LANGUAGE AND VISUAL RELATIONS ENCODING FOR VISUAL QUESTION ANSWERING

Fei Liu, Jing Liu, Zhiwei Fang, Hanqing Lu

Institute of Automation, Chinese Academy of Sciences, Beijing, China

Introduction

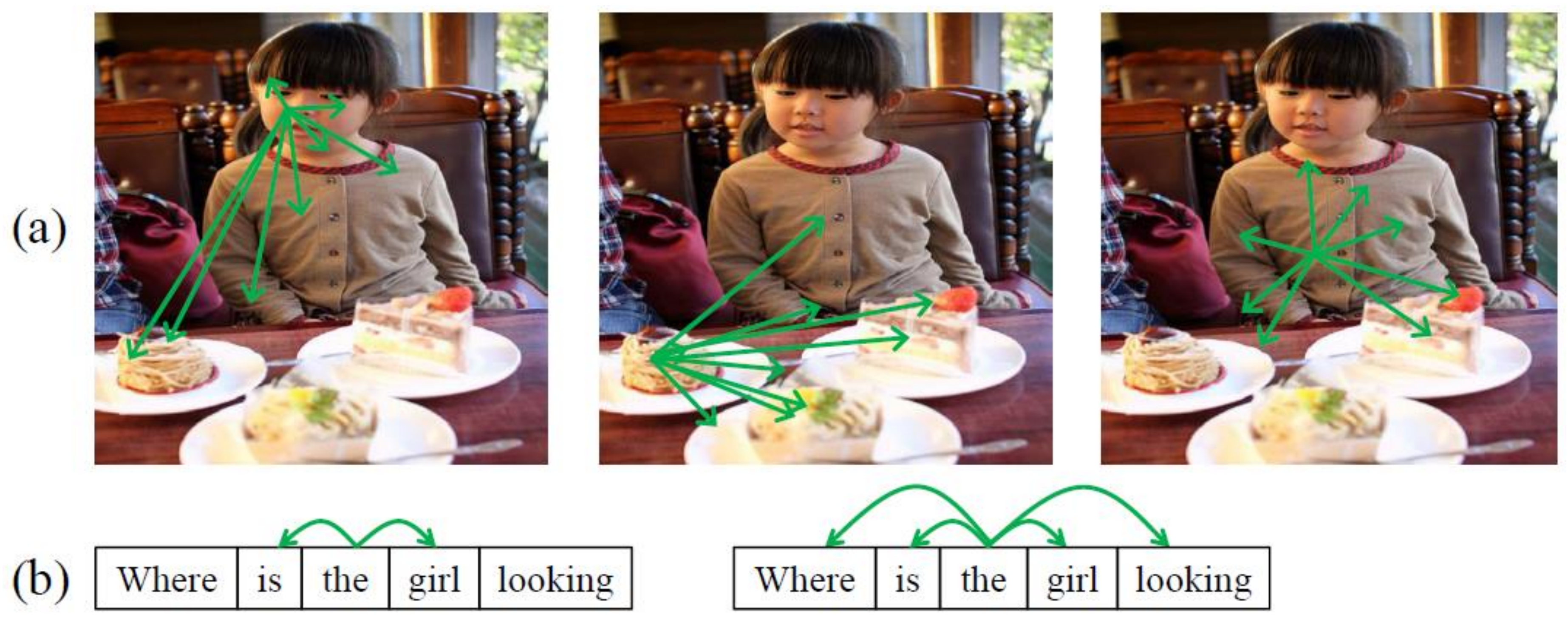
● Problem Definition

Visual Question Answering (VQA) aims at answering a natural language question about a given image.

● Contributions

1. Propose two novel modules to encode relations between words and between image regions, respectively. This is the first time to explore the relations between words and between image regions in a unified framework for the VQA task.

2. Extensive experiments show the effectiveness of the proposed relation encoding modules. Our approach achieves new state-of-the-art results on the VQA 1.0 dataset.

