

SnowMaster: Comprehensive Real-world Image Desnowing via MLLM with Multi-Model Feedback Optimization

Jianyu Lai^{1,*} Sixiang Chen^{1,*} Yunlong Lin² Tian Ye¹ Yun Liu³ Song Fei¹
Zhaohu Xing¹ Hongtao Wu¹ Weiming Wang⁴ Lei Zhu^{1,5,†}

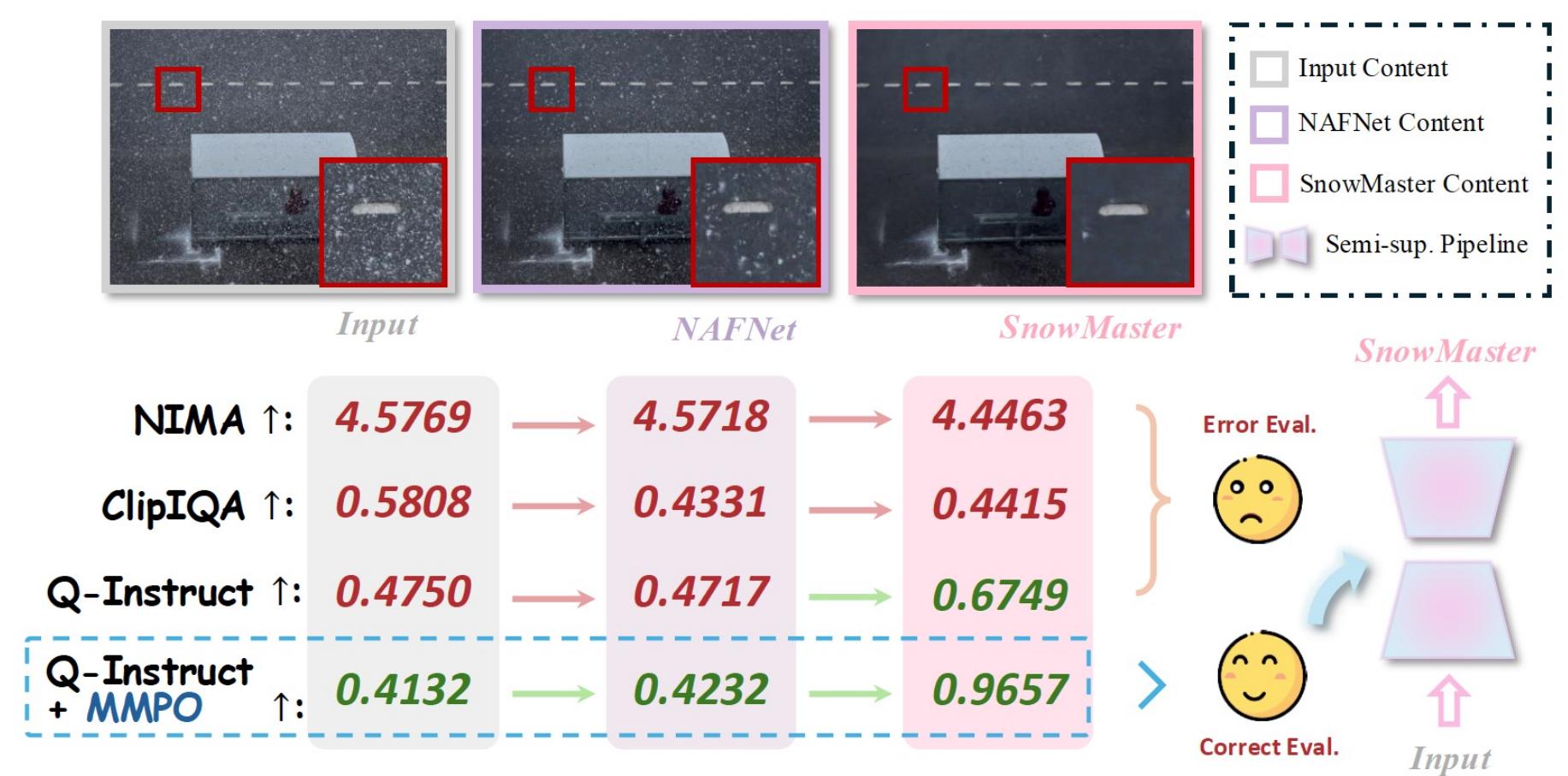
¹ The Hong Kong University of Science and Technology (GZ) ² Xiamen University ³ Southwestern University
⁴ Hong Kong Metropolitan University ⁵ The Hong Kong University of Science and Technology



Problem of Real-world Desnowing

- Existing image desnowing methods are trained on [synthetic data](#). Due to the complexity of [real-world snowfall](#), they struggle with domain gaps and generalize poorly to real-world snow images.
- Introducing real-world snowfall images into training is a viable approach to enhance model generalization capability. However, existing real-world snowfall datasets are [limited in scale](#), and [reliable evaluation methods are lacking](#).

