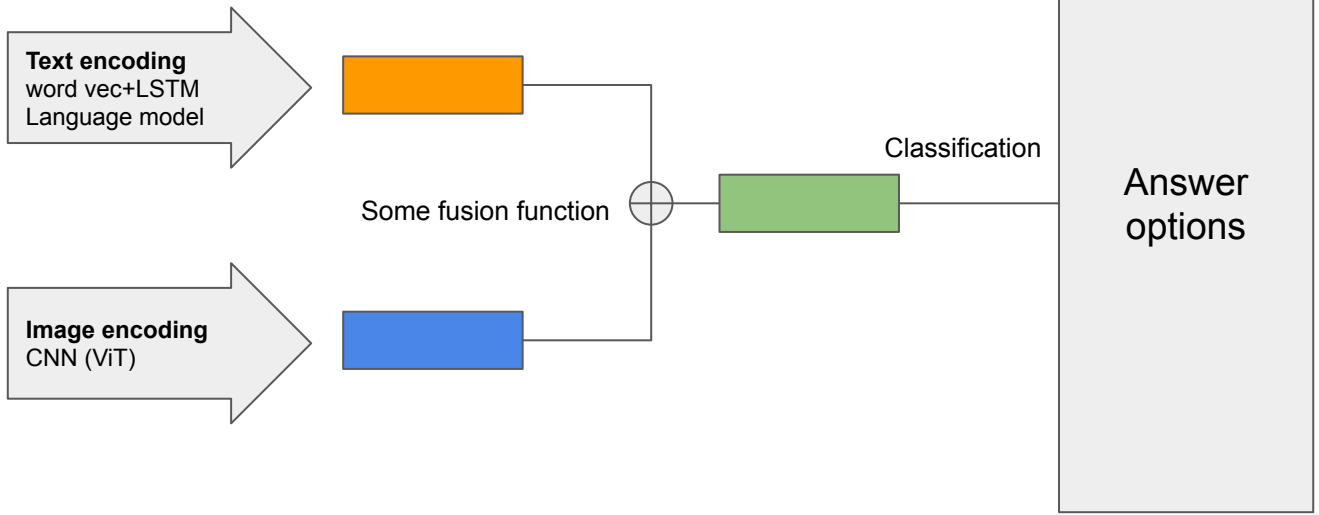
VQA Approach & Dataset Survey

Sungguk Cha

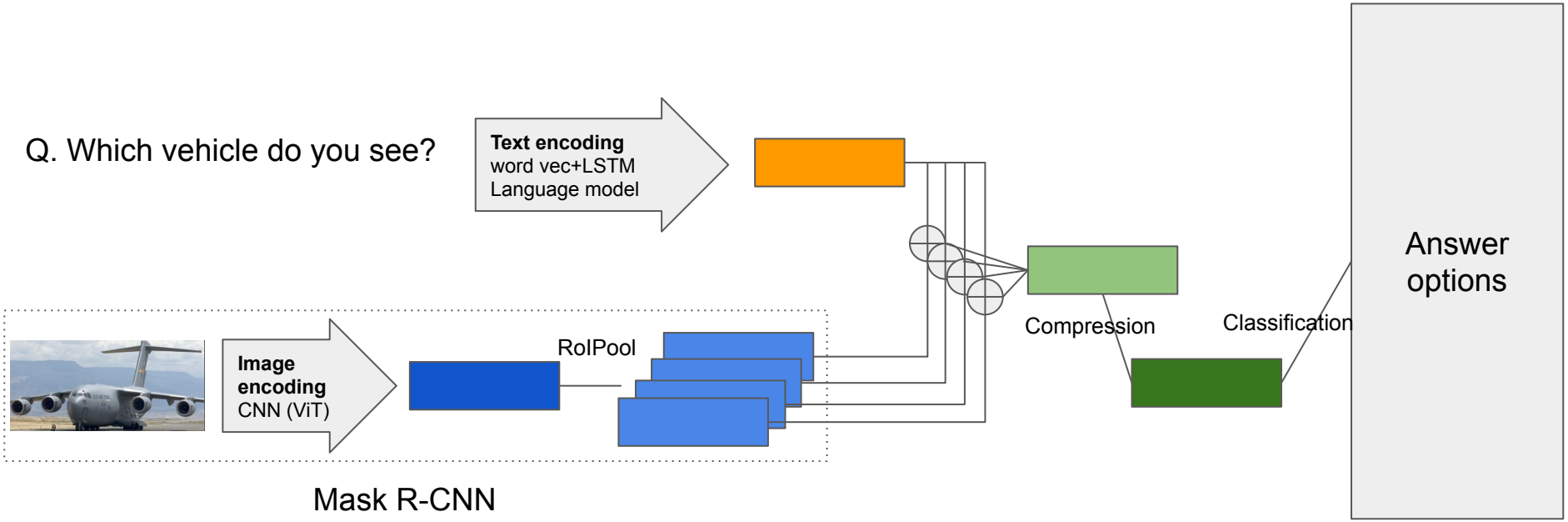
Contents

VQA approach overview

Q. Which vehicle do you see?



Bottom-up VQA Approach



VQA Datasets

- Visual Genome
  - 108K images, 1.7M question-answers
  - densely annotated with **graphs** containing object **attributes** and **relationships**

- VQA v2
  - Balanced Real Images
    - 204,721 COCO images
  - 1,105,904 questions
