

# Understanding Knowledge Gaps in Visual Question Answering: Implications for Gap Identification and Testing

Goonmeet Bajaj

Bortik Bandyopadhyay

Daniel Schmidt<sup>†</sup>

Pranav Maneriker

Christopher Myers<sup>‡</sup>

Srinivasan Parthasarathy

Fbajaj.32, bandyopadhyay.14, maneriker.1G@osu.edu, schmidt.152@wright.edu,  
christopher.myers.29@us.af.mil, srin@cse.ohio-state.edu

## Abstract

*Traditional Visual Question Answering (VQA) datasets typically contain questions related to the spatial information of objects, object attributes, or general scene questions. Recently, researchers have recognized the need to improve the balance of such datasets to reduce the system’s dependency on memorized linguistic features and statistical biases, while aiming for enhanced visual understanding. However, it is unclear whether any latent patterns exist to quantify and explain these failures. As an initial step towards better quantifying our understanding of the performance of VQA models, we use a taxonomy of Knowledge Gaps (KGs) to tag questions with one or more types of KGs. Each KG describes the reasoning abilities needed to arrive at a resolution, and failure to resolve gaps indicates an absence of the required reasoning ability. After identifying KGs for each question, we examine the skew in the distribution of questions for each KG. We then introduce a targeted question generation model to reduce this skew, which allows us to generate new types of questions for an image.*

## 1. Introduction

When compared to artificially intelligent (AI) systems, human cognition demonstrates a reasonably flexible system when faced with gaps in knowledge while executing a prescribed task. Humans often demonstrate both the ability to identify gap(s) in their knowledge and the ability to resolve these different gaps through diverse strategies (e.g., by seeking clarification, conducting research, etc.).

Informally, a knowledge gap (KG) is an instance of limited or missing information or capabilities, which leads to an AI agent being inefficient or incapable of completing a given task. AI agents, when presented a same/similar task

Figure 1: Skew in the Distribution of Questions per KG

as humans, might not always have the perfect knowledge to complete it [8]. A framework for KG identification and resolution can facilitate flexibility for an AI agent during both training and execution. As a preliminary effort to understand how to build a system for AI agents to *detect*, *identify*, and *resolve* the KGs that can occur, we leverage *visual question answering* (VQA) tasks. As a first step, we aim to detect gaps in knowledge and identify the knowledge gap type in this work. VQA sits at an intersection of three components of artificial intelligence: *language*, *vision*, and *reasoning*, thus making it a challenging task. To the best of our knowledge, we are the first to systematically manipulate VQA questions to produce knowledge gaps within AI agents.

Using a refined version of a KG taxonomy [1], we initially eight different KGs that occur in the GQA dataset [4]. Figure 1 shows the skew we observe in the distribution of the number of questions per KG category. To alleviate this skew and make questions more evenly distributed across KGs, we apply a neural framework to generate questions for specific KGs.

The Ohio State University (OSU)

<sup>†</sup>Wright State University

<sup>‡</sup>Air Force Research Laboratory (AFRL)





