# Learning by Asking Questions

Xuewen Yang

July 18 2018

## Abstract

*In this paper,the author introduce an interactive learning framework for the development and testing of intelligent visual systems, called learning-by-asking (LBA).They explore LBA in context of the Visual Question Answering (VQA) task. LBA differs from standard VQA training in that most questions are not ob- served during training time, and the learner must ask ques- tions it wants answers to. Thus, LBA more closely mim- ics natural learning and has the potential to be more data- efficient than the traditional VQA setting.They present a model that performs LBA on the CLEVR dataset, and show that it automatically discovers an easy-to-hard curriculum when learning interactively from an oracle. Our LBA gener- ated data consistently matches or outperforms the CLEVR train data and is more sample efficient.The author also show that the model asks questions that generalize to state-of-the-art VQA models and to novel test time distributions.*

## 1. Introduction

Machine learning models have led to remarkable progress in visual recognition. However, while the training data that is fed into these models is crucially important, it is typically treated as predetermined, static information. The current models are passive in nature: they rely on training data curated by humans and have no control over this supervision. This is in stark contrast to the way we humans learn ł by interacting with our environment to gain information. The interactive nature of human learning makes it sample efficient (there is less redundancy during training) and also yields a learning curriculum (we ask for more complex knowledge as we learn).

In this paper,the author argue that next-generation recognition systems need to have agency ł the ability to decide what information they need and how to get it.They explore this in the context of visual question answering (VQA). Instead of training on a fixed, large-scale dataset, we propose an alternative interactive VQA setup called learning- by-asking (LBA): at training time, the learner receives only images and decides what questions to ask. Questions asked by the learner are answered by an oracle (human supervision). At test-time, LBA is evaluated exactly like VQA using well understood metrics.

## 2. Related Work

Visual question answering (VQA) is a surrogate task designed to assess a systems ability to thoroughly understand images. It has gained popularity in recent years due to the release of several benchmark datasets [3]. Motivated by the well-studied difficulty of analyzing results on real-world VQA datasets [1] [4], Johnson et al. [23] recently proposed a more controlled, synthetic VQA dataset that we adopt in this work.

Visual question generation (VQG) was recently proposed as an alternative to image captioning. Our work is related to VQG in the sense that we require the learner to generate questions about images, however, our objective in doing so is different. Whereas VQG focuses on asking questions that are relevant to the image content, LBA requires the learner to ask questions that are both relevant and informative to the learner when answered.

Active learning (AL) involves a collection of unlabeled examples and a learner that selects which samples will be labeled by an oracle. Common selection criteria include entropy [2], boosting the margin for classifiers and expected informativeness.

## 3. Approach

They propose an LBA agent built from three modules: (1) a question proposal module that generates a set of question proposals for an input image; (2) a question answering module (or VQA model) that predicts answers from (I, q) pairs; and (3) a question selection module that looks at both the answering modules state and the proposal modules questions to pick a single question to ask the oracle. After receiving the oracles answer, the agent creates a tuple (I, q, a) that is used as the online learning signal for all three modules. Each of the modules is described in a separate subsection below; the interactions between them are illus-

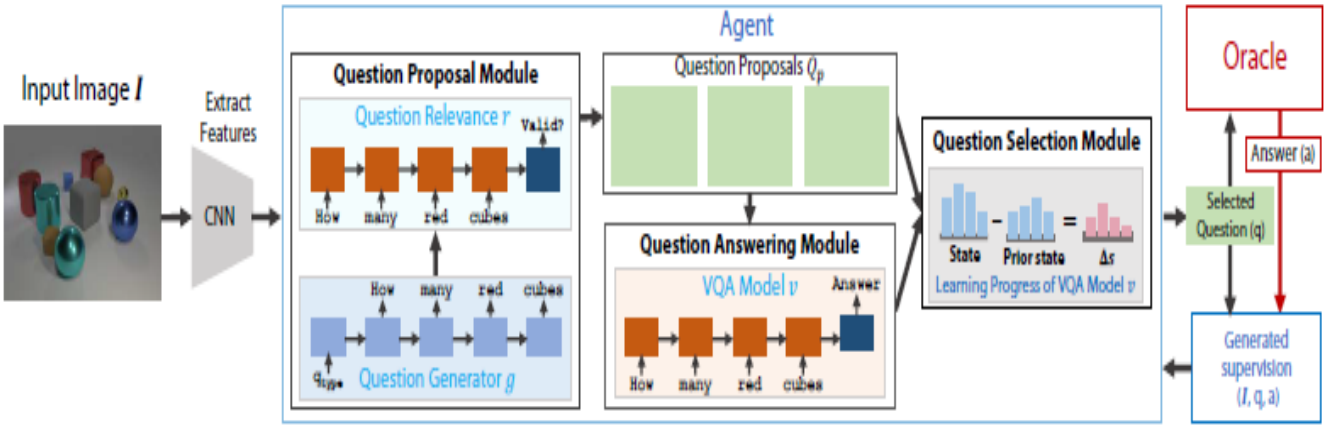


Figure 1. The approach to the learning-by-asking setting for VQA. Given an image I, the agent generates a diverse set of questions using a question generator g. It then filters out irrelevant questions using a relevance model r to produce a list of question proposals. The agent then answers its own questions using the VQA model v. With these predicted answers and its self-knowledge of past performance, it selects one question from the proposals to be answered by the oracle. The oracle provides answer-level supervision from which the agent learns to ask informative questions in subsequent iterations.

trated in Figure 1.

## References


[1] A. Jabri, A. Joulin, and L. V. D. Maaten. Revisiting visual question answering baselines. In *ECCV*, pages 727–739, 2016. 1

[2] A. J. Joshi, F. Porikli, and N. Papanikolopoulos. Multi-class active learning for image classification. In *CVPR*, 2009. 1

[3] M. Malinowski and M. Fritz. A multi-world approach to question answering about real-world scenes based on uncertain input. In *NIPS*, 2014. 1

[4] P. Zhang, Y. Goyal, D. Summersstay, D. Batra, and D. Parikh. Yin and yang: Balancing and answering binary visual questions. In *CVPR*, 2016. 1