Visual Question Answering

present by - Fei Gao, Feiyu Wang, Yifei Zhu
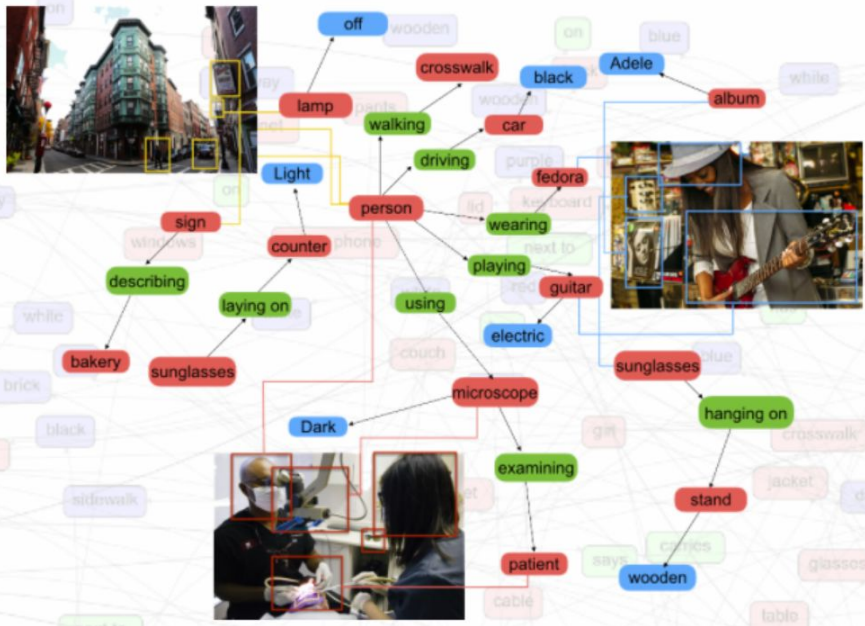
PRODUCT MISSION

The product of VQA is to set for helping the visually impaired identify daily necessities and give them essential instructions.

MINIMUM VIABLE PRODUCT

- Analyze Image: Extract the basic information of the item from the image

- Locate Items: Can identify and locate the required item

- Text Recognition: Recognize and read/process the text infomation from image

- Feedback: Able to answering questions from user

USER STORIES

- The user can use this product with the VQA function to help him identify daily necessities such as microwave food, and the product is able Identify the words on food packaging and give instructions to people how to heat food in a microwave oven.

