**Report of the DataSet**

Group Project Ⅲ

**Visual Question Answering**

Supervisor

郑锋

Team Members

11813226 张一弛

11812518 杨英特



catalogue

[Abstract 3](#_Toc86294238)

[Introduction 3](#_Toc86294239)

[Process 5](#_Toc86294240)

[Data sets 5](#_Toc86294241)

[Design Data Sets 6](#_Toc86294242)

[plan 10](#_Toc86294243)

[Design the model 10](#_Toc86294244)

[VQA program design 10](#_Toc86294245)

[Reference 11](#_Toc86294246)

# Abstract

This paper introduces the goal of this semester: Chinese VQA (CHS-VQA). And in order to design and complete the Chinese VQA, we have completed some preliminary work, such as the research of topic opening and the construction of Chinese database. We hope to bring VQA into the Chinese environment this semester.

# Introduction

For a VQA problem, computer should produce a correct answer from a given picture and a natural language question related to the picture, which is the multimodal problem of the natural language processing (NLP) and the image recognition. This concept of VQA was formally put forward in 2015, when relevant technologies achieved breakthrough development.



Figure 1: Some Common VQA Problems

VQA is a very promising and challenging field in deep learning. In recent years, VQA has developed very rapidly, but there are still many problems. Our previous two innovative practical projects challenged VQA model learning and reasoning enhancement respectively, but because these problems are too close to VQA current bottleneck, the results of learning is not good. With the recommendation of the tutor and reference to the literature this semester, we decided to challenge another part of the problems missing in the development of VQA: Chinese environment and application.

Through investigation, the QA pairs of VQA use English as the input language. Even if the development of Chinese VQA is involved, the Chinese QA pair will be translated into English and input into VQA. In fact, due to the difference between Chinese and English, the accuracy of VQA will be greatly reduced. For example, the “主席” in Figure 2 should be a “椅子”, because This feature binds word ‘chair’. If we use Chinese directly, it will bind “椅子”.

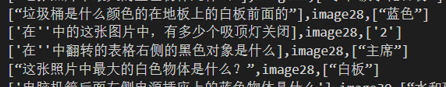


Figure 2: wrong of the QA pairs

Finally, we find that there are not many VQA application scenarios, especially those that can use Chinese. We hope to carry out this group project from two aspects: Chinese application scenario or Chinese VQA construction

# Process

## Data sets

The data set of VQA needs at least three elements: a picture, the question relying on the picture and the answer to the question. At the same time, the answer should not be too complex, and the question must be related to the characteristics of the picture. Because of the needs of QA pairs, general computer vision data sets like COCO, Visual Genome cannot meet the needs of VQA. Therefore, VQA data sets are processed on these data sets, like VQA2.0 and VQA1.0 come from COCO data sets. At present, in VQA data sets, only Baidu artificially generated Chinese QA pairs with COCO data sets in 2015 (FM-IQA), but their use still translates them into English.

