Overcoming Language Priors for Visual Question Answering via Loss Rebalancing Label and Global Context

Runlin Cao[1,2], Zhixin Li*[1,2]

[1]Key Lab of Education Blockchain and Intelligent Technology, Ministry of Education, Guangxi Normal University, Guilin 541004, China

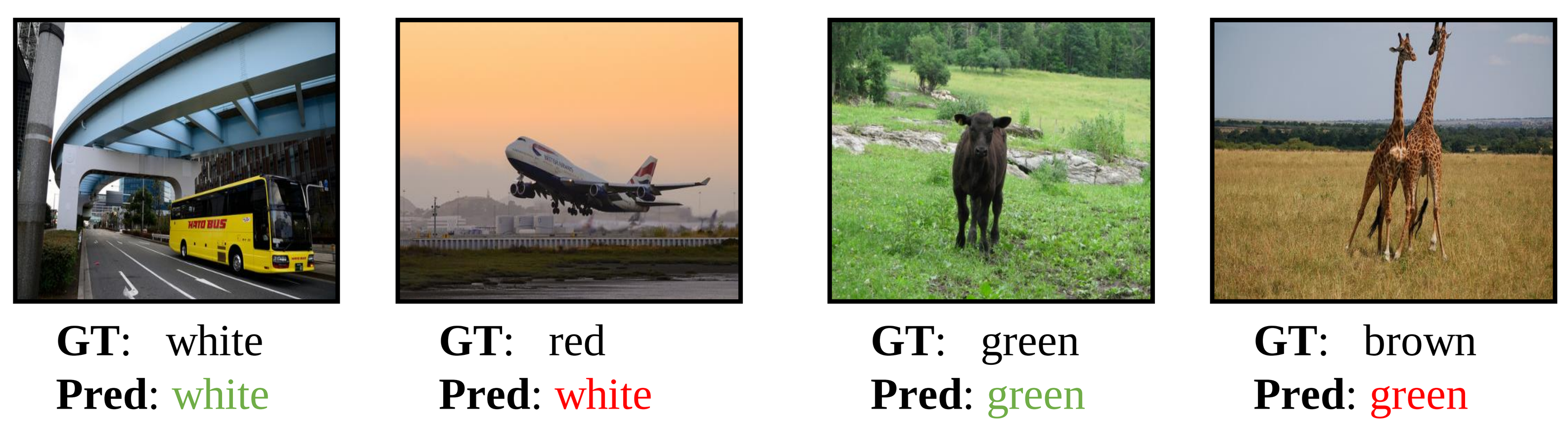
[2]Guangxi Key Lab of Multi-source Information Mining and Security, Guangxi Normal University, Guilin 541004, China

