ICCV OCT 19-23, 2025 HONOLULU HAWAII

Fuzzy Contrastive Decoding to Alleviate Object Hallucination in Large Vision-Language Model

Jieun Kim, Jimyeong Kim, Yoonji Kim, Sung-Bae Cho

Yonsei University

Contact Paper

SOFT COMPUTING LABORATORY

YONSEI UNIVERSITY

What’s object Hallucination?

- Model describes non-existent objects in an image
- Model misses objects that actually exist
- Results in false or misleading visual descriptions

What is a vicuna standing in the sand looking at?

A vicuna standing in the sand looking at a tree branch with green leaves.

What's object Hallucination?

- Model describes non-existent objects in an image
- Model misses objects that actually exist
- Results in false or misleading visual descriptions

What is a vicuna standing in the sand looking at?

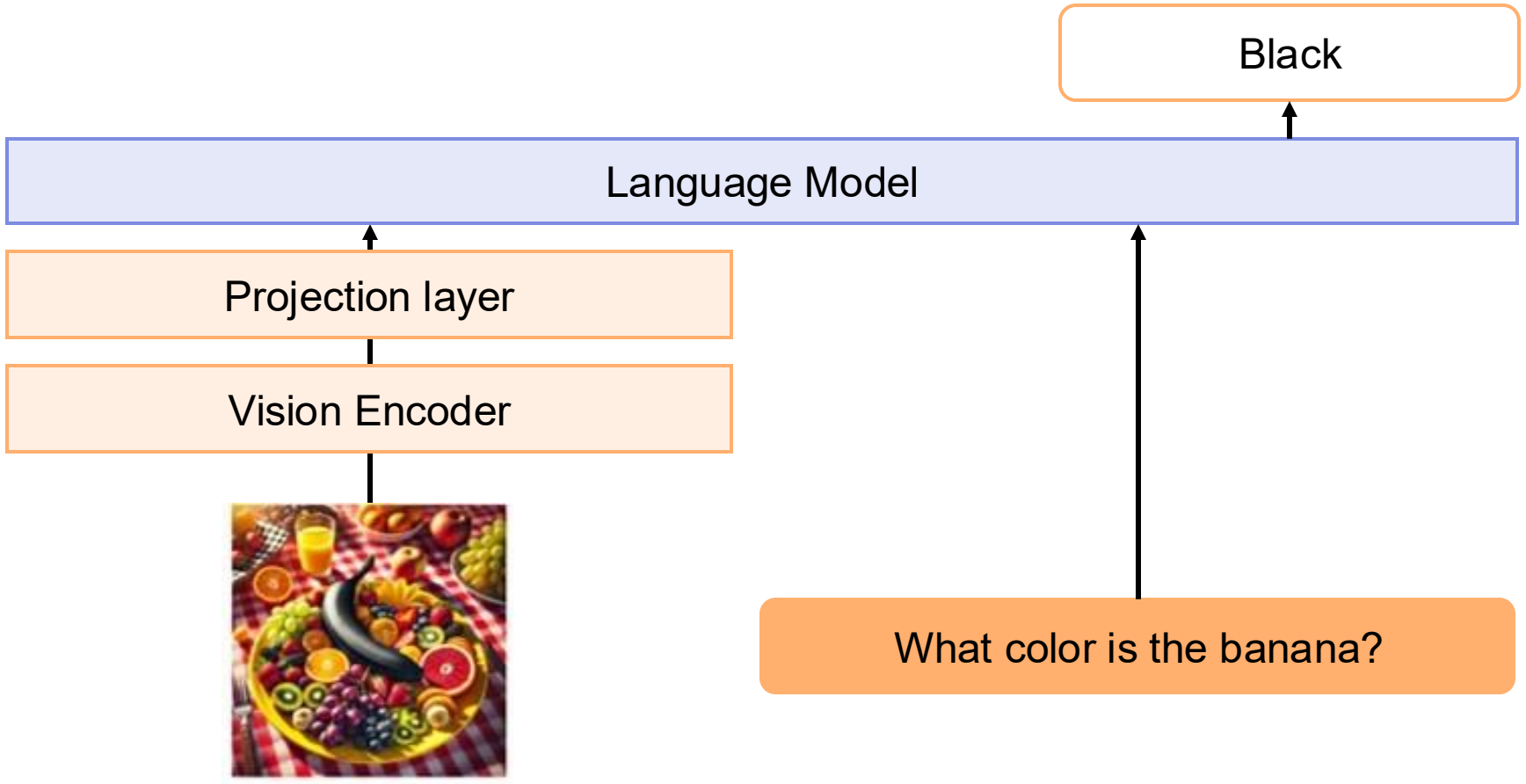
A vicuna standing in the sand looking at a tree branch with green leaves.