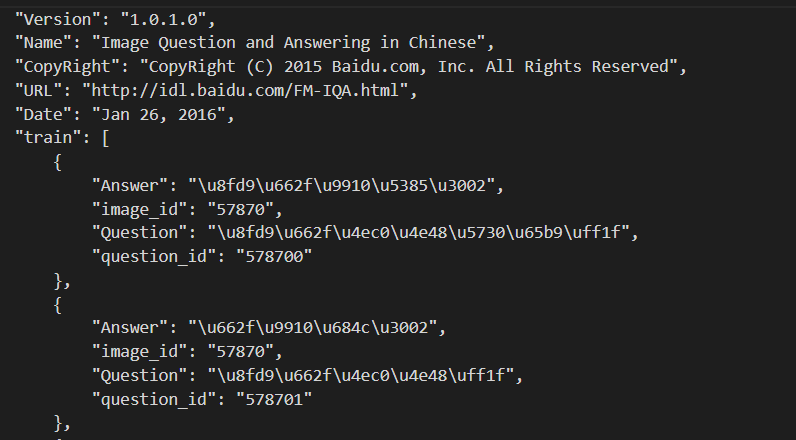


Figure 3: QA pair of FM-IQA, Chinese is load by Unicode

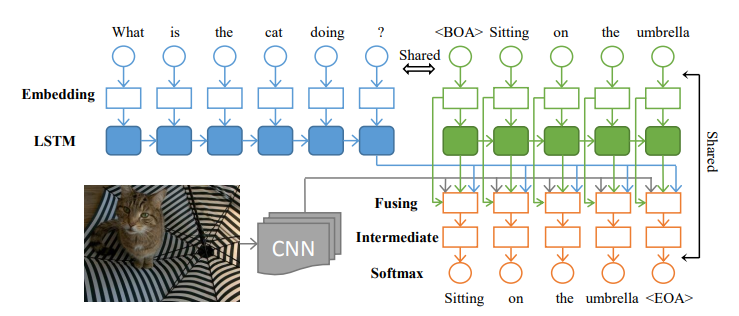


Figure 4: VQA of FM-IQA, Chinese QA pairs is translated to English

## Design Data Sets

At the same time, we compared the FM-IQA database and found that there are quite a lot of complex answers, and many pictures do not correspond to QA. But because the amount of data is huge enough, Baidu runs out with good results. We need a small, accurate, and simple database to facilitate us to improve the test. So, we designed a Chinese VQA database during this period. At the same time, we learned some differences between Chinese and English in the process of our design. Including word segmentation and semantic association.

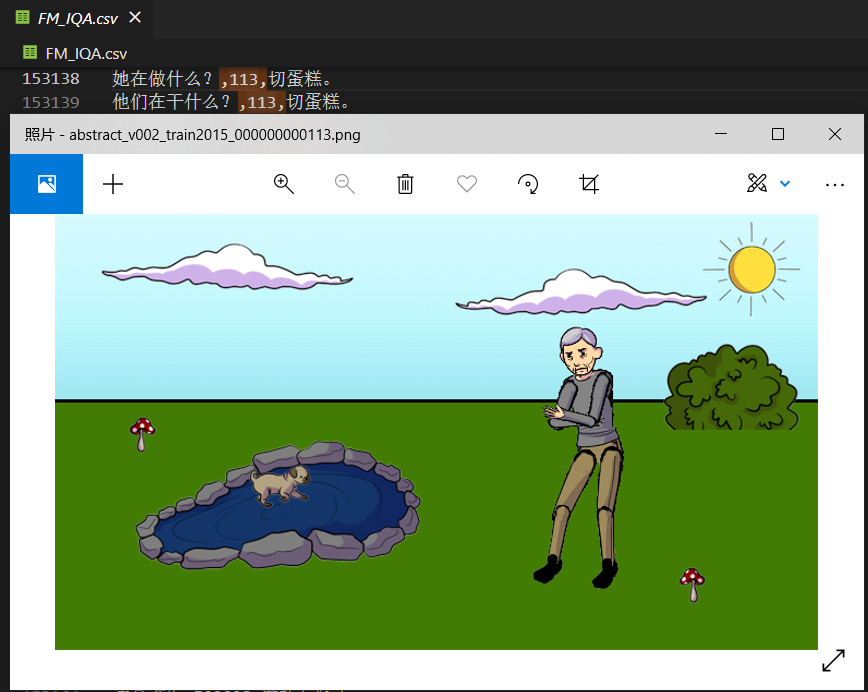


Figure 5: in image113, it gets Irrelevant QA pairs

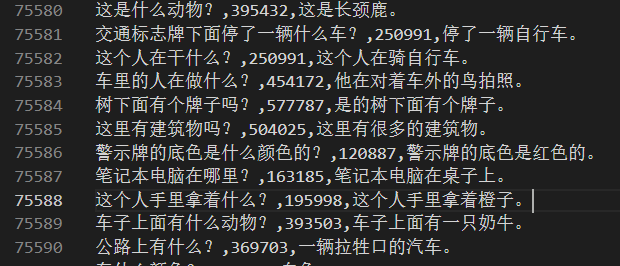


Figure 6: complex answers

For our data set, we use COCO image and English QA pairs of VQA v2. Because COCO-QA is generated by NLP algorithm, and the result is not good, VQA v1 has many kinds of language bias, CLEVR has a lot of reasoning and is too complex.