* Zhixin Li is the corresponding author

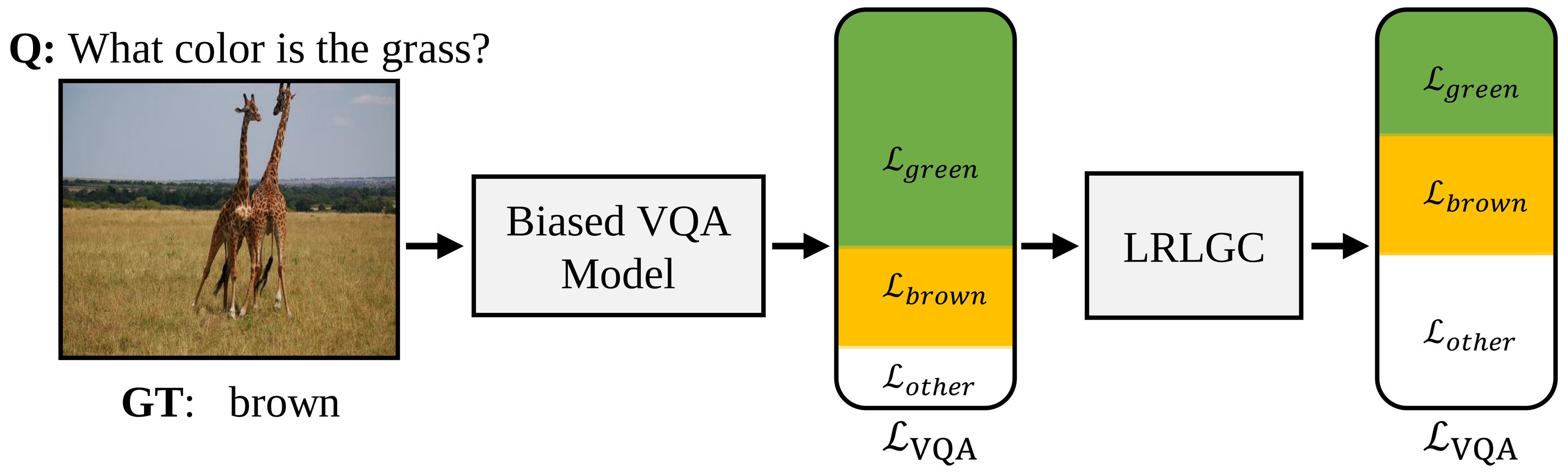
Language priors in VQA

- Despite the advances in Visual Question Answering (VQA), many VQA models currently suffer from language priors (i.e. generating answers directly from questions without using images)
- LRLGC can overcome the class imbalance in the VQA dataset by rescaling the total VQA loss to a more balanced form.

Question: What color is the sky? **Question**: What color is the grass?



(A)



(B)

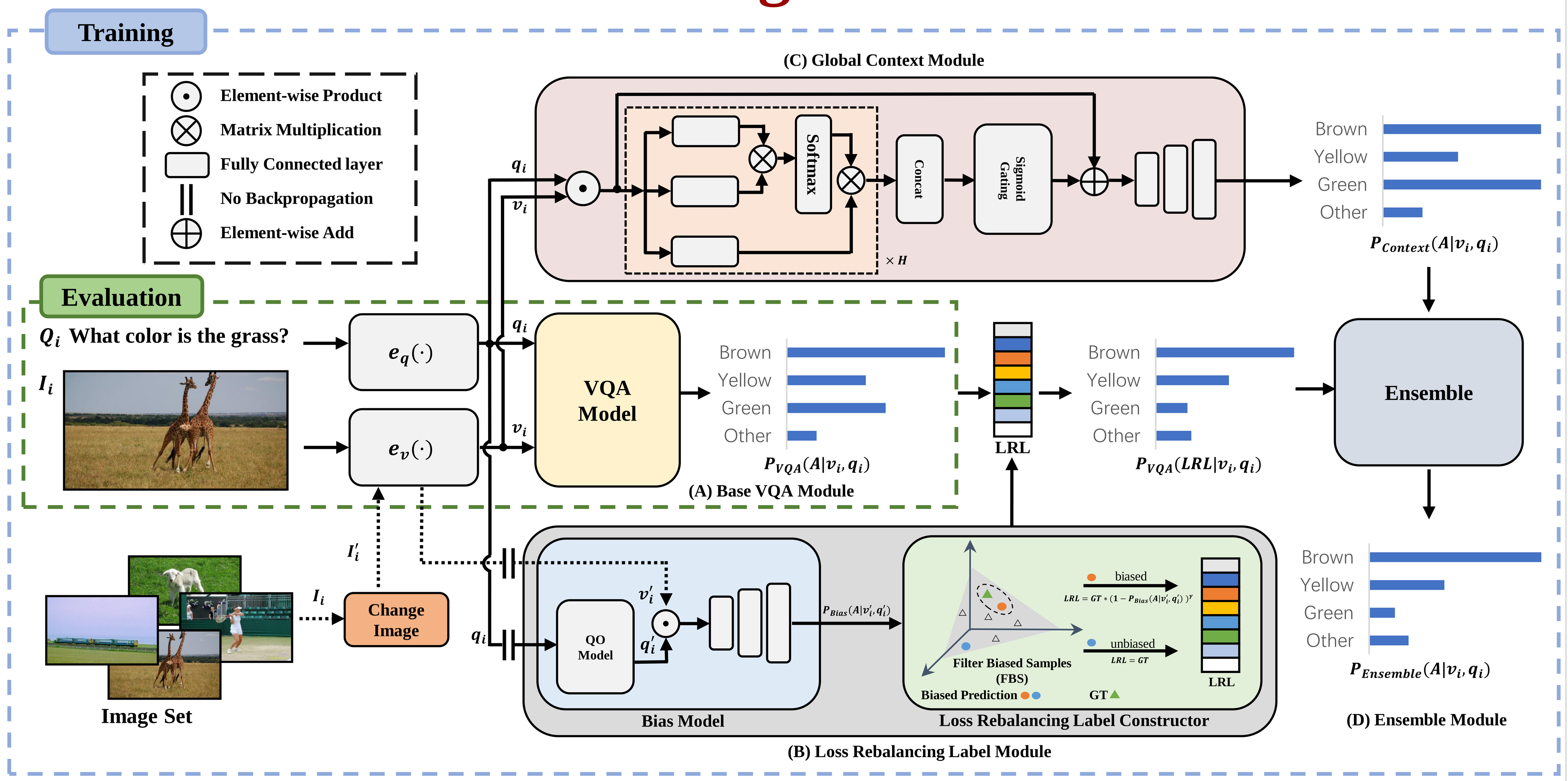
Our contributions:

- We propose a novel model-agnostic generic framework LRLGC that enables end-to-end training and can be easily integrated into various VQA models.
- We propose LRL and Global Context Module, which can effectively help the model overcome the language priors while preserving the contextual information.
- Experimental results show that LRLGC achieves competitive performance on the bias-sensitive VQA-CP v2 (60.91%) without sacrificing performance on the in-distribution VQA v2 (60.81%).

LRLGC vs. Other Re-Weighting Methods

Model	Adaptive	q	v	FBS	GC	VQA-CP v2 test (%)
Loss-Rescaling [Guo et al., 2022]		✓				53.26
LPF [Liang et al., 2021]	✓	✓				55.34
LP-Focal [Lao et al., 2021]	✓	✓				58.45
LRLGC (Ours)	✓	✓	✓	✓	✓	60.91

LRLGC: Loss Rebalancing Label and Global Context



Overview of LRLGC training strategy：

- (A) An arbitrary VQA model.
- (B) A Bias Model captures language biases, and the Loss Rebalancing Label Constructor dynamically generates Loss Rebalancing Labels (LRL) for each biased sample.
- (C) A gated multi-headed self-attention mechanism captures global context.
- (D) Learning by the ensemble.

Experimental results

- Comparison results for the VQA-CP v2 test split and the VQA v2 validation split.