Motivation

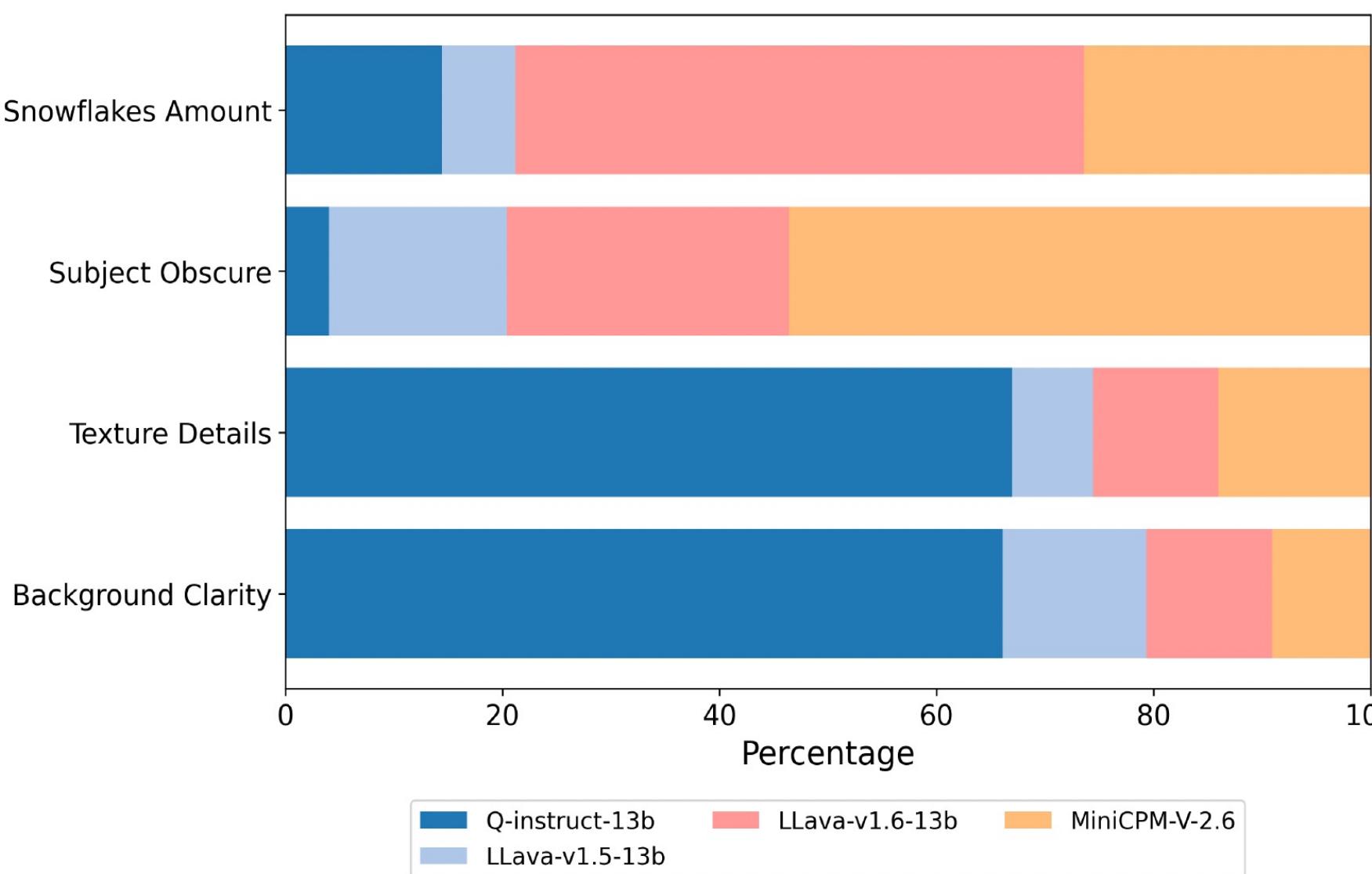


Are existing evaluation metrics for snowfall images reliable?:

- Due to the complexity of real-world snowfall, existing NR-IQA methods are [unsuitable](#) for assessing snowfall images.

