LPF: A Language-Prior Feedback Objective Function for De-biased Visual Question Answering

SIGIR 2021

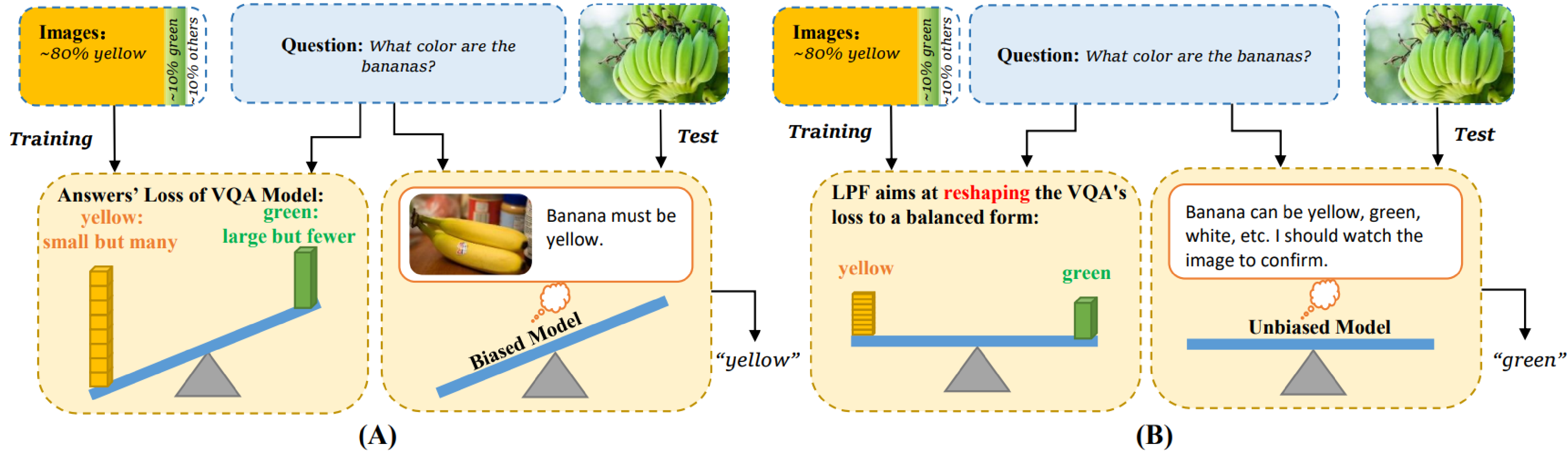
Zujie Liang, Haifeng Hu, Jiaying Zhu

Sun Yat-Sen University, China

Background – Bias in VQA

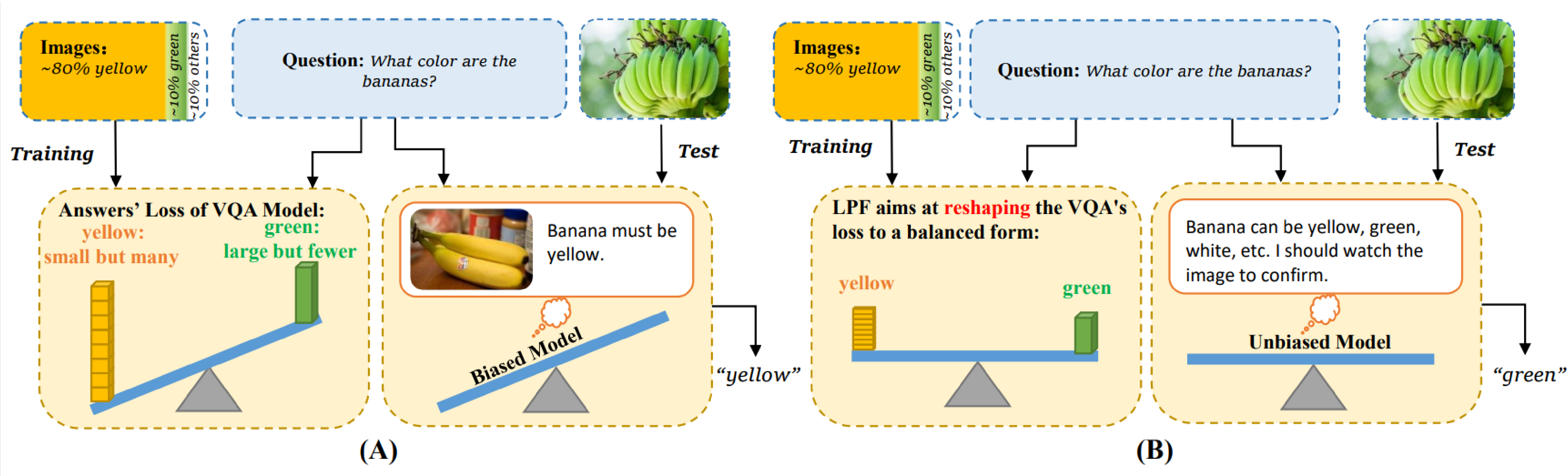
sigir21

- Strong superficial linguistic correlations in the training set



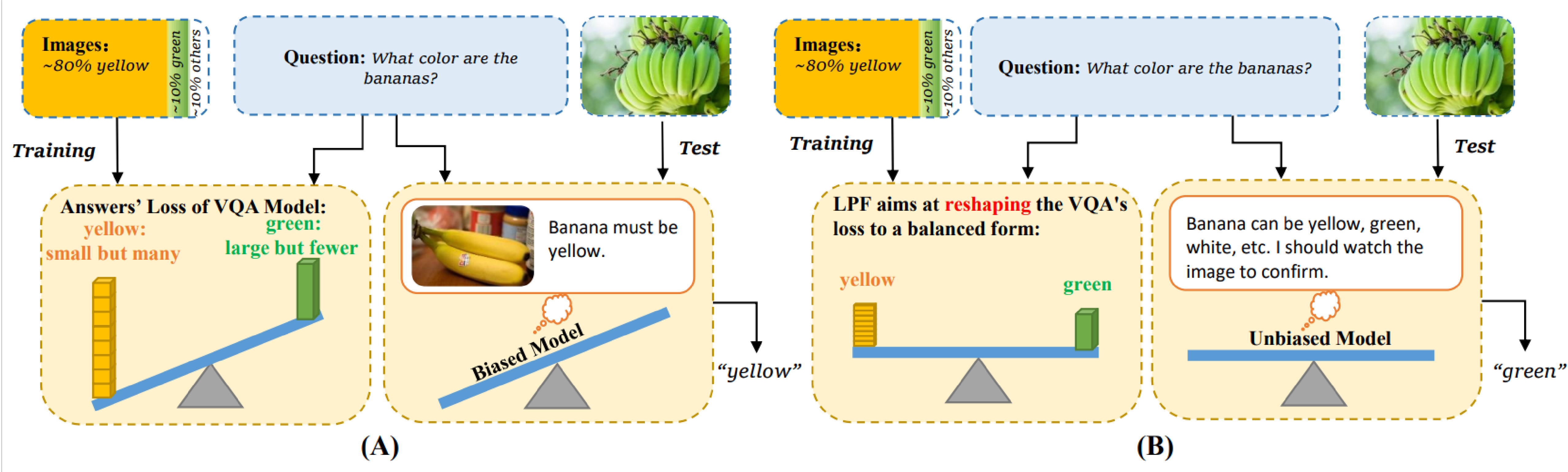