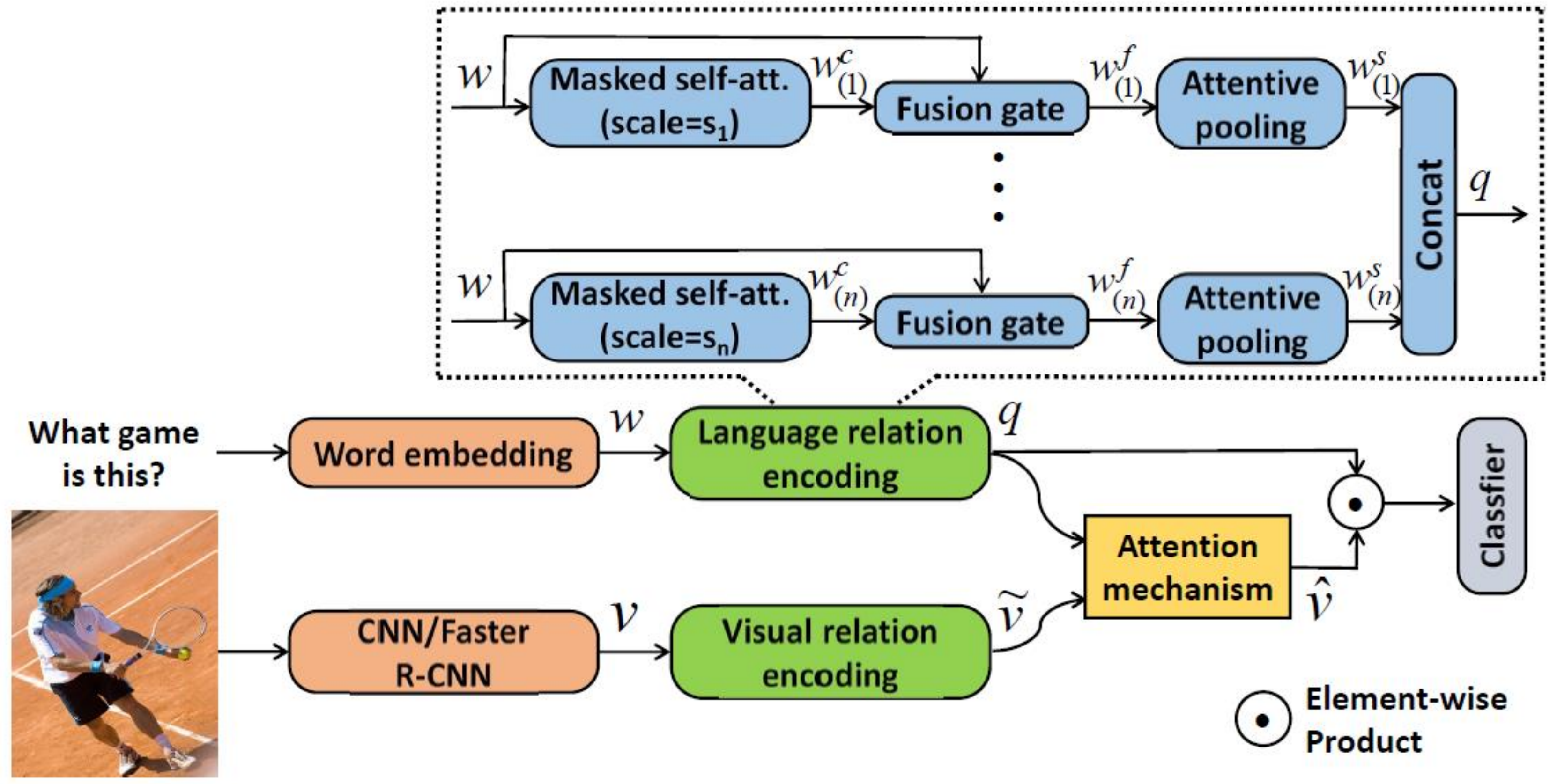


(a) Visual relation encoding. It encodes the relations between image regions.

