Question-Agnostic Attention for Visual Question Answering

Presenter: Moshiur Farazi

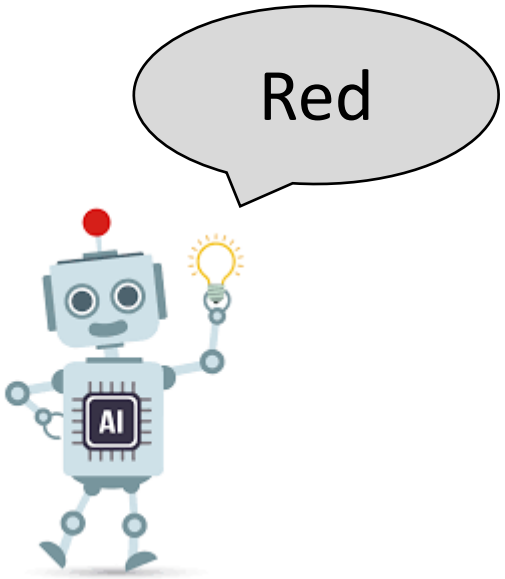
PS T3.4, DAY 2 – January 13, 2021, 12:30 PM CET

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Authors:
Moshiur Farazi, ANU and Data61-CSIRO, Canberra, Australia
Salman H Khan, Mohamed bin Zayed University of AI, Abu Dhabi, UAE
Nick Barnes, ANU, Canberra, Australia

Image: Official White House Photo by Pete Souza, Aug. 9, 2010

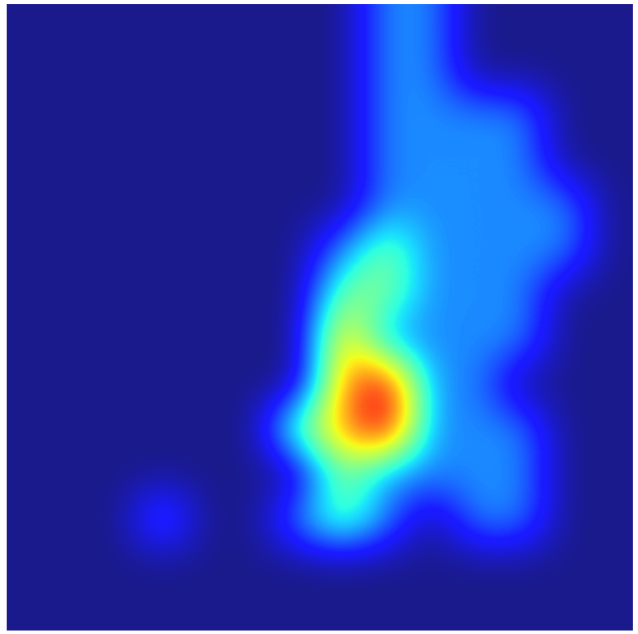
Visual Question Answering (VQA)



What color is the women's shirt?

Image from Visual Genome Dataset [Krishna et al., 2016]

Question Guided Attention



VQA Model

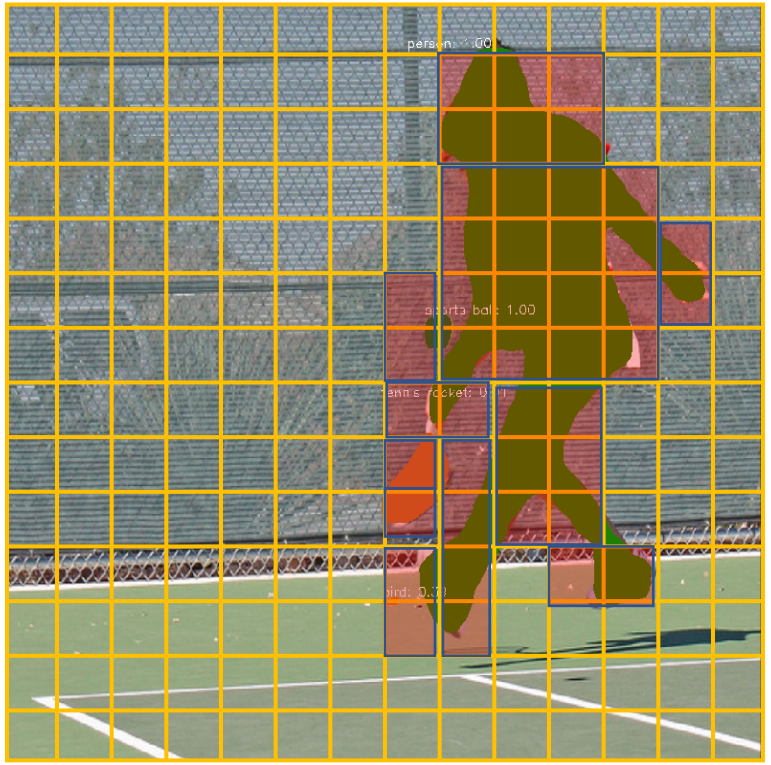
Red

What color is the women's shirt?