Background – Bias in VQA

- Strong superficial linguistic correlations in the training set
  - VQA models tend to guess the answer based on the language prior



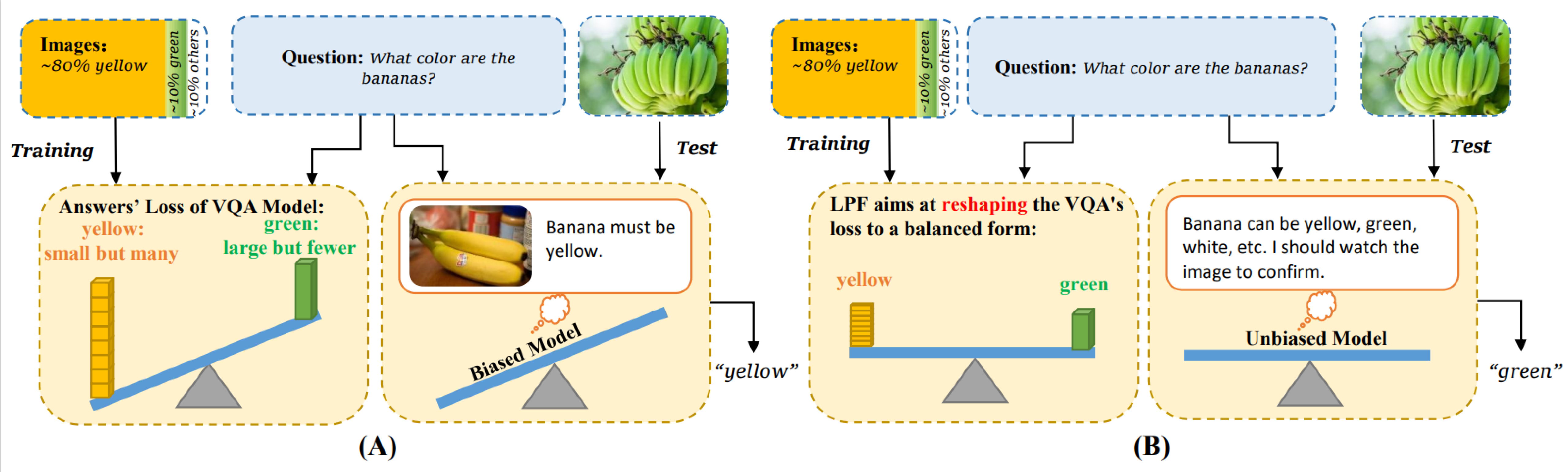
Background – Bias in VQA

- Strong superficial linguistic correlations in the training set
  - VQA models tend to guess the answer based on the language prior
  - Poor robustness and generalization



