Image representation
Text representation
Semantic Space
Unified-Modal Transformer
[IMG]
[CLS]
The
man
umpires
looks
[SEP]
Any baseball game involves one or more umpires, who make rulings on the outcome of each play. At a minimum, one umpire will stand behind the catcher, to have a good view of the strike zone, and call balls and strikes. Additional umpires may be stationed near the other bases …
The baseball player readies to swing at the pitch while the umpire behind him looks on.
Image Collections
Text Corpus
Image-Text Pairs