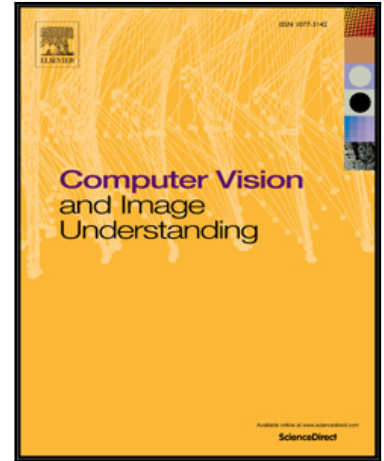


## Accepted Manuscript

Visual Question Answering: Datasets, Algorithms, and Future Challenges

Kushal Kafle, Christopher Kanan

PII: S1077-3142(17)30117-0  
DOI: [10.1016/j.cviu.2017.06.005](https://doi.org/10.1016/j.cviu.2017.06.005)  
Reference: YCVIU 2589



To appear in: *Computer Vision and Image Understanding*

Received date: 14 September 2016  
Revised date: 27 May 2017  
Accepted date: 10 June 2017

Please cite this article as: Kushal Kafle, Christopher Kanan, Visual Question Answering: Datasets, Algorithms, and Future Challenges, *Computer Vision and Image Understanding* (2017), doi: [10.1016/j.cviu.2017.06.005](https://doi.org/10.1016/j.cviu.2017.06.005)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Visual Question Answering: Datasets, Algorithms, and Future Challenges

Kushal Kafle<sup>a</sup>, Christopher Kanan<sup>a,\*</sup>

<sup>a</sup>*Chester F. Carlson Center for Imaging Science,  
Rochester Institute of Technology,  
Rochester, NY, 14623, USA*

---

## Abstract

Visual Question Answering (VQA) is a recent problem in computer vision and natural language processing that has garnered a large amount of interest from the deep learning, computer vision, and natural language processing communities. In VQA, an algorithm needs to answer text-based questions about images. Since the release of the first VQA dataset in 2014, additional datasets have been released and many algorithms have been proposed. In this review, we critically examine the current state of VQA in terms of problem formulation, existing datasets, evaluation metrics, and algorithms. In particular, we discuss the limitations of current datasets with regard to their ability to properly train and assess VQA algorithms. We then exhaustively review existing algorithms for VQA. Finally, we discuss possible future directions for VQA and image understanding research.

*Keywords:* Image Understanding, Natural Language Processing, Vision and Language

---

## 1. Introduction

Recent advancements in computer vision and deep learning research have enabled enormous progress in many computer vision tasks, such as image clas-

---

\*Corresponding author

*Email addresses:* [kk6055@rit.edu](mailto:kk6055@rit.edu) (Kushal Kafle), [christopher.kanan@rit.edu](mailto:christopher.kanan@rit.edu) (Christopher Kanan)

Download English Version:

<https://daneshyari.com/en/article/6937488>

Download Persian Version:

<https://daneshyari.com/article/6937488>

[Daneshyari.com](https://daneshyari.com)