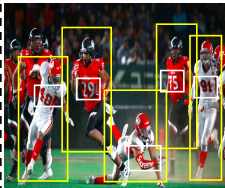


Input

what is the number of the player to the left of the ball carrier?

Tokenize

What is the number of the player to ...



Faster R-CNN



Microsoft-OCR

OCR Augmentation

OCR_iText

Semantically Consistent Encoding

LXMERT

Question

Question-related objects

Uni-Modal Transformer

Question Embedding



Object Embedding



Tanh

\oplus



UMT_output

OCR Embedding



Previous prediction embedding

Answer Generation

Pre-training Tasks

Multimodal Transformer Layers

Loss Functions

Question

L_{MLM}

Objects

L_{RPP}

Image/Text Match

L_{ITM}

OCR Token

L_{TWC}

Fine-tuning Tasks

Pointer Network

Linear Layer

Pointer Network

OCR Token Prediction

Vocabulary Prediction

$\begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix}$

Vocabulary Embedding

Answer: 48