  - 11,059,040 ground truth answers
  - open-ended questions about images
  - require an understanding of vision, language and **commonsense** knowledge to answer.

VQA Datasets

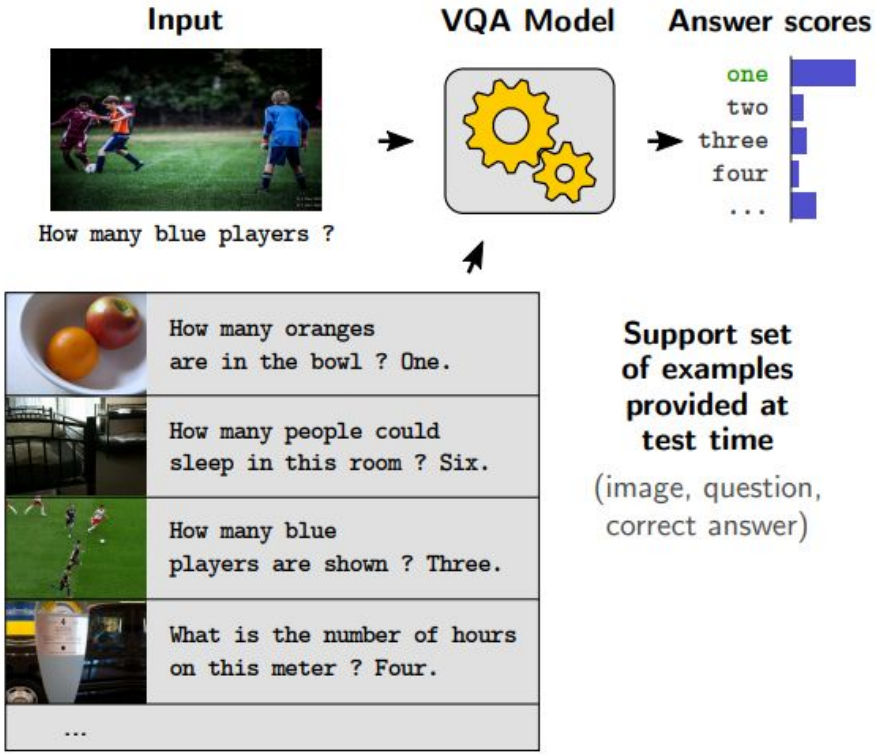
- VQA v2
  - VQA Number
  - VQA CP
  - …
- CLEVR
  - Train: 70,000 images, 699,989 questions
  - Val: 15,000 images, 149,991 questions
  - Test: 15,000 images, 14,988 questions
  - Scene **graph** annotations: locations, attributes and relationships