Question Agnostic Attention

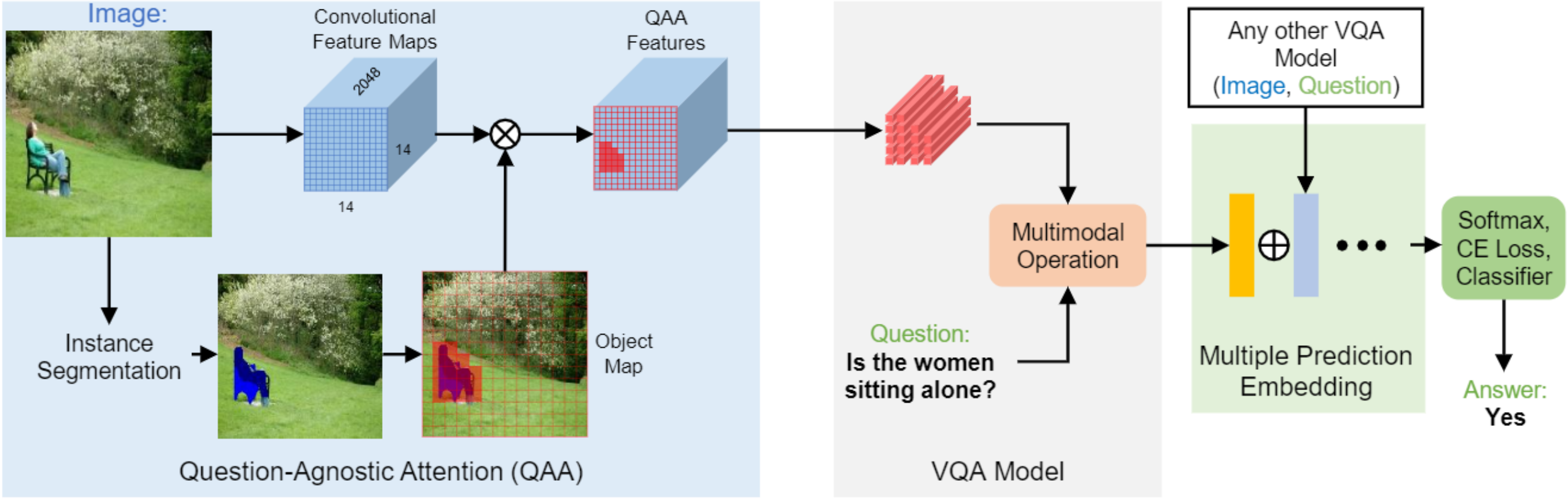
Original Image

Segmentation



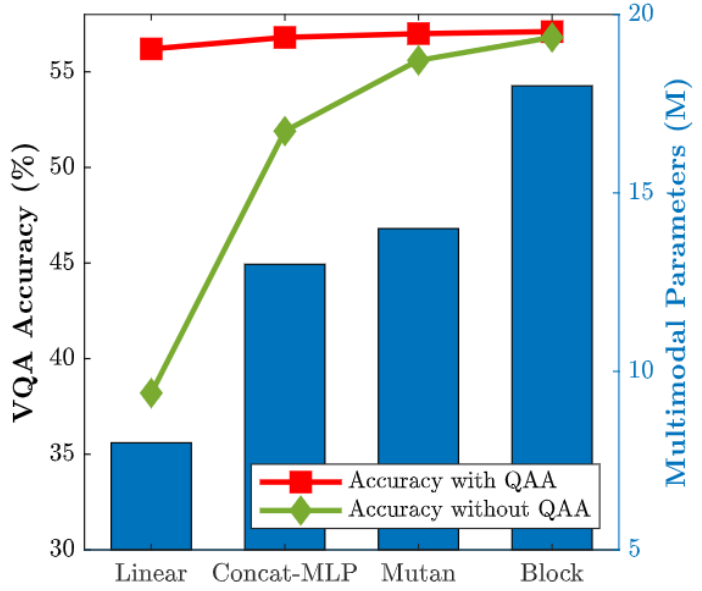
Object Map

Question Agnostic Attention (QAA)



Simplistic VQA models get a significant performance boost with QAA

	Visual Feature	Spatial Attention	VQAv1 Dataset Multimodal Operation Linear	C-MLP	Mutan	Block	VQAv2 Dataset Multimodal Operation Linear	C-MLP	Mutan	Block
(1)	Spatial Grid (SG)	✗	39.7	57.2	56.3	58.2	38.2	51.9	55.6	56.8
(2)	QAA	✗	41.4	40.5	57.3	58.4	39.7	53.2	56.3	56.5
(3)	Ours(SG+QAA)	✗✗	**57.9**	**58.3**	**57.5**	**58.4**	**56.2**	**56.8**	**57.0**	**57.1**
			18.2 ↑	1.1 ↑	1.2 ↑	0.2 ↑	18.2 ↑	4.9 ↑	1.4 ↑	0.3 ↑
(4)	Spatial Grid (SG)	✓	41.8	60.4	58.6	61.2	41.0	54.4	57.9	60.1
(5)	QAA	✓	41.4	59.6	57.9	60.5	37.3	57.3	56.5	59.3
(6)	Ours(SG+QAA)	✓✓	**60.6**	**60.7**	**59.2**	**61.6**	**59.1**	**59.5**	**58.2**	**60.5**
			18.8 ↑	0.3 ↑	0.6 ↑	0.3 ↑	18.1 ↑	5.2 ↑	0.3 ↑	0.4 ↑
	Multimodal Parameters		8M	13M	14M	18M	8M	13M	14M	18M



Summary

- Question agnostic attention can be used in complement with most VQA models.

- QAA helps simple VQA model achieve SOTA performance.

- Object Maps inferred using QAA can be helpful for other vision and language tasks.

Poster Session

T3.4, DAY 2 – January 13, 2021, 12:30 PM CET