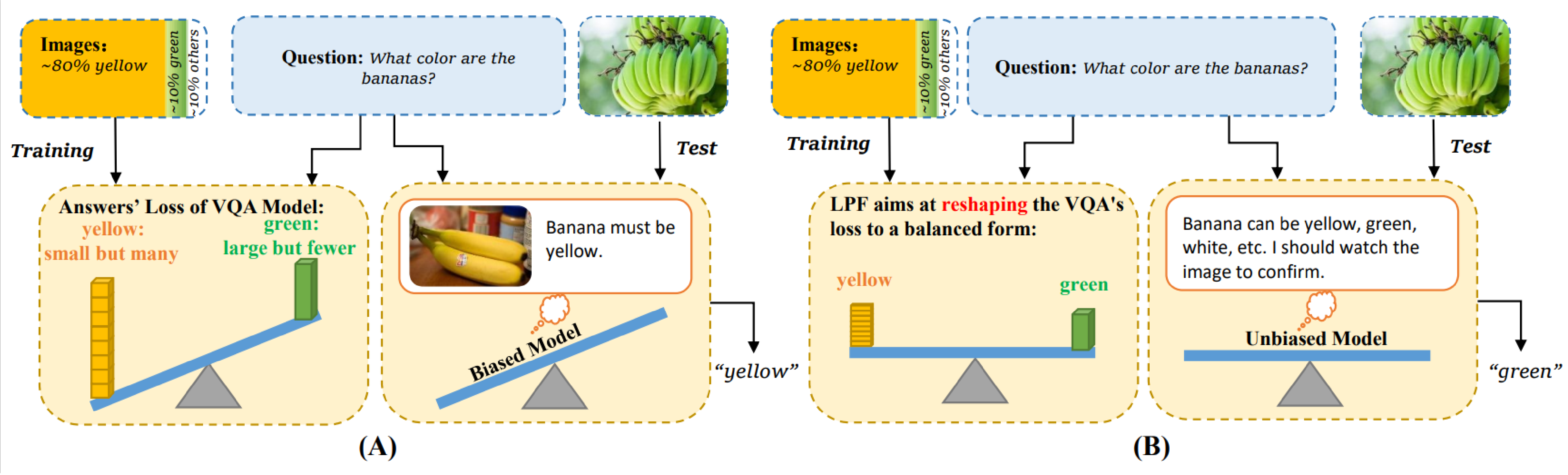
Motivation

- Long-tailed answer distribution could lead to unbalanced training objective



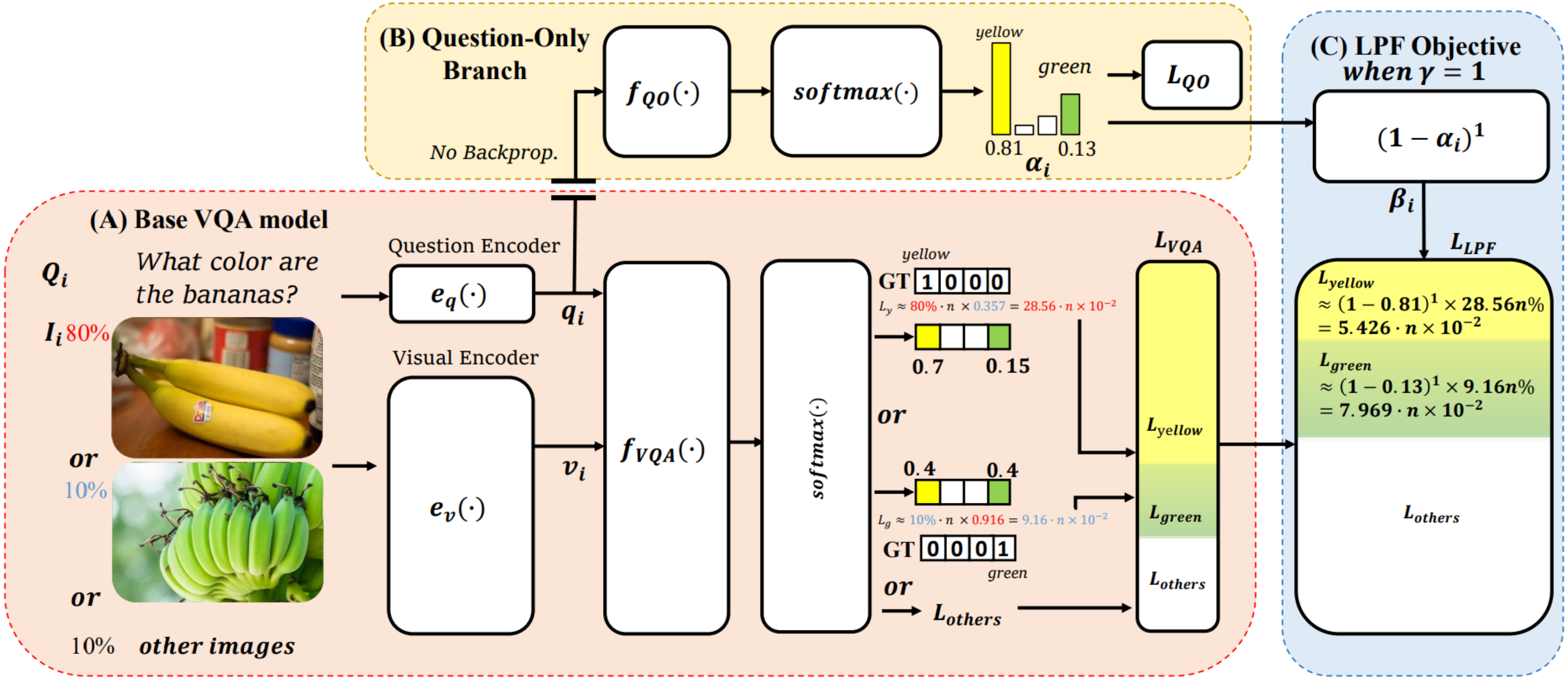
Motivation

- Long-tailed answer distribution could lead to unbalanced training objective
- LPF aims at automatically reshaping the training loss to a balanced form



