MANGO: A Mask Attention Guided One-Stage Scene Text Spotting

Liang Qiao¹  Ying Cheng²  Zhanzhan Cheng³  Yunlu Xu¹  Yi Niu¹  Shiliang Pu¹  Fei Wu³
¹ Hikvision Research Institute  ² Tongji University  ³ Zhejiang University

MOTIVATION & CONTRIBUTIONS

• Existing two-staged end-to-end text spotting methods adopt complicated RoI operations with time-consuming NMS.

• Recognition highly relies on detection results and its annotations. Accurate boundaries need to be depicted during detection.

• We propose a compact and robust one-stage text spotting framework named MANGO that can be trained in an end-to-end manner.

• We develop the position-aware mask attention module to generate the text instance features into a batch, and build the one-to-one mapping with final character sequences. The module can be trained with only rough text position information and text annotations.

METHOD

• Existing two-staged end-to-end text spotting methods adopt complicated RoI operations with time-consuming NMS.

• Recognition highly relies on detection results and its annotations. Accurate boundaries need to be depicted during detection.

• We propose a compact and robust one-stage text spotting framework named MANGO that can be trained in an end-to-end manner.

• We develop the position-aware mask attention module to generate the text instance features into a batch, and build the one-to-one mapping with final character sequences. The module can be trained with only rough text position information and text annotations.

EVALUATION & VISUALIZATION