recognition Queries
- Detection Queries
- Prompt Queries

Video-level text spotting



Word-level text spotting (TotalText)



Line-level text spotting (CTW1500)



OR

Word-level text spotting

Prompt Generation Module

ResNet50

REM

Transformer Encoder

Query Initialization

Fusion

Adapter

Prompt Generation Module

Line-level text spotting
OR
Video-level text spotting

VimTS

Task-aware Adapter

Decoder

Word-level text spotting



Line-level text spotting