Can MLLMs accurately perceive and evaluate snow-covered images from multiple perspectives?

- Unfortunately, our experiment demonstrates that [none of the MLLMs](#) could simultaneously provide accurate descriptions of both low-level visual features and high-level semantic features in snowfall images.



Our Method :

- Multi-Perspectives Evaluation :** Comprehensively evaluate snow removal effects.
- Multi-Model Feedback :** Collect feedback on four models.
- Preference Optimization :** Aggregate the advantages of different models into one model.

Dataset



- The choice are listed below:
 A. ... heavy snow falling ...
 B. ... light snow falling, ...
 C. ... the visibility is poor ...
 D. ... the visibility is good...
 E. ... not a photograph ...

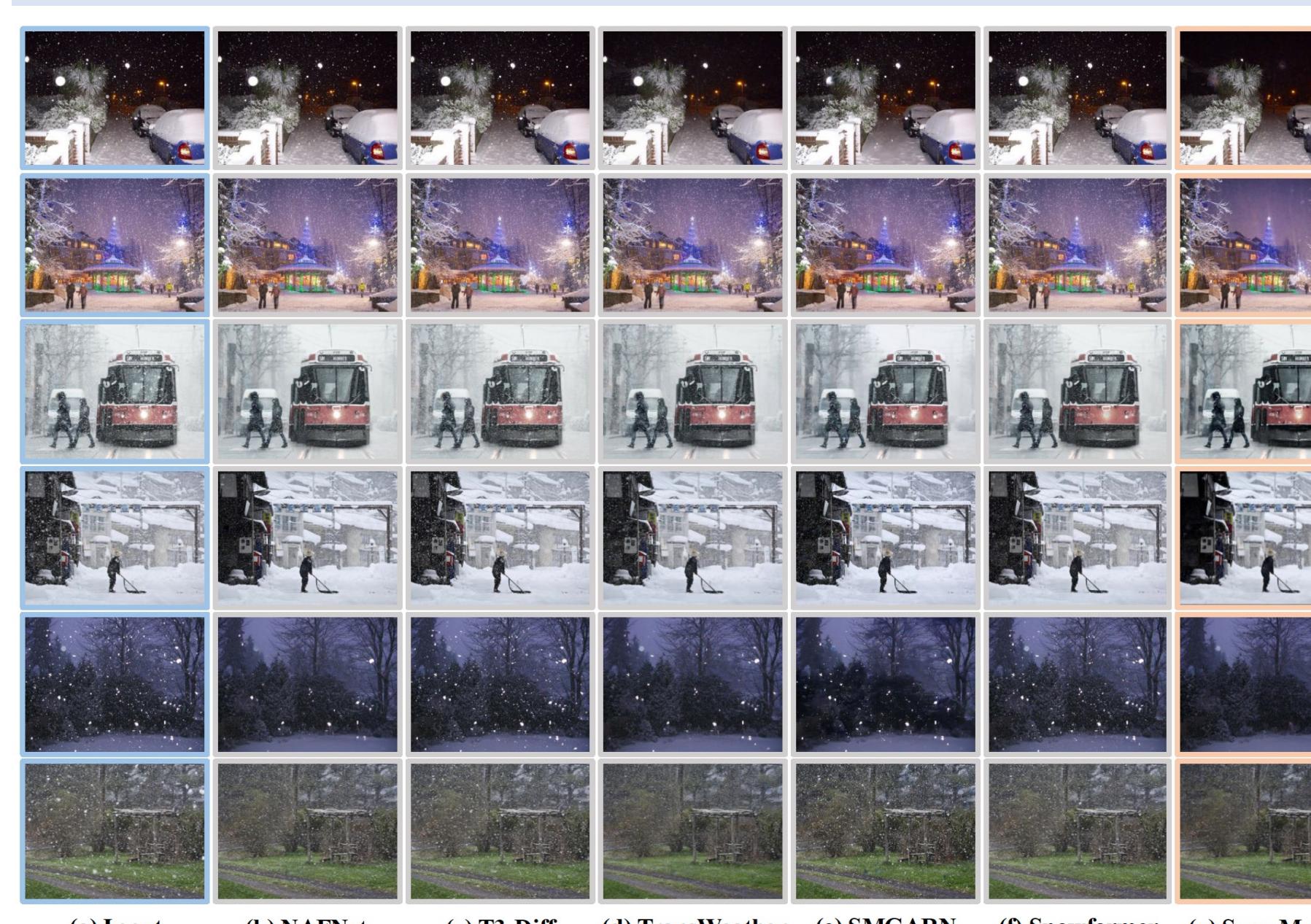
Choose the one closest to the image ...