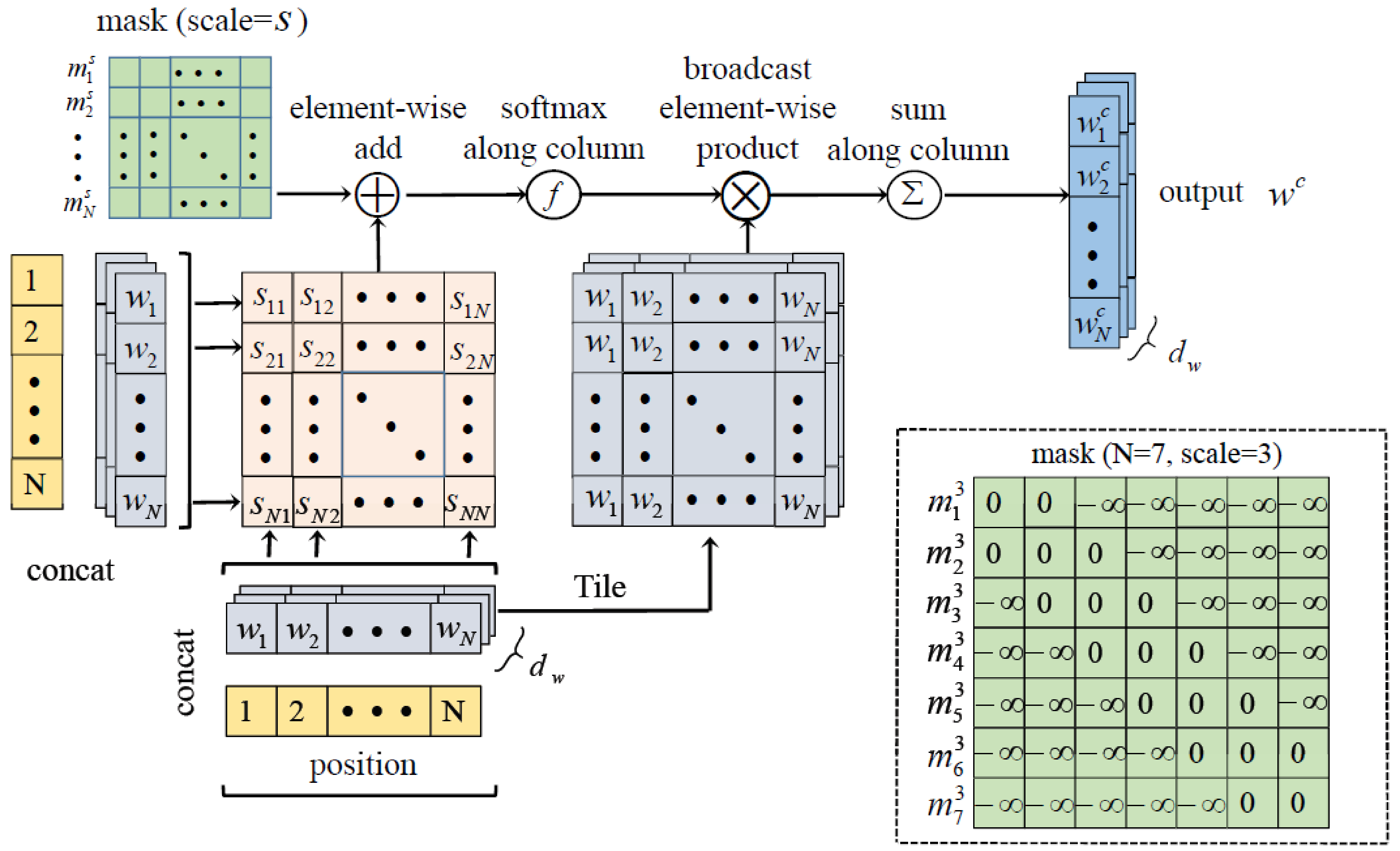
(b) Language relation encoding. It encodes multi-scale relations between words.

Approach

● Overview of the proposed VQA model



● Masked self-attention



mask (N=7, scale=3)

m^3_1	0	0	−∞	−∞	−∞	−∞	−∞
m^3_2	0	0	0	−∞	−∞	−∞	−∞
m^3_3	−∞	0	0	0	−∞	−∞	−∞
m^3_4	−∞	−∞	0	0	0	−∞	−∞
m^3_5	−∞	−∞	−∞	0	0	0	−∞
m^3_6	−∞	−∞	−∞	−∞	0	0	0
m^3_7	−∞	−∞	−∞	−∞	−∞	0	0

Experiments

● Comparison with the state-of-the-arts

Model	Test-dev				Test-std
	All	Other	No.	Y/N	All
QGHC [17]	65.9	57.1	38.1	83.5	65.9
VKMN [18]	66.0	57.0	37.9	83.7	66.1
MFH [19]	66.8	57.4	39.7	85.0	66.9
DCN [20]	66.9	57.3	42.4	84.6	67.0
DA-NTN [21]	67.9	58.6	41.9	85.8	68.1
CoR [22]	68.4	59.1	**44.1**	85.7	68.5
Ours	67.2	57.5	40.6	85.6	67.4
Ours + BU	**69.1**	**59.5**	**44.1**	**86.8**	**69.3**

● Ablation studies

Model	Accuracy
Our model	62.9
Our model w/o position information	62.6
Our model w/o masked self-attention	61.8
Our model w/o fusion gate	62.5
Our model w/o attentive pooling	62.5
Our model w/o visual relation encoding	62.0

More experiments can be found in our paper.

Contact us: Fei Liu (liufei2017@ia.ac.cn)
Jing Liu (jliu@nlpr.ia.ac.cn)
Zhiwei Fang (zhiwei.fang@nlpr.ia.ac.cn)
Hanqing Lu (luhq@nlpr.ia.ac.cn)