Model Overview

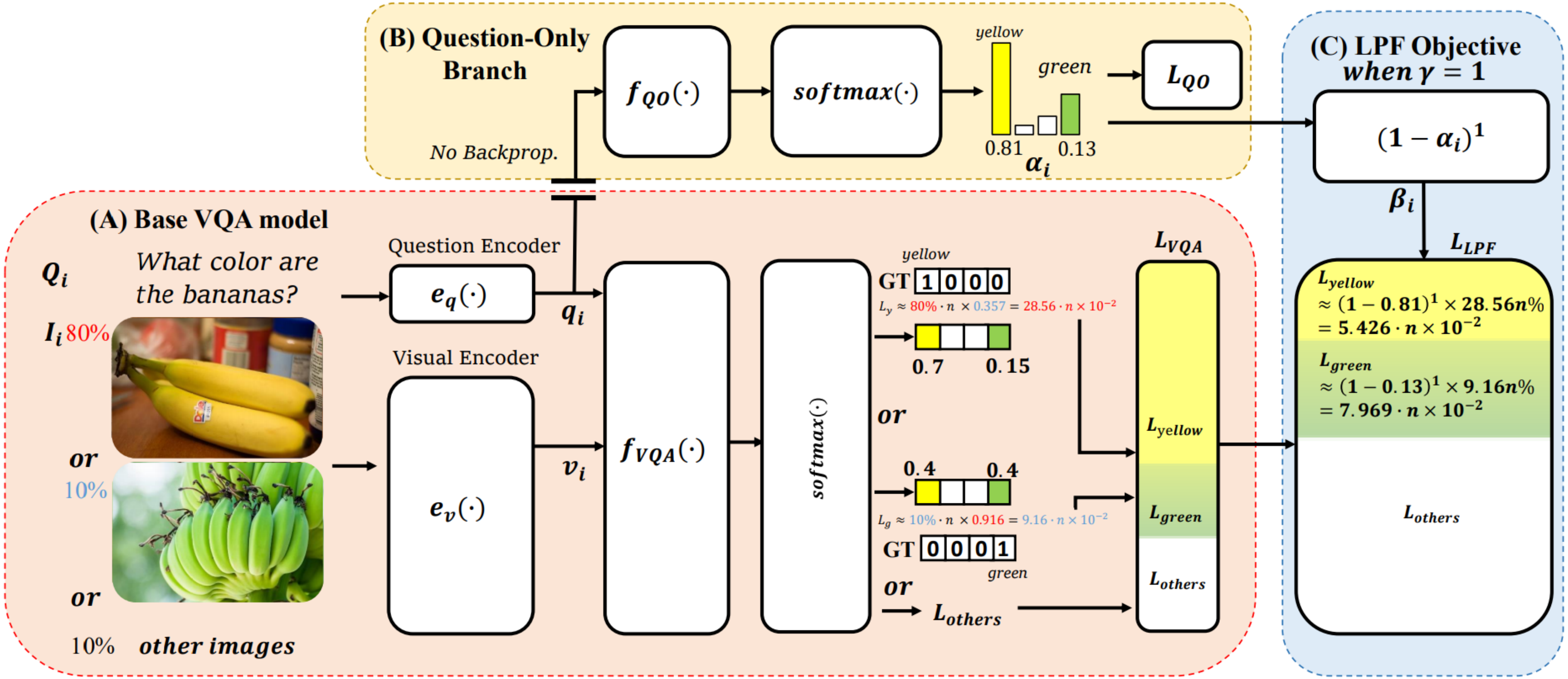
- (A) VQA model



Model Overview

sigir21

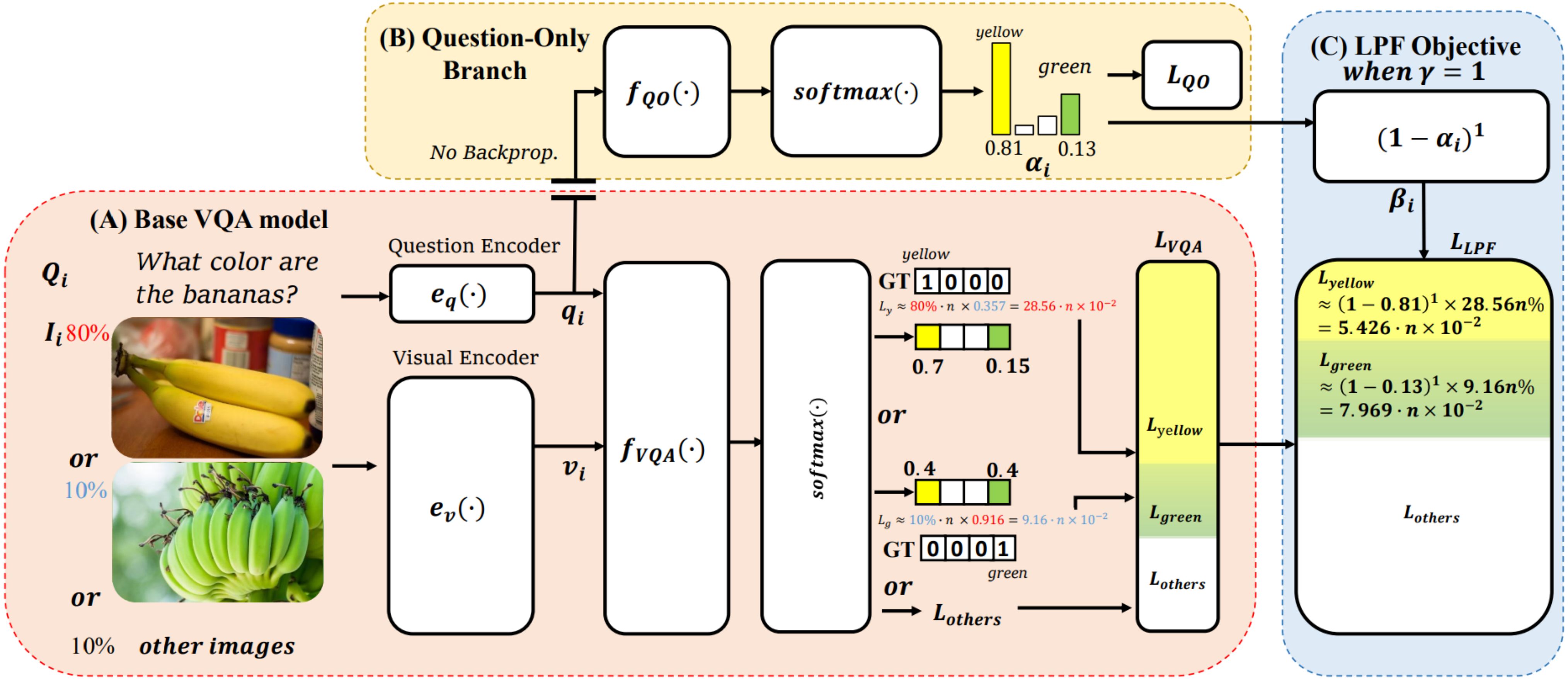
- (A) VQA model
- (B) Question-Only Branch: modeling the bias