$$\text{score} = \sum_{i=1}^N \text{prob}_i * w_i$$

$$\text{prob}_i = e^{l_i} / \sum_{j=1}^N e^{l_j}$$

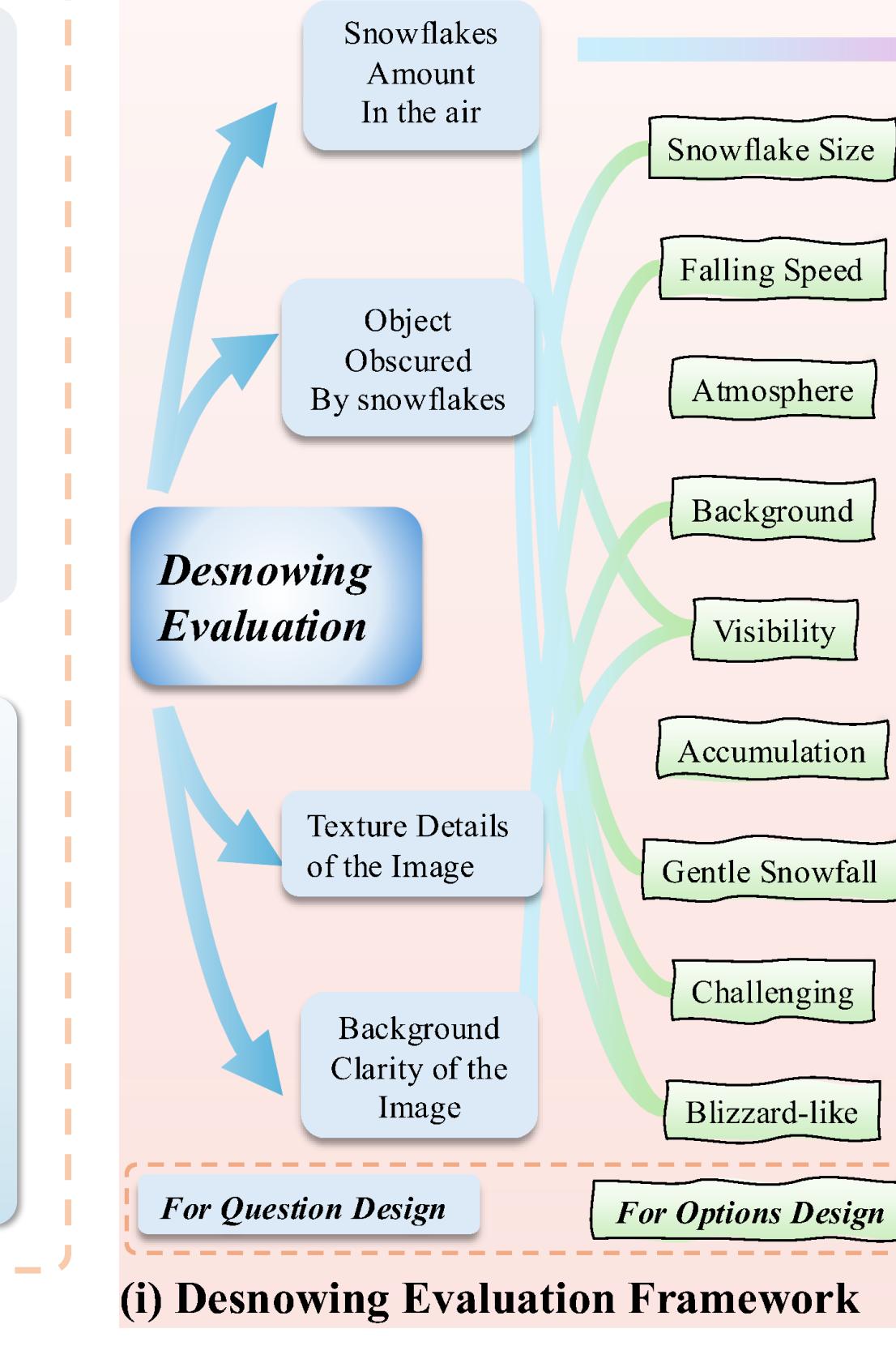
	A	B	C	D	E
w _i	5	4	3	2	1

Dataset: We propose [Realsnow10K](#), the [largest real-world snowfall image dataset](#), containing more than [10K](#) high-quality snowfall images. We utilize [MLLM with Multiple Choice Questions](#) for data cleaning and annotation, where higher score indicate greater snowfall intensity.