- When people with impaired vision need to take medicine, the product can also find the words on the medicine bottle such as medicine name, ingredients, expiration date and usage. Give specific answers based on the questions people ask.
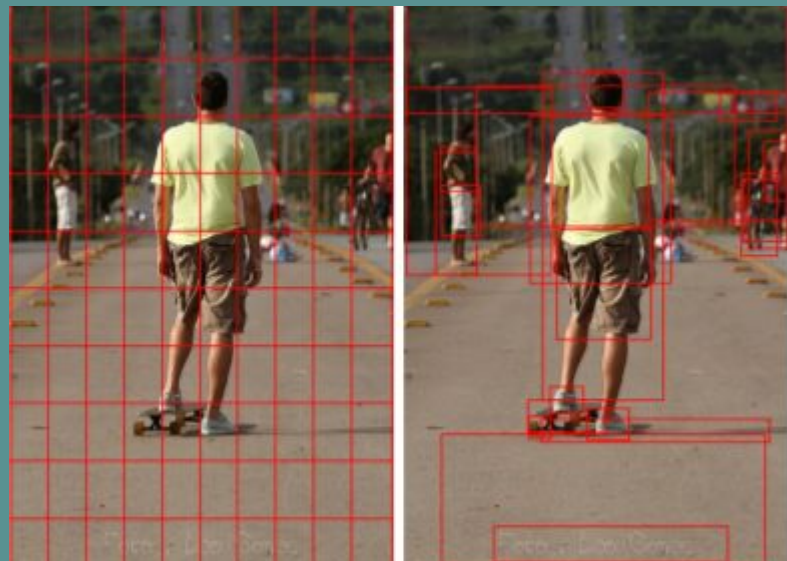
Technologies
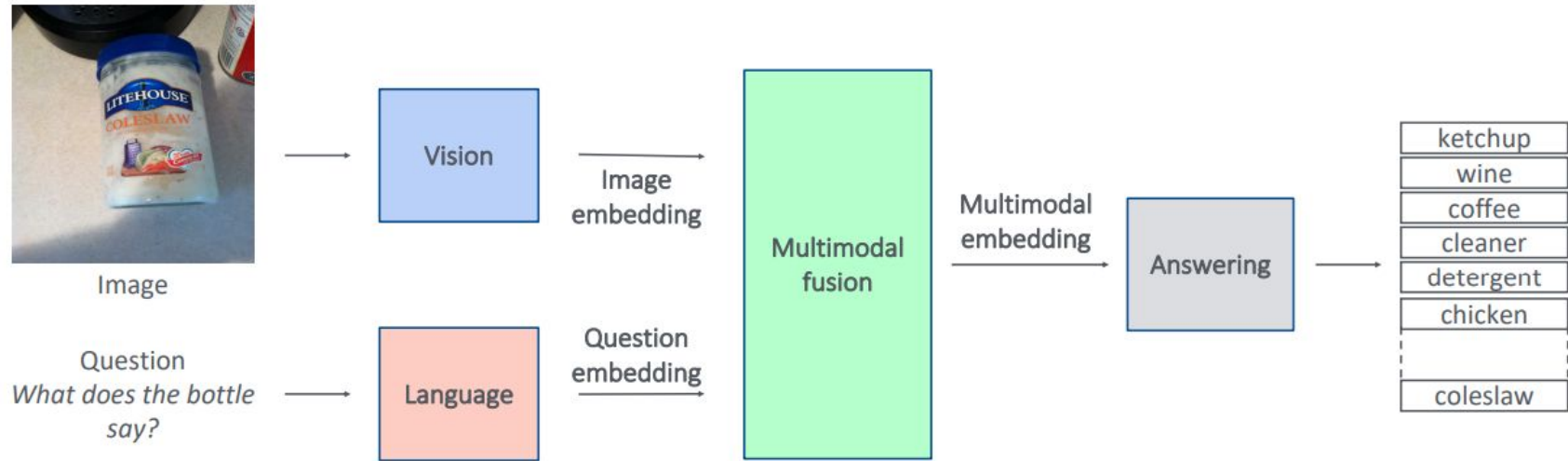
Picture from: VizWiz 2018 Challenge winner (Pythia)

Models for VQA

- Joint embedding approaches

- Attention mechanisms - Bottom-Up Top-Down

  enables attention to be calculated at the level of objects

  and other salient image regions. Focus on objects

- Compositional Models

- Models using external knowledge base

Technologies

- Bottom-up mechanism: Faster R-CNN (based on CNN, the way process imagines based on filters) proposes image regions
  - Bottom-up Attention Model

- Top-down mechanism: use task-specific context to determines feature weightings
  - Top-down Attention Model
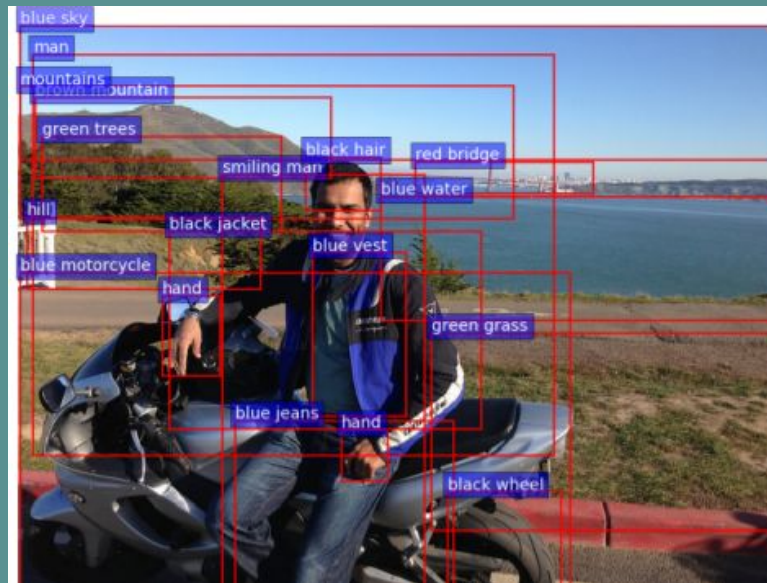   - Captioning Model
    - VQA Model