Model Overview

sigir21

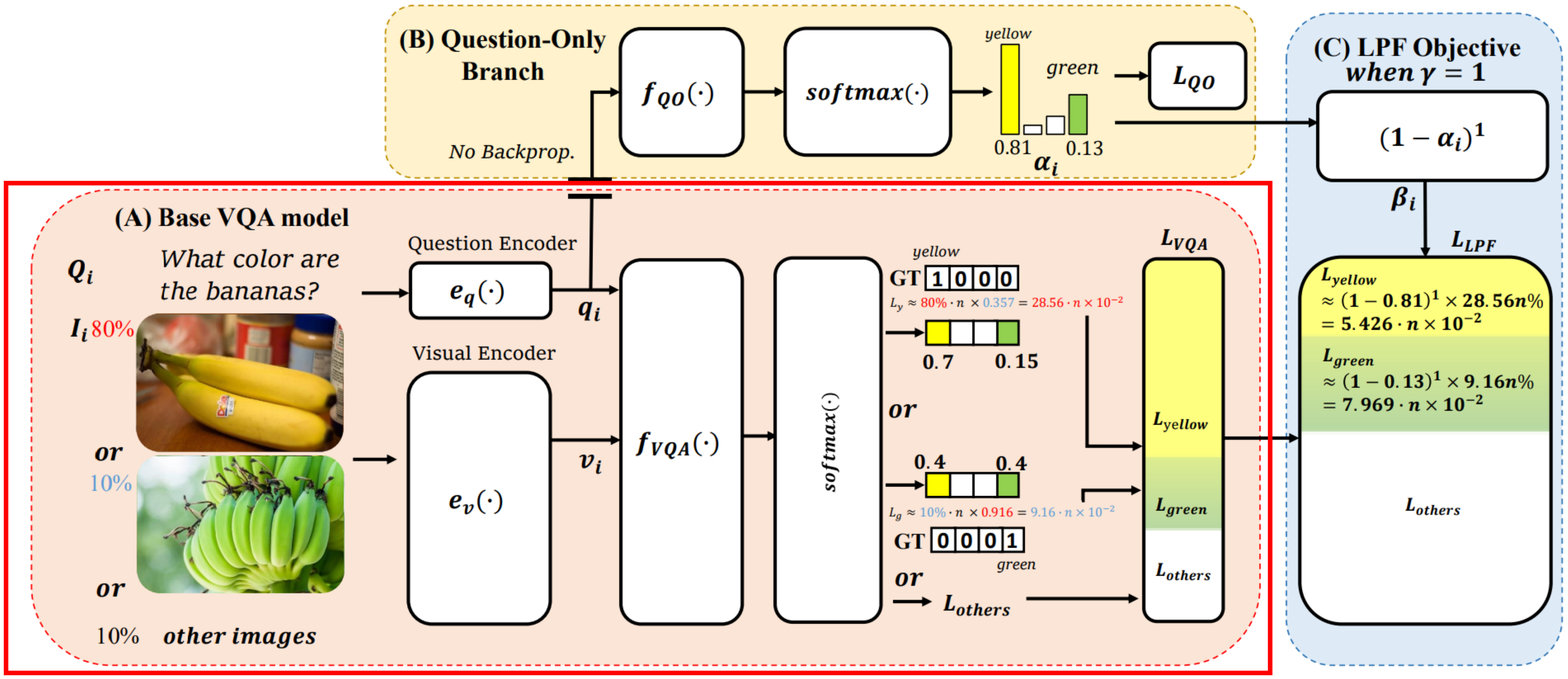
- (A) VQA model
- (B) Question-Only Branch: modeling the bias
- (C) LPF objective: removing the bias through re-weighting



Model Overview

- (A) VQA model

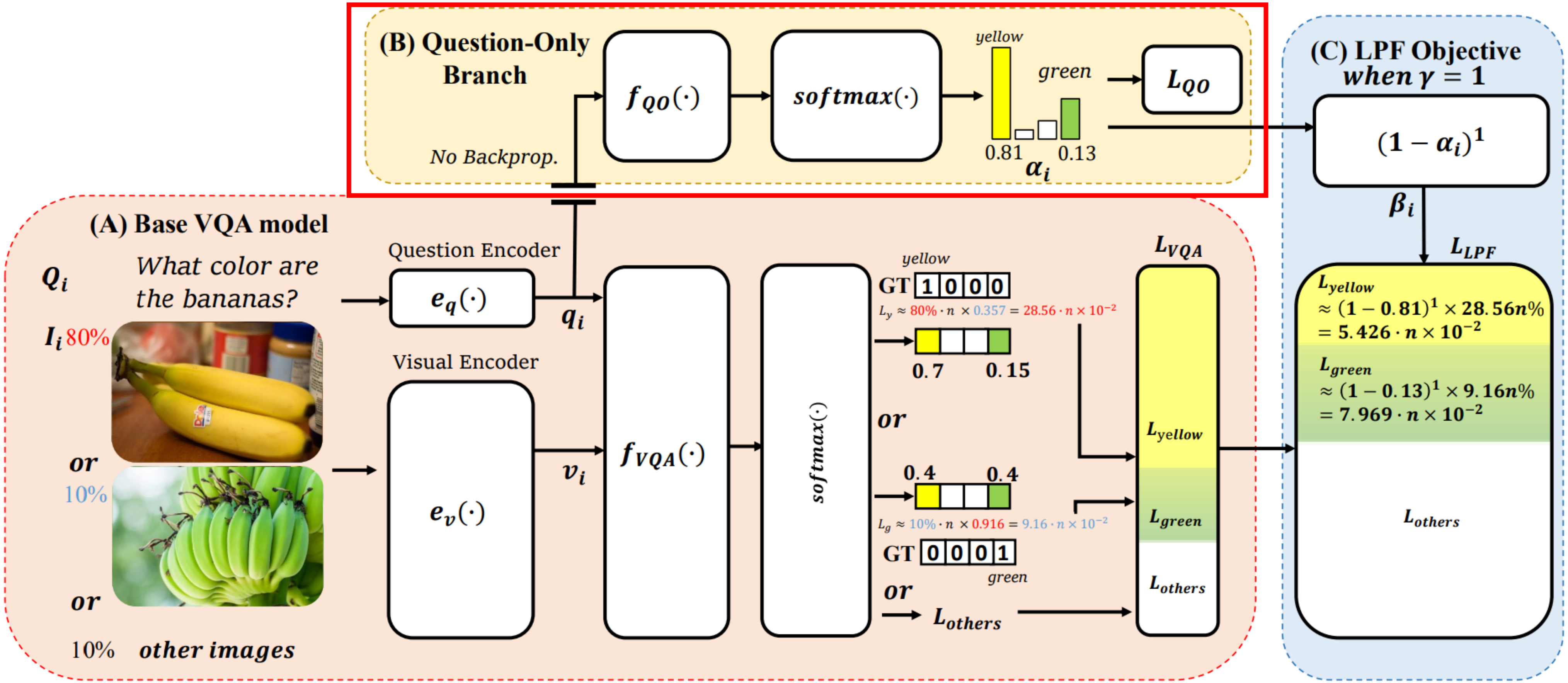
$$\mathcal{L}_{VQA} = -\frac{1}{N}\sum_{i=1}^{N} \log\left(softmax\left(f_{VQA}\left(\mathbf{v_i}, \mathbf{q}_i\right)\right)\right)[a_i]$$



