

Recent VQA Approaches

Wentao Mo¹

¹Department of Machine Intelligence
Peking University

2021 年 4 月 27 日

1 Visual Feature Extractor

- VC R-CNN
- Grid Feature
- VinVL

2 Feature Fusion Methods

- MFH
- MCAN
- TRRNet
- BERT/Transformer-based

Visual Commonsense R-CNN

In Defense of Grid Features for VQA

MFH: Multimodal Factorized High-order Attention

MCAN: Deep Modular Co-Attention Networks

TRRNet: Tiered Relation Reasoning for Compositional VQA

OSCAR & UNITER: Transformer-like Fusion