Technologies

Bottom-up Attention Model
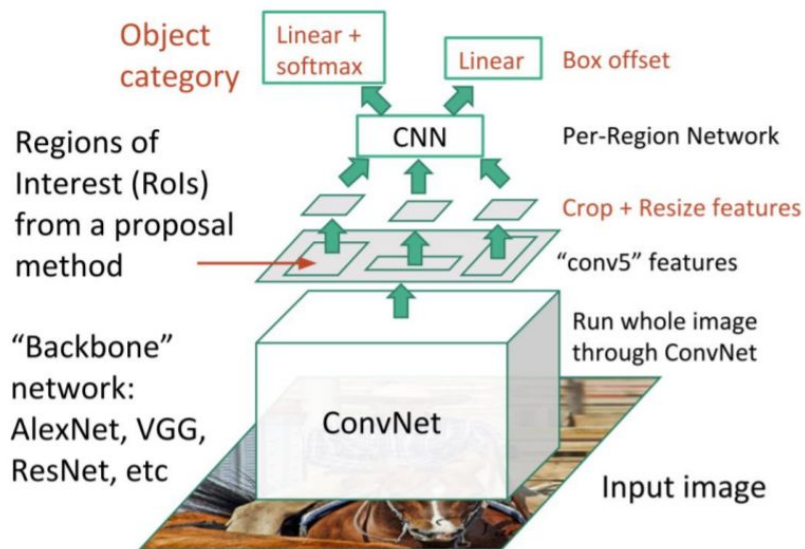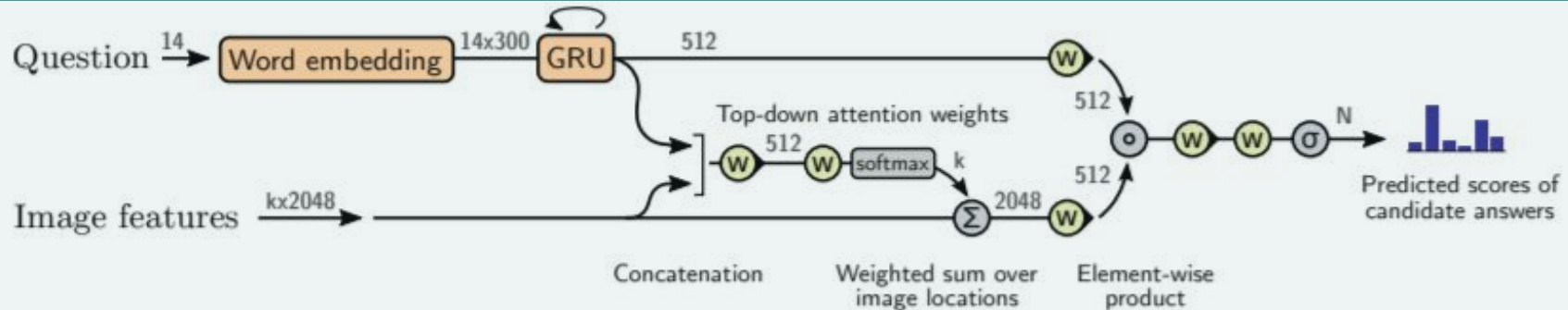
Technologies

Captioning Model

Picture from: Bottom-Up and Top-Down Attention for
Image Captioning and Visual Question Answering

Technologies

VQA Model