Model Overview

- (B) Question-Only Branch: modeling the bias

$$\mathcal{L}_{QO} = -\frac{1}{N}\sum_{i=1}^{N} \log\left(softmax\left(f_{QO}\left(\mathbf{q}_i\right)\right)\right)[a_i]$$



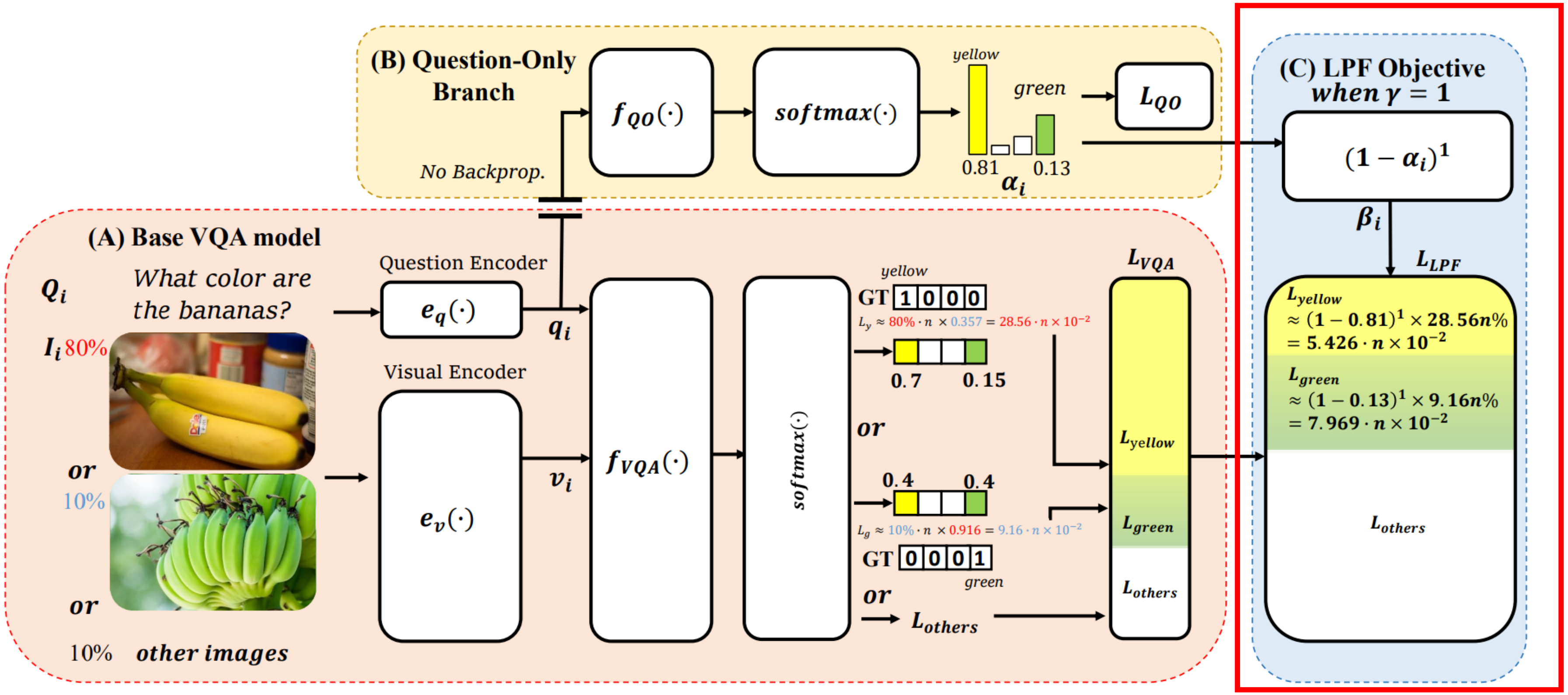
Model Overview

- (C) LPF objective: removing the bias through re-weighting

$$\alpha_i = softmax\left(f_{QO}\left(\mathbf{q}_i\right)\right)[a_i] = \frac{\exp\left(f_{QO}\left(\mathbf{q}_i\right)\right)[a_i]}{\sum_{j=1}^{|\mathcal{A}|}\exp\left(f_{QO}\left(\mathbf{q}_i\right)\right)[a_j]}$$

$$\mathcal{L}_{LPF} = -\frac{1}{N}\sum_{i=1}^{N}\underbrace{\left(1-\alpha_i\right)^{\gamma}}_{=\beta_i}\log\left(softmax\left(f_{VQA}\left(\mathbf{v}_i,\mathbf{q}_i\right)\right)\right)[a_i]$$

$$\mathcal{L}_{total} = \mathcal{L}_{LPF} + \mathcal{L}_{QO}$$



Experiments

- Comparison with the SOTA systems on bias-sensitive VQA-CP v2 *test* and VQA v2 *val* set
  - LPF shows significant improvements over UpDn baseline on VQA-CP v2
  - LPF achieves competitive performance