Image Captioning

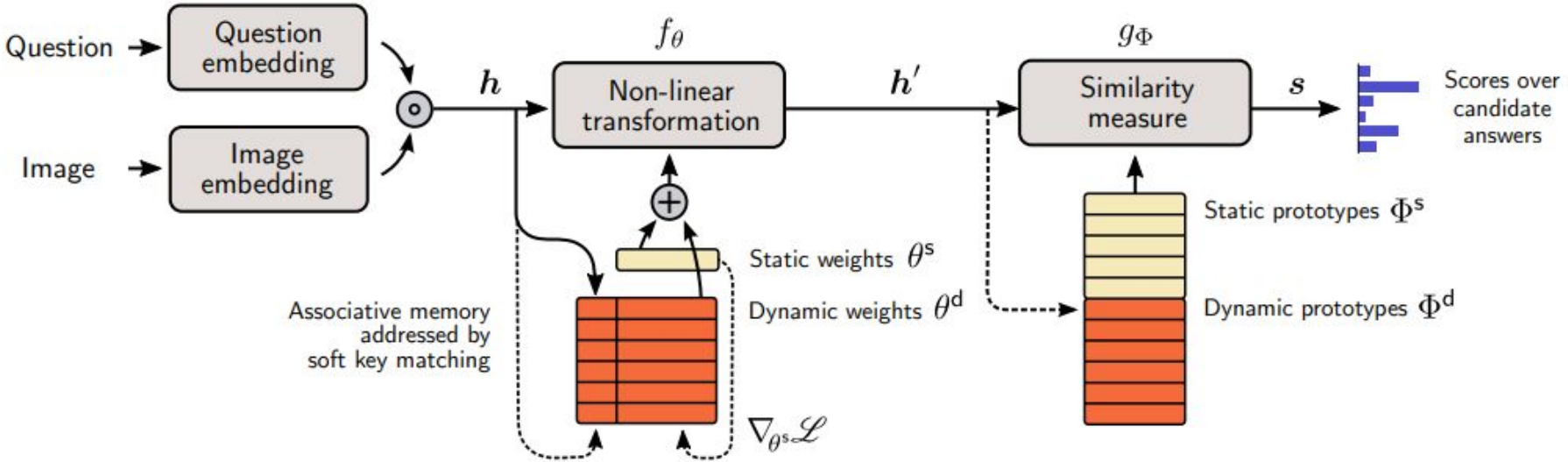
- MS COCO 2014
  - 113,287 training images with five captions each
  - 5k val/test images

References

Visual Question Answering as a Meta Learning Task
Damien Teney and Anton van den Hengel
ECCV 2018

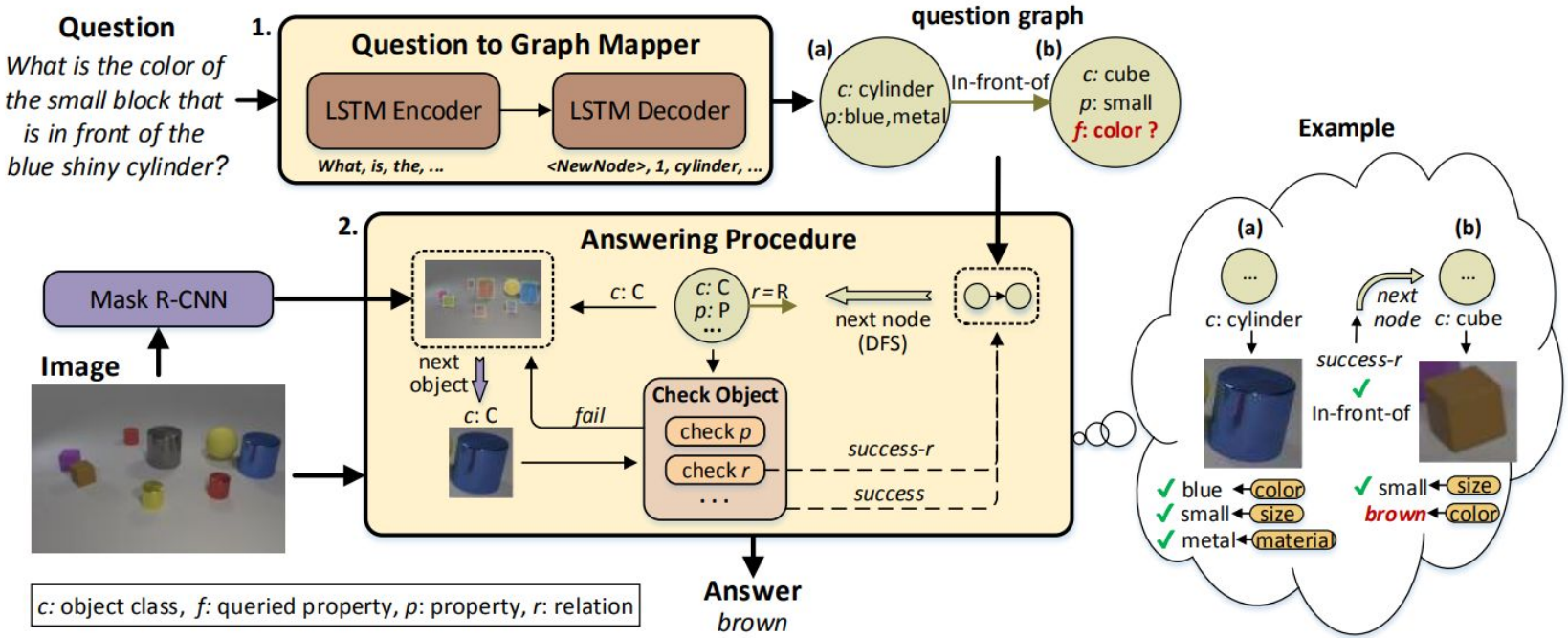


Visual Question Answering as a Meta Learning Task
Damien Teney and Anton van den Hengel
ECCV 2018

