CS 444 Project Proposal

Yuqing Su, Yuan Su, Shan Huang

TOTAL POINTS

1 / 1

QUESTION 1

1 Project Proposal **1 / 1**

✓ **- 0 pts** *Correct*

Concrete project and reasonable minimum and maximum goals! One thing to note: COCO is not a VQA dataset. You may want to use VQAv2 (https://visualqa.org/download.html) or a subset of it instead.

A Replication study of the combination model of deeper LSTM and CNN dealing with VQA

Group members

Shan Huang - sh69@illinois, Yuqing Su - yuqings3@illinois, Yuan Su - yuansu3@illinois

Project description and goals

This project is about the visual questions answering (VQA). We will 1) run the given code to verify the result in the paper, 2) follow the paper to implement the model from scratch, and 3) compare our model results with the paper's results.

For 1), we will use the code available here and compare it with the results in the paper's result section.

For 2), we will follow the steps in the paper's VQA Baselines and Methods section:

- Implement the baseline models, which include random, prior("yes"), per Q-type prior, and nearest neighbor.
- Implement the image channel and question channel.
  - Image channel: 1) I and 2)normal I
  - Question channel: 1) Bag-of-Words Question, 2) LSTM Q, and 3) deeper LSTM Q
- Implement Multi-Layer Perceptron (MLP) separately for 1) BoW Q + I and 2) LSTM Q+I

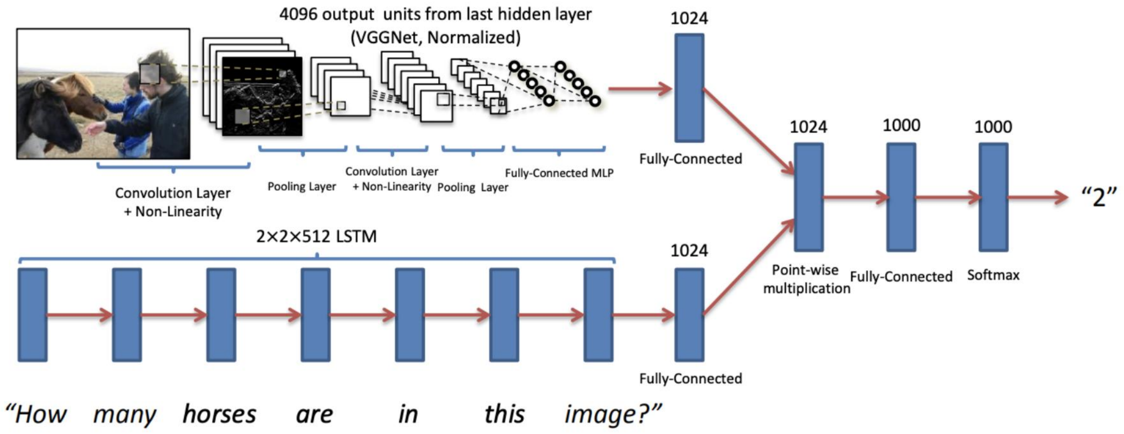

Figure 1. Copy from Figure 8 in the paper that describes the overall network

For 3), we will run our model from 2) and compare the results we gained from 1) with the subset of COCO dataset

Based on the above plan, we have our goals and outcomes as follows:

Minimum goals: Finish 1) and finish 2) and 3) without the baselines comparisons.

Maximum goals: Finish the baseline comparisons. If there is more time, generate our own statistical graphs(eg. bar chart) dataset, train our model and paper's model on the newly created dataset, and evaluate the performance of both models on the new dataset.

Final outcomes: Github repository with code and the final evaluation table.

Resources

Paper: http://arxiv.org/pdf/1505.00468.pdf
Code resource 1:https://github.com/GT-Vision-Lab/VQA
Code resource 2: https://github.com/Cadene/vqa.pytorch

Member roles

This group will meet twice per week for discussion. The model will be implemented collaboratively through Github.

Background Research: All the group members will read through the paper and related topics including Natural Language Processing (NLP), Computer Vision (CV), and Knowledge Representation and Reasoning (KR).

Dataset Processing: Yuan would be in charge of generating a subset of the VQA dataset for training, testing, and evaluation. In addition, he will investigate the dataset to gain an understanding of the types of questions, typical answers, lengths of answers, etc.

Model Implementation:

- **Image channel implementation:** Shan
- **Question channel implementation:** Yuan
- **Multi-Layer Perceptron (MLP) and Model Integration:** Yuqing

Evaluation and data visualization: Shan will be responsible for running the results of our model and paper model, and comparing the two results.

Further study (if more more time):

- **Baselines Implementation(if more time):** Yuan
- **Applying to statistical graphs:** Shan, Yuqing

Relationship to your background

The original paper used a combination of deep learning techniques to perform Visual Question Answering (VQA) which contains Computer Vision (CV), Natural Language Processing (NLP), and Knowledge Representation and Reasoning (KR). The image channel implements VGGNet, which have learned in class. When implementing the baseline of VQA, the authors use multiple other simpler models for comparison such as KNN.

Aside from what we have learned in class, we will also explore techniques that we are not familiar with in NLP and KR. For these parts, we will use the packages already implemented in PyTorch (e.g. torch.nn.LSTM) to implement the model. One of our group members is currently taking CS410:Text Information System which might relate to the NLP.

1 Project Proposal 1 / 1

✓ **- 0 pts** *Correct*

Concrete project and reasonable minimum and maximum goals! One thing to note: COCO is not a VQA dataset. You may want to use VQAv2 (https://visualqa.org/download.html) or a subset of it instead.