Model		VQA-CP v2 *test*				VQA v2 *val*			
	γ	Overall	Yes/No	Number	Other	Overall	Yes/No	Number	Other
GQA [2]		31.30	57.99	13.68	22.14	48.24	72.03	31.17	34.65
UpDn [3]		39.49	45.21	11.96	42.98	63.48	81.18	42.14	55.66
UpDn+HINT [25]		46.73	67.27	10.61	45.88	63.38	81.18	42.99	55.56
UpDn+SCR [32]		49.17	71.55	10.72	**47.49**	62.20	78.90	41.40	54.30
UpDn+SSL(CE) [33]		52.63	87.75	26.40	41.42	63.73	-	-	-
UpDn+CSS [9]		58.95	84.37	49.42	48.21	59.91	73.25	39.77	55.11
UpDn+AdvReg [24]		41.17	65.49	15.48	35.48	62.75	79.84	42.35	55.16
UpDn+GRL [15]		42.33	59.74	14.78	40.76	51.92	-	-	-
UpDn+RUBi [8]		44.23	67.05	17.48	39.61	61.16	-	-	-
UpDn+VGQE [13]		48.75	-	-	-	64.04	-	-	-
UpDn+DLR [18]		48.87	70.99	18.72	45.57	57.96	76.82	39.33	48.54
UpDn+LMH [10]		52.01	72.58	**31.11**	46.96	56.34	65.05	37.63	54.68
UpDn+LPF(ours)	1	51.57	87.33	12.25	43.61	62.63	79.51	42.90	55.02
UpDn+LPF(ours)	5	**55.34**	**88.61**	23.78	46.57	55.01	64.87	37.45	52.08

Experiments

- LPF is model-agnostic: generalizing well on different VQA architectures

Model	Overall	Y/N	Number	Other	GapΔ ↑
S-MRL [8]	38.46	42.85	12.81	**43.20**	
S-MRL+RUBi [8]	47.11	68.65	20.28	43.18	+8.65
S-MRL+LPF(ours)	**53.38**	**88.06**	**25.00**	42.99	**+14.92**
UpDn [3]	39.49	45.21	11.96	42.98	
UpDn+RUBi [8]	44.23	67.05	17.48	39.61	+4.74
UpDn+LPF(ours)	**55.34**	**88.61**	**23.78**	**46.57**	**+15.85**
BAN [19]	37.03	41.55	12.43	**41.40**	
BAN+LPF(ours)	**50.76**	**88.13**	**18.59**	40.03	**+13.73**

Experiments

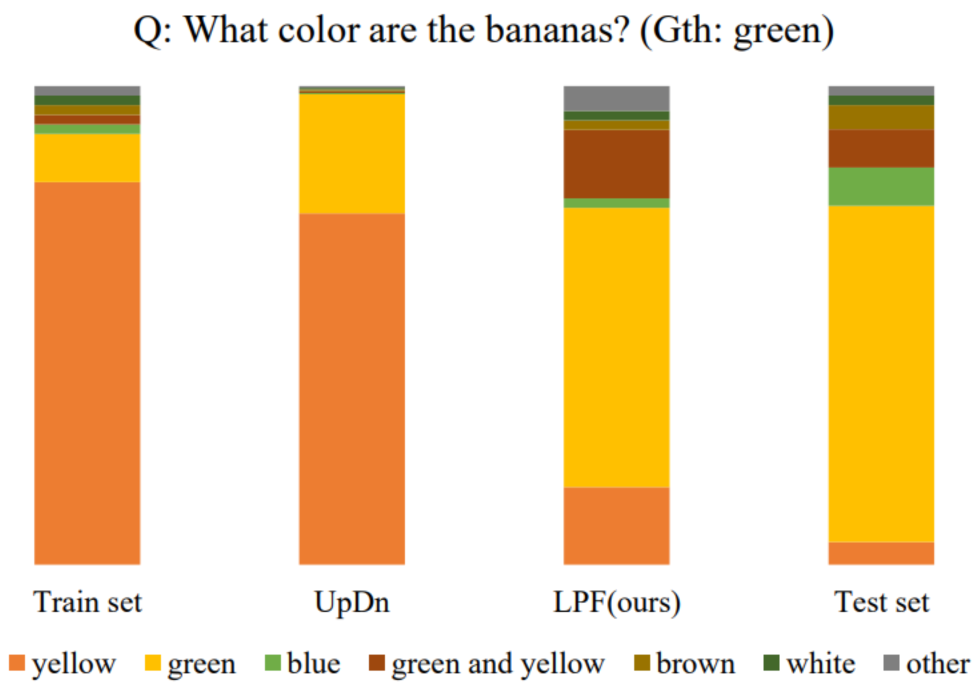
- Discussion on different variants of re-weighting based methods
  - Pre-computing the prior distribution on training set
  - Focal loss

