

**ROV: REASONING ON VISION, VISUAL QUESTION
ANSWERING MODEL BASED ON IMAGE TEXT**

Jinali Wijetunge

A dissertation submitted in partial fulfilment of the requirement for Bachelor of
Engineer (Honours) degree in Software Engineering

**Department of Computing
Informatics Institute of Technology, Sri Lanka
in collaboration with
University of Westminster, UK**

2020

Abstract

Visual Question Answering is answering natural language questions according to a given visual content. Humans can understand visual content effortlessly with one glance at an image where computers find this extremely hard which needs high amount of cognition to recognize about the world. Visual Question Answer, the name itself define three major components a solver should contain. They are a visual, a question and an answer, where the visual and the question be the input and the answer be the output. It allows machines to reason across language and vision like human beings and is viewed as one key measurement for machine intelligence.

Visual understanding is far away from object recognition using computer systems. Answering a natural language question related to an image and then providing a rationale justification about the answer is defined as Visual Common-sense Reasoning. To fulfill this task machines, require high order cognition and common-sense reasoning about the world.

RoV is a visual question answering approach base on image question. RoV provides a novel design algorithm to be used in VQA. Developing such system is helpful various way in order to overcome the day today activities

Keywords: Visual Question Answering, Image Text, NLP, Computer Vision