

# RESEARCH & TECHNICAL PRESENTATION ASSIGNMENT 3

Ali Gholami

Department of Computer Engineering & Information Technology  
Amirkabir University of Technology

<http://ceit.aut.ac.ir/~aligholamee>  
[aligholamee@aut.ac.ir](mailto:aligholamee@aut.ac.ir)

## Abstract

This assignment provides a new subject for this course. This is my research subject which is illustrated in a *fish-bone* diagram.

**Keywords.** *Research Topic, Dissertation, Ideal Goals, Generic Goals, Domain-specific Goals, Practical-Goals.*

## 1 My Research Topic Illustration & Goals

After investing 2 months of effort on finding a proper research topic, I came up with the following results. Note that at the time of writing this article, I'm totally impressed by the applications of *Computer Vision* in human life. It might seem a little bit too much for a bachelor's degree student like me, but I will be focusing on the *Visual Question Answering* as the main theme of my bachelor's project. In this section I'm supposed to provide my detailed research topic title and goals that support this idea. I'll be deeply going through the *Ideal Goals, Generic Goals, Domain-specific Goals* and *Practical Goals* for this topic.

### Research Title

"Visual Question Answering using Deep Convolutional & Recurrent Neural Networks."

### Ideal Goal

"Improvement of Scene Understanding and Visual Perception."

### Generic Goal

"Integration of Convolutional and Recurrent Neural Networks to Answer Perceptual Scene-oriented Questions."

## Domain-specific Goals

- Dimensionality reduction of high-resolution images using Manifold techniques.
- Object detection using convolutional neural networks.
- Word embedding of questions using LSTM and recurrent neural networks.

## Practical Goals

“Visual assistance tools for the blind.”

## 2 The Problem VQA Solves

Wouldn't it be nice if machines could understand content in images and communicate this understanding as effectively as humans? Such technology would be immensely powerful, be it for aiding a visually-impaired user navigate a world built by the sighted, assisting an analyst in extracting relevant information from a surveillance feed, educating a child playing a game on a touch screen, providing information to a spectator at an art gallery, or interacting with a robot. As computer vision and natural language processing techniques are maturing, we are closer to achieving this dream than we have ever been.

Visual Question Answering (VQA) is one step in this direction. Given an image and a natural language question about the image (e.g., “What kind of store is this?”, “How many people are waiting in the queue?”, “Is it safe to cross the street?”), the machine's task is to automatically produce an accurate natural language answer (“bakery”, “5”, “Yes”). [1]

## 3 Fish-bone Diagram of VQA

The diagram is attached as a separate A4 paper to this assignment.

## 4 VQA Mind Map

The diagram is attached as a separate A4 paper to this assignment.

## References

- [1] Devi Parikh. *Visual Question Answering (VQA)*. CS 294-131: Special Topics in Deep Learning.