Visual Question Answering with Graph Matching and Reasoning

Anonymous ICCV submission

Paper ID ****

Abstract

Visual Question Answering (VQA) is of great significance in offering people convenience: one can raise a question for details of objects, or high-level understanding about the scene, over an image. This paper proposes a novel method to address the VQA problem. In contrast to prior works, our method that targets single scene VQA, replies on graph-based techniques and involves reasoning. In a nutshell, our approach is centered on three graphs. The first graph, referred to as inference graph G_I, is constructed via learning over labeled data. The other two graphs, referred to as query graph Q and entity-attribute graph G_{EA}, are generated from natural language query Q_{nl} and image Img, that are issued from users, respectively. As G_{EA} often does not take sufficient information to answer Q, we develop techniques to infer missing information of G_{EA} with G_I. Based on G_{EA} and Q, we provide techniques to find matches of Q in G_{EA}, as the answer of Q_{nl} in Img. Unlike commonly used VQA methods that are based on end-to-end neural networks, our graph-based method shows well-designed reasoning capability, and thus is highly interpretable. We also create a dataset on soccer match (SoccerVQA) with rich annotations. The experimental results show that our approach outperforms the state-of-the-art method and has high potential for future investigation.

1. Reasoning

1.1. Inference Graph

$$P_{\mathfrak{B}}(X_1, \cdots, X_n) = \prod_{i=1}^{n} P_{\mathfrak{B}}(X_i|\text{Pa}(X_i)) = \prod_{i=1}^{n} \theta_{X_i|\text{Pa}(X_i)} \tag{1}$$

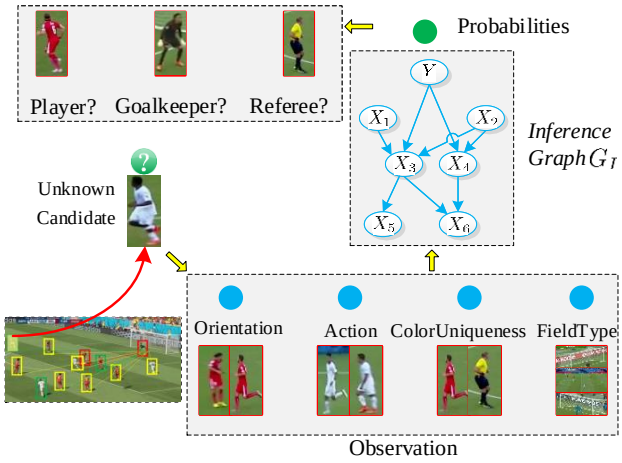

Figure 1: The pipeline of inference graph used for inferring the role of a person object.

$$\begin{aligned} P_{\mathfrak{B}}(Y|X) &= \frac{P_{\mathfrak{B}}(Y)P_{\mathfrak{B}}(X|Y)}{P_{\mathfrak{B}}(X)} \\ &= \frac{\theta_{Y|\text{Pa}(X)} \prod_{i=1}^{n} \theta_{X_i|Y,\text{Pa}(X_i)}}{\sum_{y' \in \mathcal{Y}} \theta_{y'|\text{Pa}(X_0)} \prod_{i=1}^{n} \theta_{X_i|y',\text{Pa}(X_i)}} \end{aligned} \tag{2}$$

1.2. Learning the Inference Graph

$$P_{\mathfrak{B}}(Y|X) = c \cdot \theta_Y \prod_{i=1}^{n} \theta_{X_i|Y} \tag{3}$$

References