Case	Model	VQA-CP v2 test Overall	VQA-CP v2 test Yes/No	VQA-CP v2 test Number	VQA-CP v2 test Other	VQA v2 val Overall	VQA v2 val Yes/No	VQA v2 val Number	VQA v2 val Other	Comparison Gap↓	Comparison Mean
I	SAN [Yang et al., 2016]	24.96	38.35	11.10	21.74	52.41	70.06	39.28	47.84	27.45	38.69
	BAN [Kim et al., 2018]	37.03	41.55	12.43	41.40	63.90	81.42	**45.18**	55.54	26.87	50.47
	UpDn [Anderson et al., 2018]	39.74	42.27	11.93	46.05	63.48	81.18	42.14	**55.66**	23.74	51.61
II	AttAlign [Selvaraju et al., 2019]	39.37	43.02	11.89	45.00	63.24	80.99	42.55	55.22	23.87	51.31
	HINT [Selvaraju et al., 2019]	46.73	67.27	10.61	45.88	63.38	81.18	42.99	55.56	16.65	55.06
	SCR [Wu and Mooney, 2019]	49.45	72.36	10.93	48.02	62.20	78.80	41.60	54.50	12.75	55.83
III	Unshuffling [Teney et al., 2021]	42.39	47.72	14.43	47.24	61.08	78.32	42.16	52.71	18.69	51.74
	RandImg [Teney et al., 2020b]	55.37	83.89	41.60	44.20	57.24	76.53	33.87	48.57	1.87	56.31
	CSS [Chen et al., 2020]	58.95	84.37	49.42	48.21	59.91	73.25	39.77	55.11	0.96	59.43
	CL-VQA [Liang et al., 2020]	59.18	86.99	49.89	47.16	57.29	67.27	38.40	54.71	1.89	58.24
	SSL-VQA [Zhu et al., 2020]	57.59	86.53	29.87	**50.03**	**63.73**	-	-	-	6.14	60.66
IV	AdvReg [Ramakrishnan et al., 2018]	41.17	65.49	15.48	35.48	62.75	79.84	42.35	55.16	21.58	51.96
	RUBi [Cadene et al., 2019]	45.42	63.03	11.91	44.33	58.19	63.04	41.00	54.43	12.77	51.81
	LMH [Clark et al., 2019]	52.01	72.58	31.12	46.97	56.35	65.06	37.63	54.69	4.34	54.18
	CF-VQA [Niu et al., 2021]	53.55	**91.15**	13.03	44.97	63.54	**82.51**	43.96	54.30	9.99	58.55
	GGE-DQ [Han et al., 2021]	57.32	87.04	27.75	49.59	59.11	73.27	39.99	54.39	1.79	58.22
	LPF [Liang et al., 2021]	55.34	88.61	23.78	46.57	55.01	64.87	37.45	52.08	0.33	55.18
	Loss-Rescaling [Guo et al., 2022]	53.26	72.82	48.00	44.46	56.81	68.21	36.37	52.29	3.55	55.04
	LP-Focal [Lao et al., 2021]	58.45	88.34	34.67	49.32	62.45	-	-	-	4.00	60.45
	CCB-VQA [Yang et al., 2021]	59.12	89.12	51.04	45.62	59.17	77.28	33.71	52.14	**0.05**	59.15
	SBS [Ouyang et al., 2022]	59.57	87.44	**52.96**	46.79	61.97	78.80	42.17	54.41	2.40	60.77
	LRLGC (Ours)	**60.91**	89.95	45.13	**50.03**	60.81	77.65	39.25	53.71	0.10	**60.86**

- The effect of different backbones on model performance on the VQA-CP v2 test set.

Model	Yes/No	Number	Other	Overall	Gap↑
SAN† [Yang et al., 2016]	40.86	13.43	46.98	40.08	
SAN+LRLGC	88.03	42.05	47.65	58.56	**+18.48**
BAN† [Kim et al., 2018]	43.53	13.60	46.35	40.53	
BAN+LRLGC	89.85	42.74	47.64	59.19	**+18.66**
UpDn† [Anderson et al., 2018]	43.32	13.41	48.32	41.54	
UpDn+LRLGC	89.95	45.13	50.03	60.91	**+19.37**

- LRLGC results on VQA-CP v2 test set with varying training split proportions.

Model	Proportion of Training Set 20%	40%	60%	80%	100%
SAN† [Yang et al., 2016]	33.15	36.62	39.11	39.71	40.08
SAN+LRLGC	43.80	53.19	56.67	57.13	**58.56**
BAN† [Kim et al., 2018]	33.05	37.28	38.52	40.00	40.53
BAN+LRLGC	42.66	54.16	56.91	58.65	**59.19**
UpDn† [Anderson et al., 2018]	36.37	38.72	39.91	40.53	41.54
UpDn+LRLGC	54.10	57.57	59.02	59.96	**60.91**

- Each LRLGC module's effect on the model performance.

	LRL	FBS	GC	VQA-CP v2 test (%)
1				41.54
2	q			57.83
3	qv			58.90
4	q	✓		58.17
5	qv	✓		58.77
6	q		✓	59.43
7	qv		✓	59.81
8	q	✓	✓	59.84
9	qv	✓	✓	60.91

- Results for various α and β combinations.

Model	α vs. β	VQA-CP v2 test (%)
LRLGC	0.1 : 4	59.58
	0.3 : 4	60.08
	0.5 : 4	**60.91**
	0.7 : 4	60.28
	0.5 : 3	60.65
	0.5 : 5	60.08

- Quantitative analysis