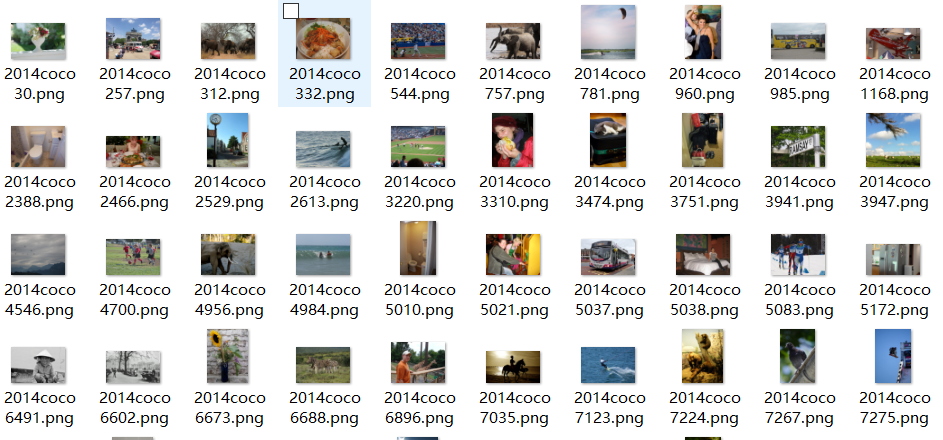


Figure 7: images of our data set

To get Chinese QA-pair, we get the simple QA-pair from VQA v2 firstly. Then I extracted the QA questions of 4000 real pictures and 6000 abstract pictures, and roughly translated them through Baidu API. A Chinese database with many defects is obtained. Then we designed a simple program to quickly proofread and modify these QA pairs.

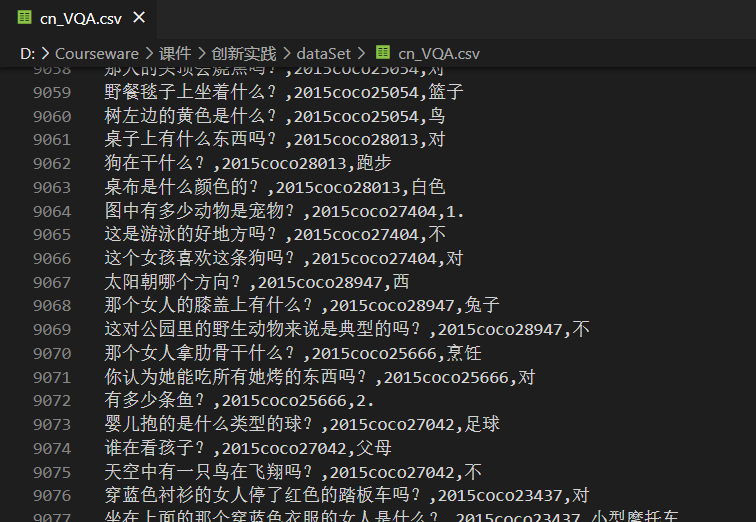


Figure 8: A Chinese database with many defects



Figure 9: the usage of Baidu API



Figure 10: the program to modify these QA pairs

# plan

## Design the model

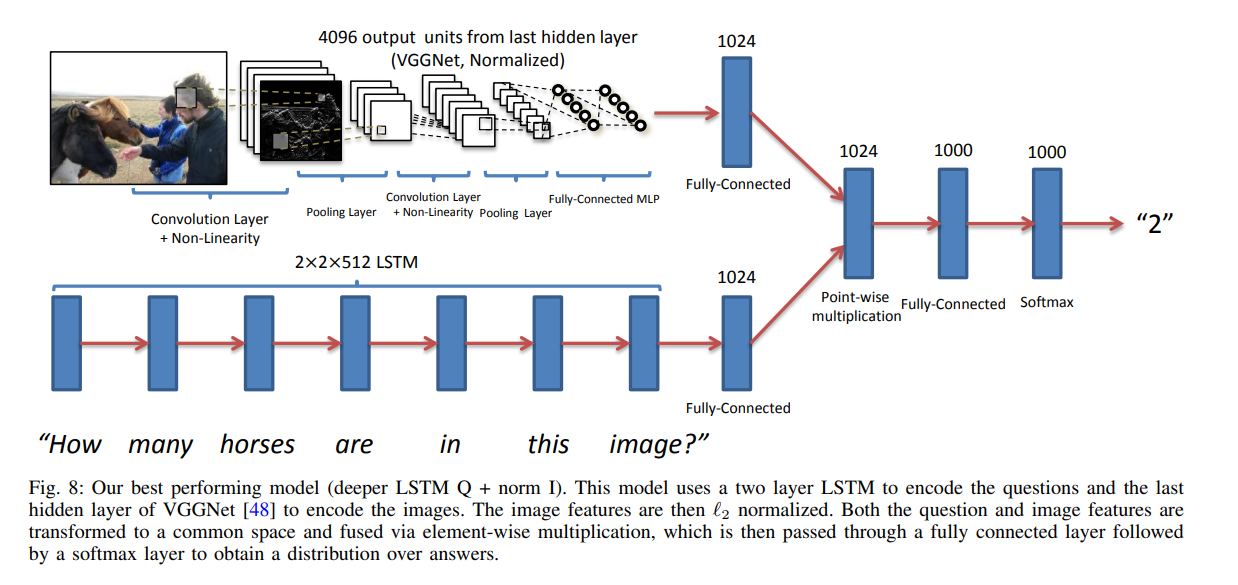


Figure 11: the model of VQA

At present, we are still in the idea stage at this step. We consider embedding and semantic transformation to try these. As shown in the figure 10, we plan to complete the construction of Chinese VQA in the semantic reading of question input and subsequent embedding.