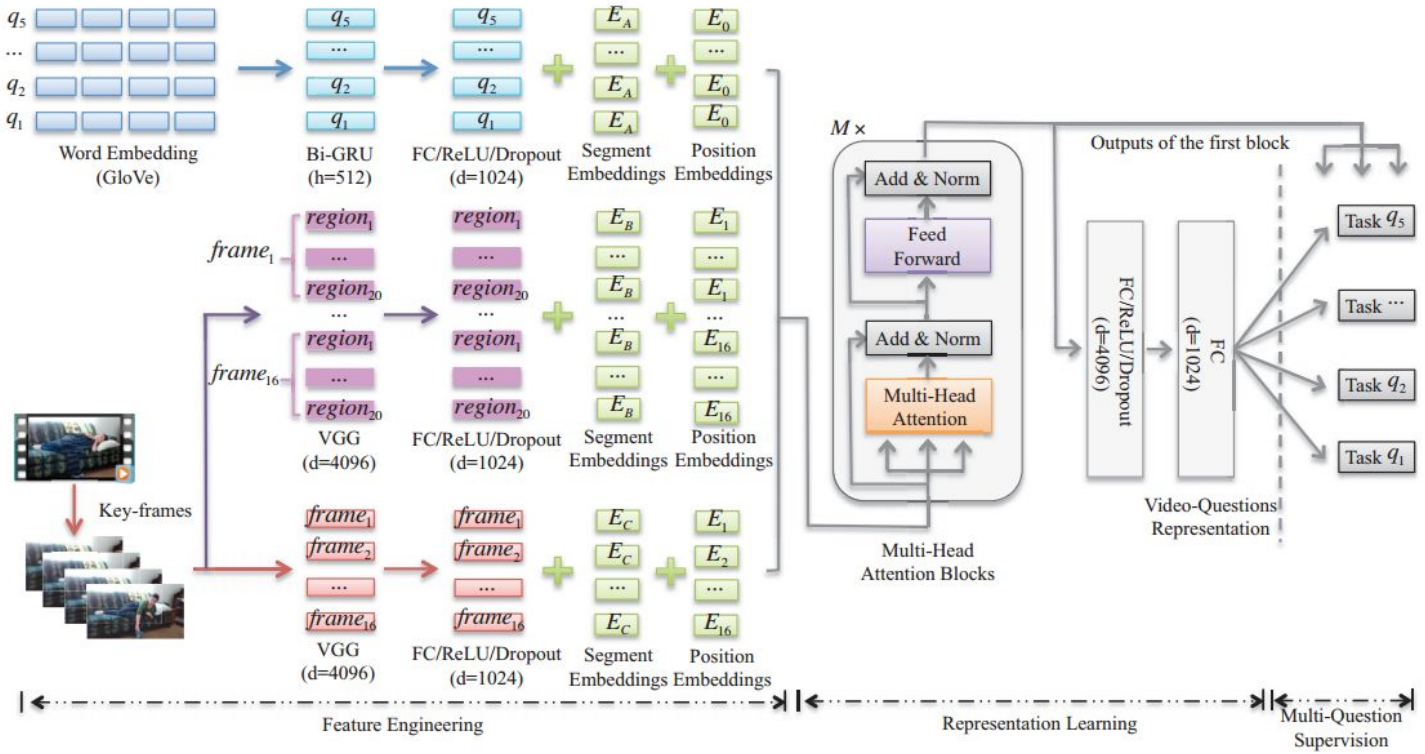


VQA with No Questions-Answer Training
Ben-Zion Vatashsky and Shimon Ullman
CVPR2020

너무 domain specific하고 우리의 목표와 다름



Multi-Question Learning for Visual Question Answering
Chenyi Lei *et al.*
AAAI2020



Multi-Question Learning for Visual Question Answering
Chenyi Lei *et al.*
AAAI2020