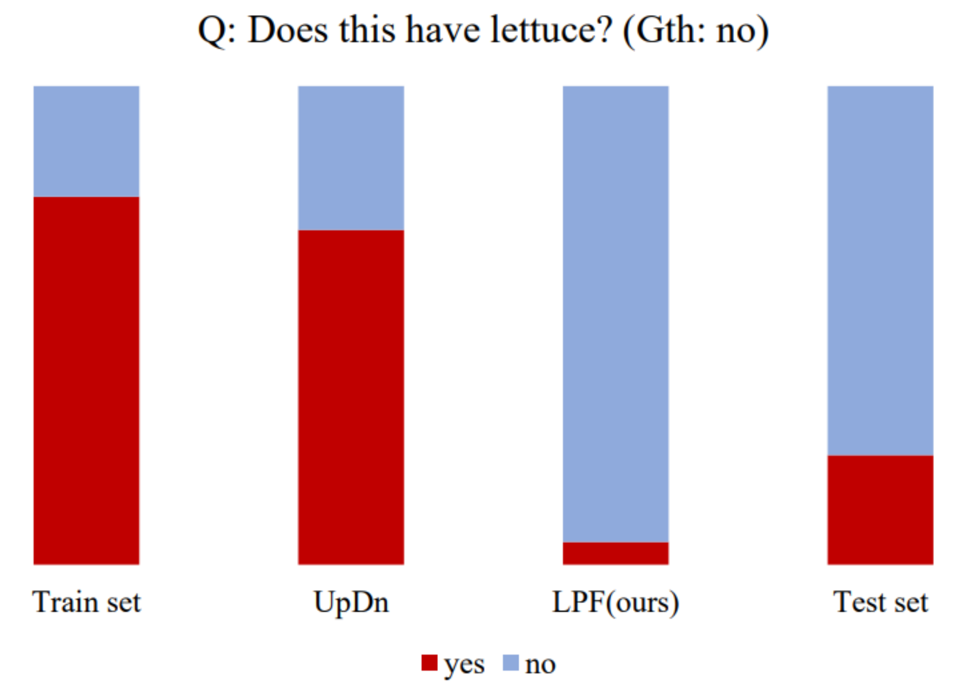
Model	Overall	Yes/No	Number	Other	Gap$\Delta\uparrow$
UpDn [3]	39.49	45.21	11.96	42.98	
UpDn+Precomputing	40.04	44.81	11.73	**45.31**	+0.55
UpDn+Focal	38.52	42.38	**12.38**	43.67	−0.97
UpDn+LPF	**51.57**	**87.33**	12.25	43.61	**+12.08**

Qualitative Analysis

sigir21

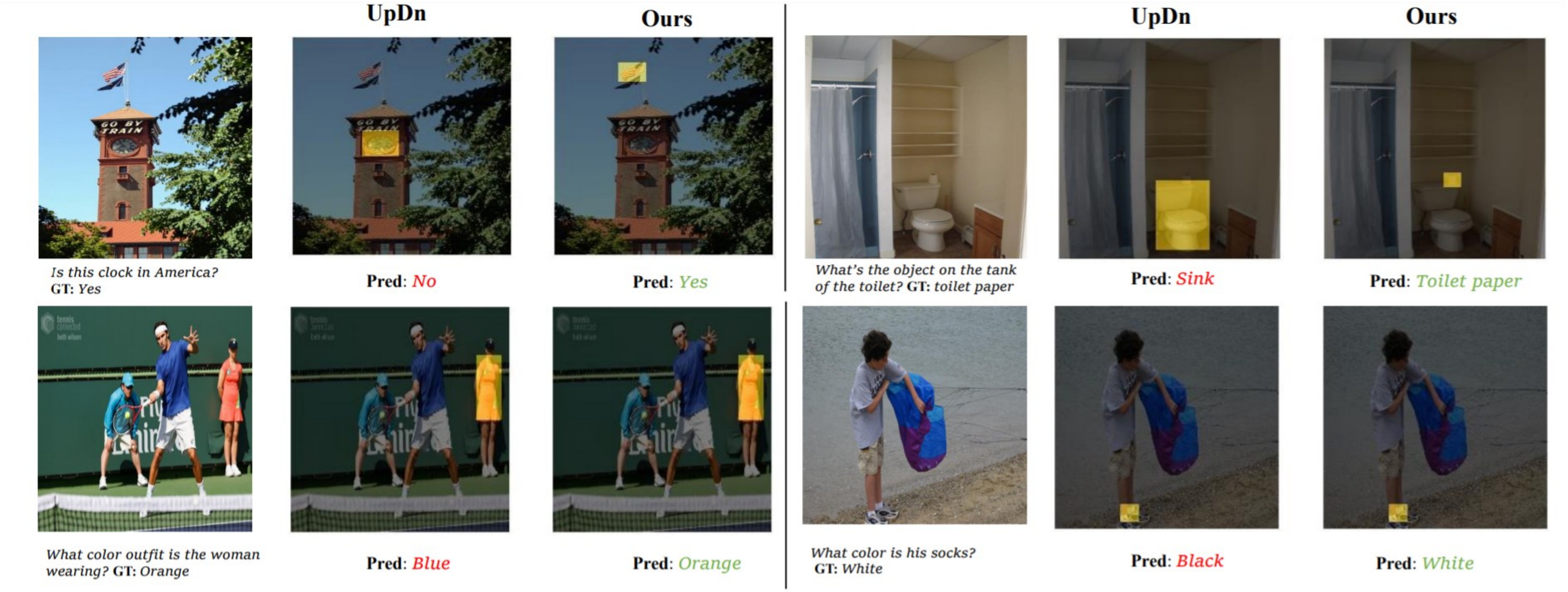
- Baseline UpDn suffers from the language prior on the training set
- LPF helps to overcome language prior and enables the model to be more grounded on the image content





Qualitative Analysis

- LPF helps model to attend to a more reasonable visual region
- LPF enables the model to be more robust instead of biasing to the common answer



Thanks for listening!

- Paper: https://arxiv.org/abs/2105.14300
- Code: https://github.com/jokieleung/LPF-VQA
- Personal homepage: https://jokieleung.github.io/