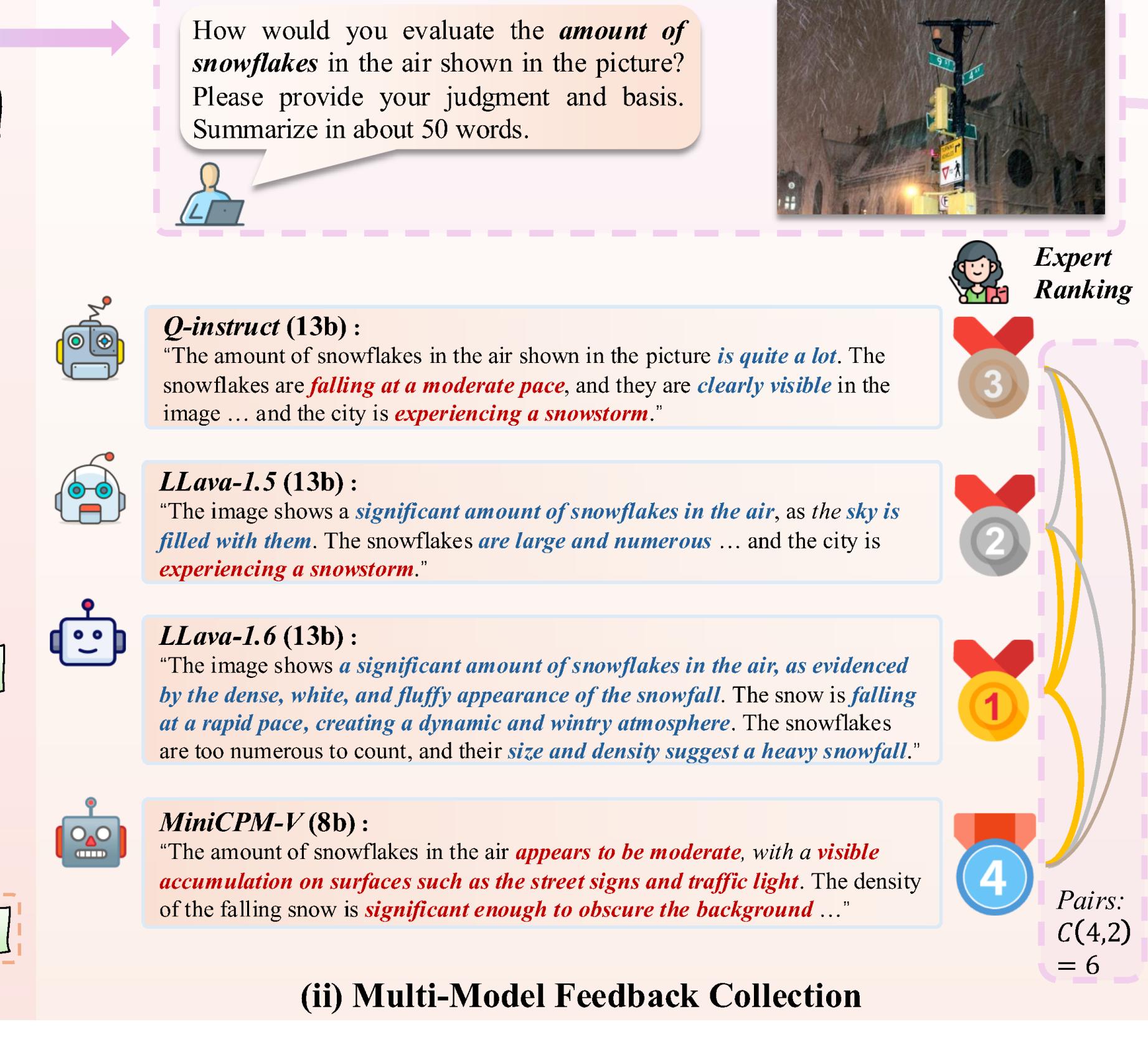


■ Visual Comparison of different models

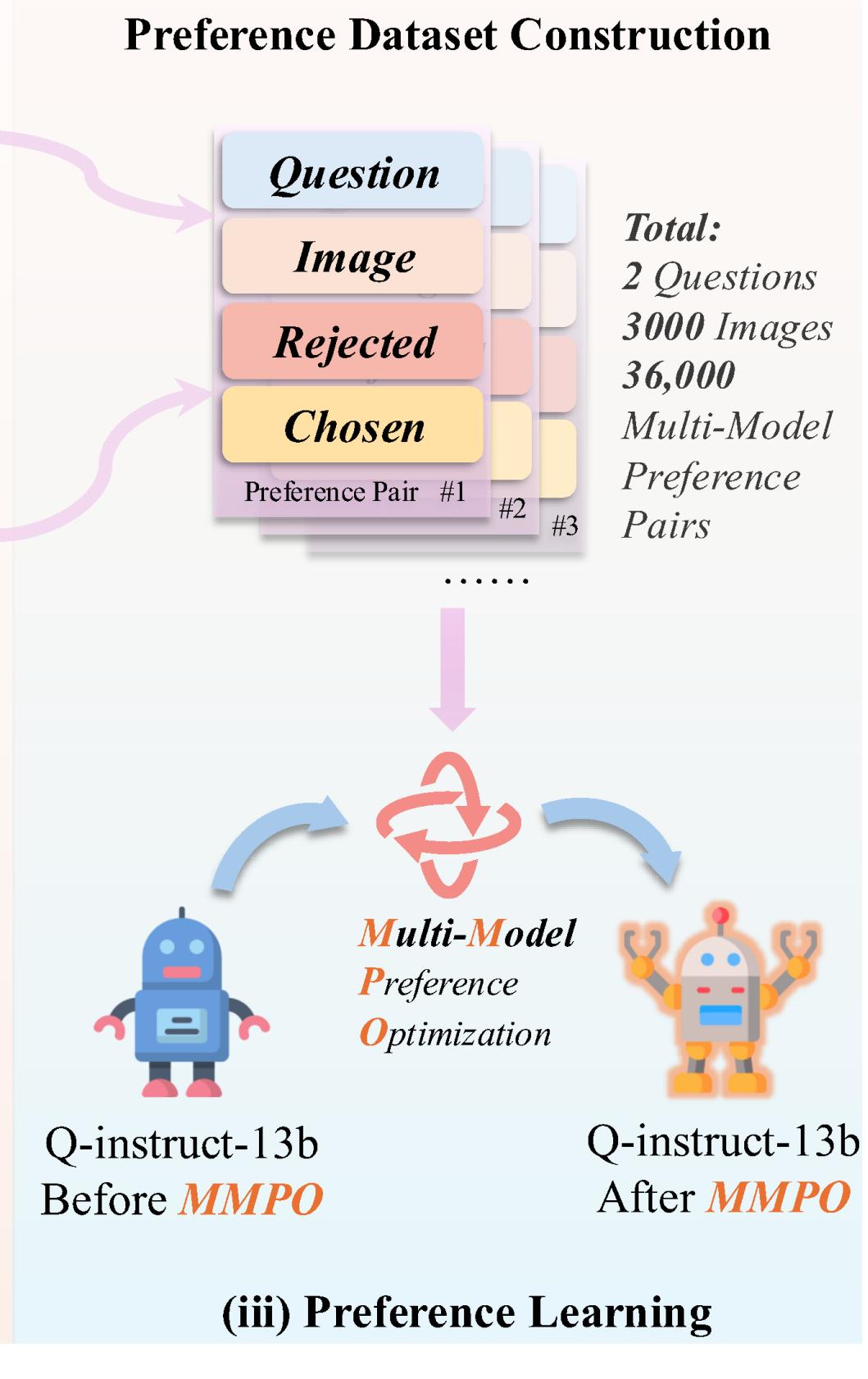
Methodology



(i) Desnowing Evaluation Framework

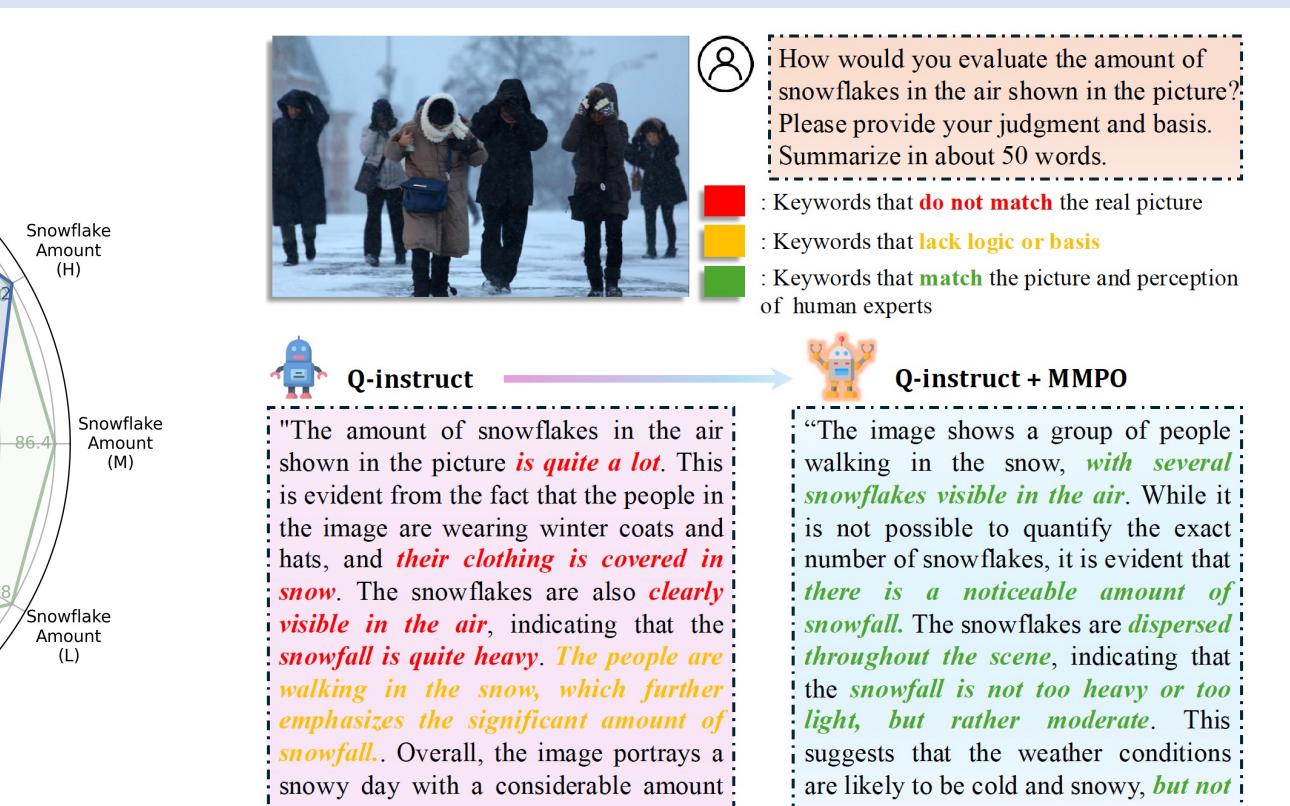


(ii) Multi-Model Feedback Collection



(iii) Preference Learning

Experiments



■ The evaluation accuracy and snowfall perception improvement after MMPO

Method	BRISQUE ↓	NIQE ↓	PIQE ↓	PAQ2-PIQ ↓	BRISQUE ↓	RealSnow10K ↓	NIQE ↓	PIQE ↓	PAQ2-PIQ ↓
SMGARN [ICCV'21]	19.289	3.132	31.927	69.767	23.005	3.580	31.210	68.919	
TransWeather [CVPR'22]	16.544	3.025	30.302	69.821	18.564	3.404	29.772	69.671	
NAFNet [ICCV'22]	17.566	2.903	30.843	69.951	20.425	3.291	29.877	69.388	
SnowFormer [MMF'23]	19.389	2.954	31.902	70.015	17.744	3.429	31.99	69.929	
T ³ -DiffWeather [CVPR'24]	18.98	3.19	31.689	70.042	20.211	3.137	32.991	70.211	
SnowMaster (Ours)	17.55	3.036			16.305	2.797	30.162	70.042	17.546

■ Quantitative Comparison and Ablation study



■ More Visual Comparison. The right side shows the image after desnowing by SnowMaster.