

CO-ATTENTION MECHANISM WITH MULTI-MODAL FACTORIZED BILINEAR POOLING FOR MEDICAL IMAGE QUESTION ANSWERING

Volviane S. Mfogo,^{1,2} Georgia Gkioxari,³ Xinlei Chen,³ Jeremiah Fadugba.¹

¹African Institute for Mathematical Science/African Masters in Machine Intelligence (AIMS/AMMI), Accra, GHANA, ²Department of Mathematics and Computer Sciences, University of Dschang, and ³Facebook AI Research (FAIR), USA.

MOTIVATION

- ▶ Lack of medical specialist
- ▶ General Practitioner need AI for more accurate diagnostic

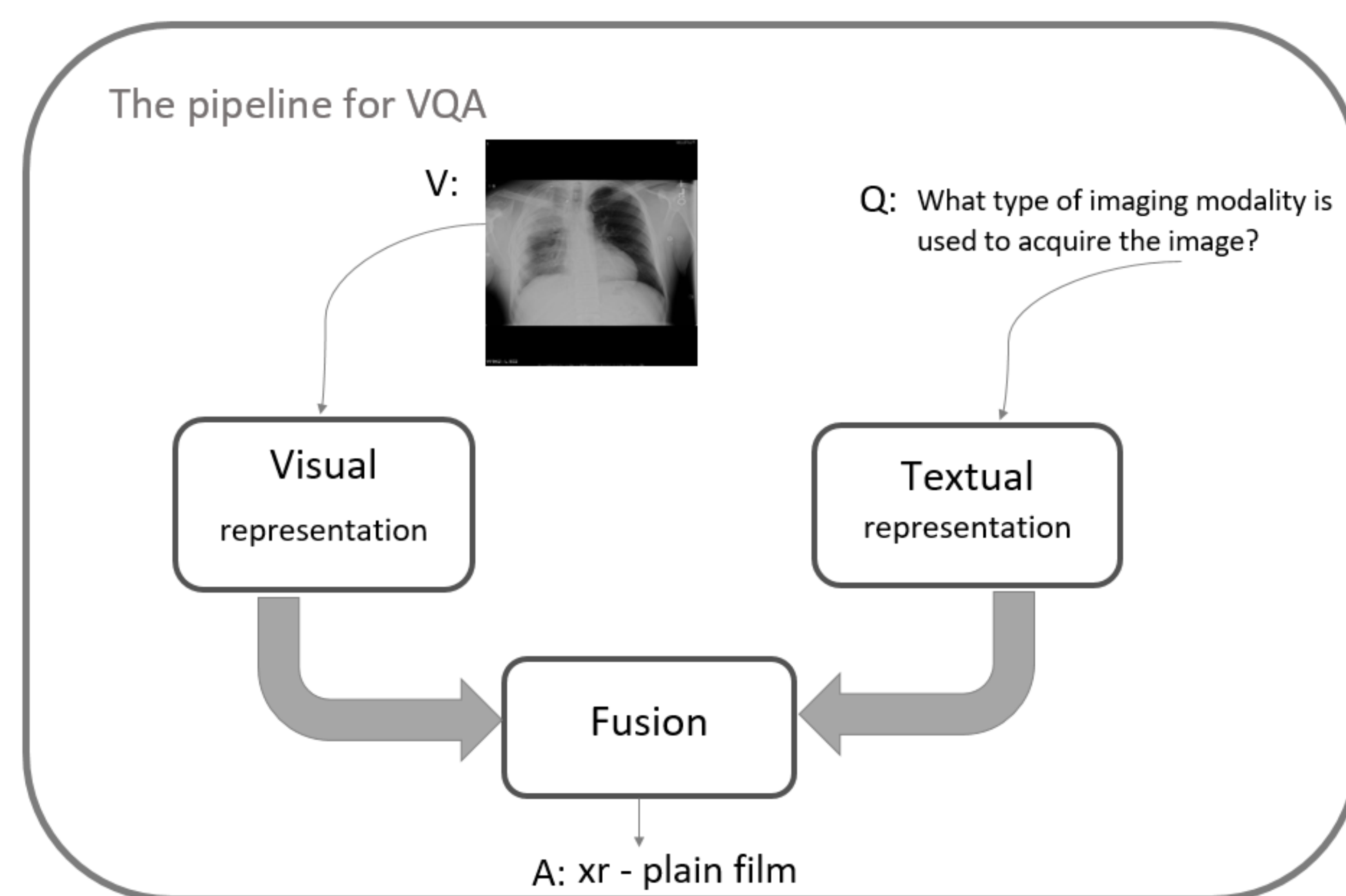


VQA-MED DATASET

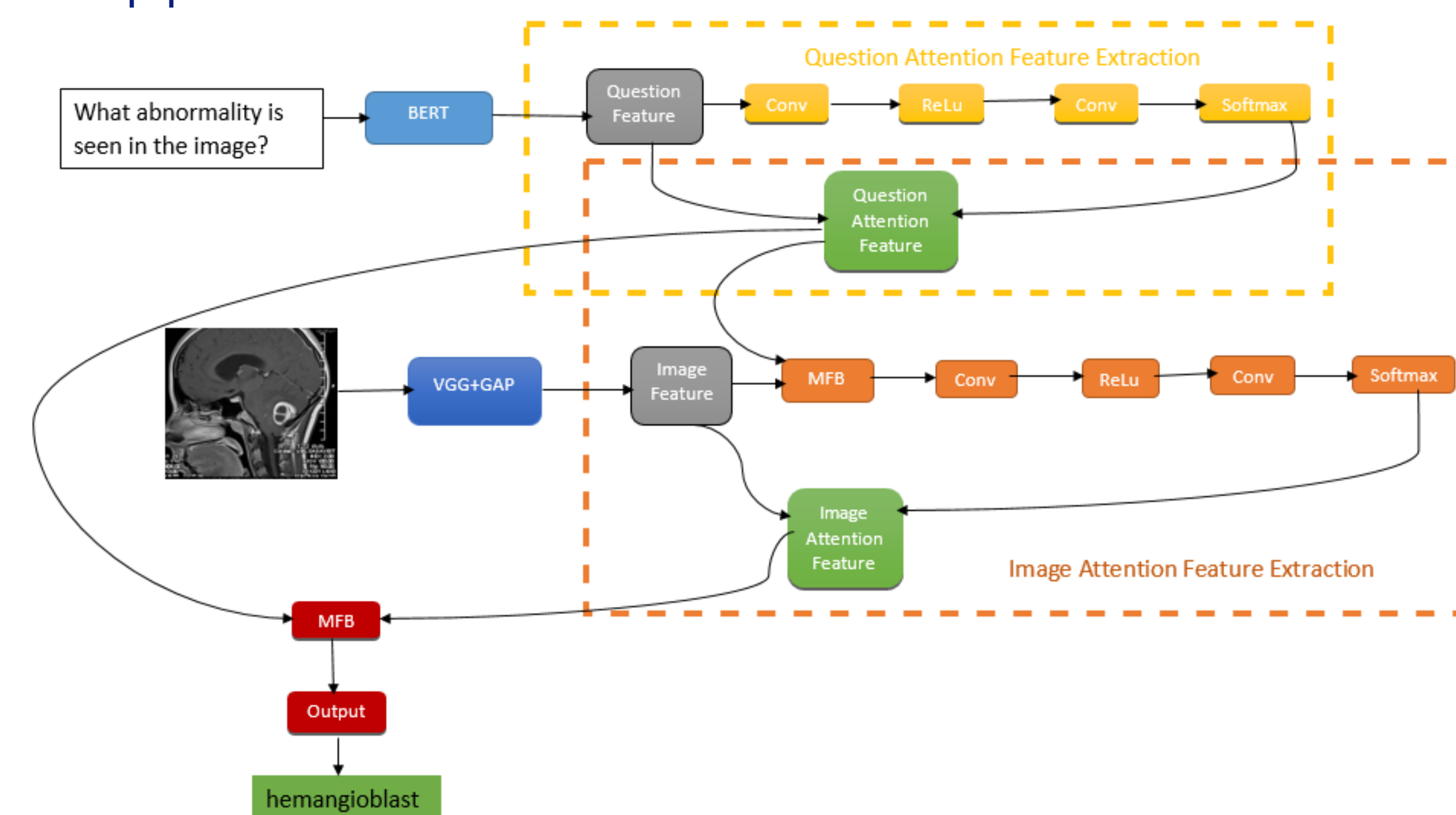
- ▶ Applying several filters to select relevant images and associated annotations;
- ▶ Creating patterns to generate the questions and their answers.
- ▶ The test set was manually validated by two medical doctors.
- ▶ The dataset is publicly available.

Image		
Question	what modality was used to take this image?	in what plane was this image taken?
Answer	mammograph	coronal

MODEL



The pipeline for VQA



MFB with Co-Attention network architecture for VQA.

RESULTS

Team Halim (Fukui et al., 2016) and team UMMS (Shi, Liu, and Rosen, 2019)

Model	Modality	Plane	Organ	Abnormality	Overall
Team Halim (State-of-the-art)	0.808	0.768	0.736	0.184	0.624
Team UMMS (fifth at VQA-Med 2019)	0.672	0.760	0.736	0.096	0.566
Our model	0.862	0.752	0.687	0.088	0.597

- ▶ The good performance on Modality question is due to the fact that the number of class here is reasonable (44 classes) and the frequency of each class on the dataset is at least 43 times, which make the model learn to know how to predict each of the class answer.

REFERENCES

- Fukui, Akira et al. (2016). "Multimodal compact bilinear pooling for visual question answering and visual grounding". In.
- Shi, Lei, Feifan Liu, and Max P Rosen (2019). "Deep Multimodal Learning for Medical Visual Question Answering.". In.

ACKNOWLEDGEMENTS

AIMS, AMMI, FAIR, Deep Learning Indaba, IDRC, AI4D