## VQA program design

In fact, even the demo of VQA's official website has many problems. We hope to make an interactive program in Chinese to better apply VQA.

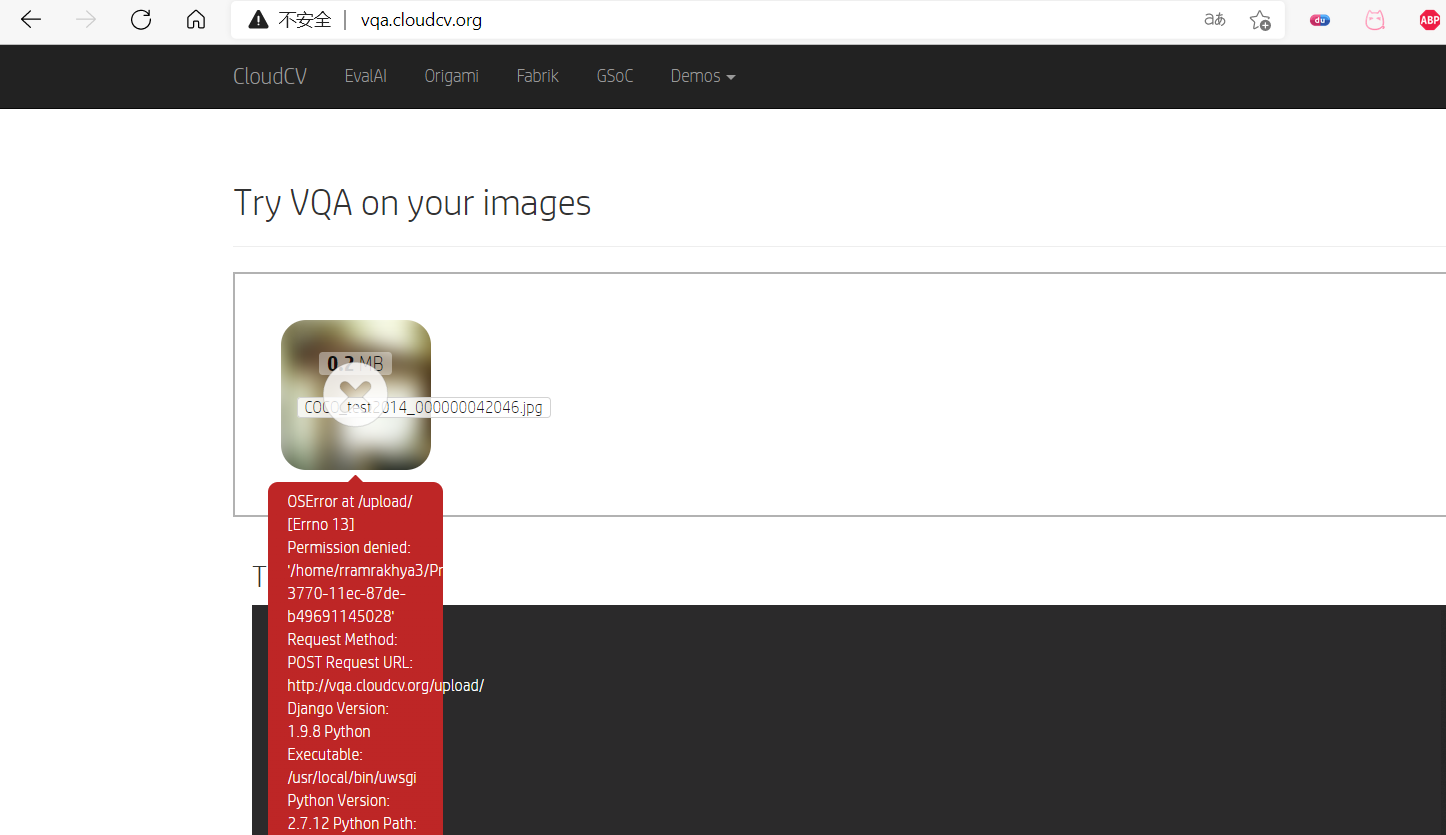


Figure 12: many error in demo of VQA

# Reference

**[1] Zhou, B. , Tian, Y. , Sukhbaatar, S. , Szlam, A. , & Fergus, R. . VQA: Visual Question Answering.**

**[2] H. Gao, J. Mao, J. Zhou, Z. Huang, L. Wang, and W. Xu, \Are you talking to a machine? Dataset and methods for multilingual image question answering," in Advances in Neural Information Processing Systems (NIPS), 2015.**