There is NO vicuna in the image .. !!

vicuna

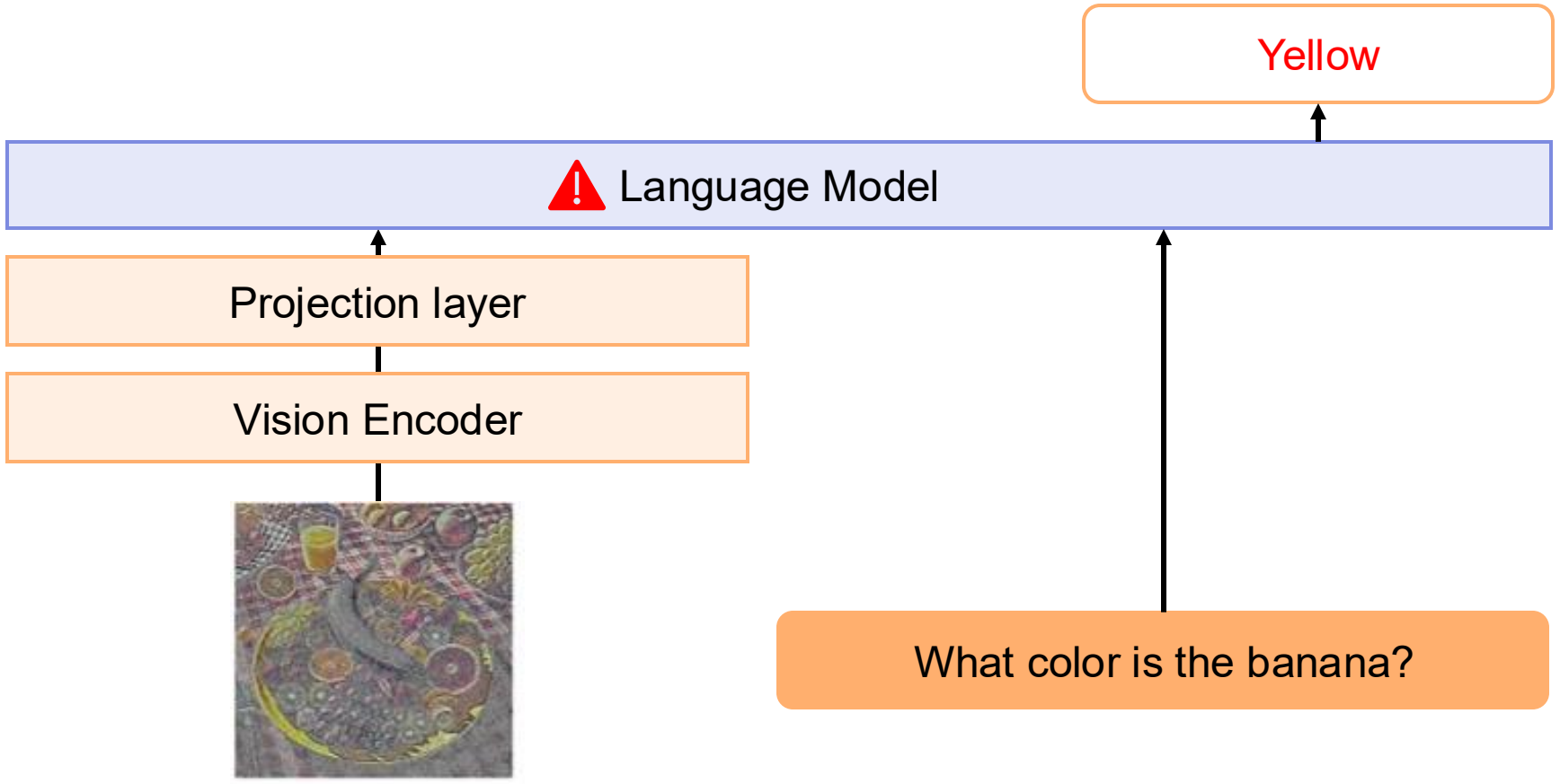
Why this happen?

- Clear visual features allow the model to focus on image evidence



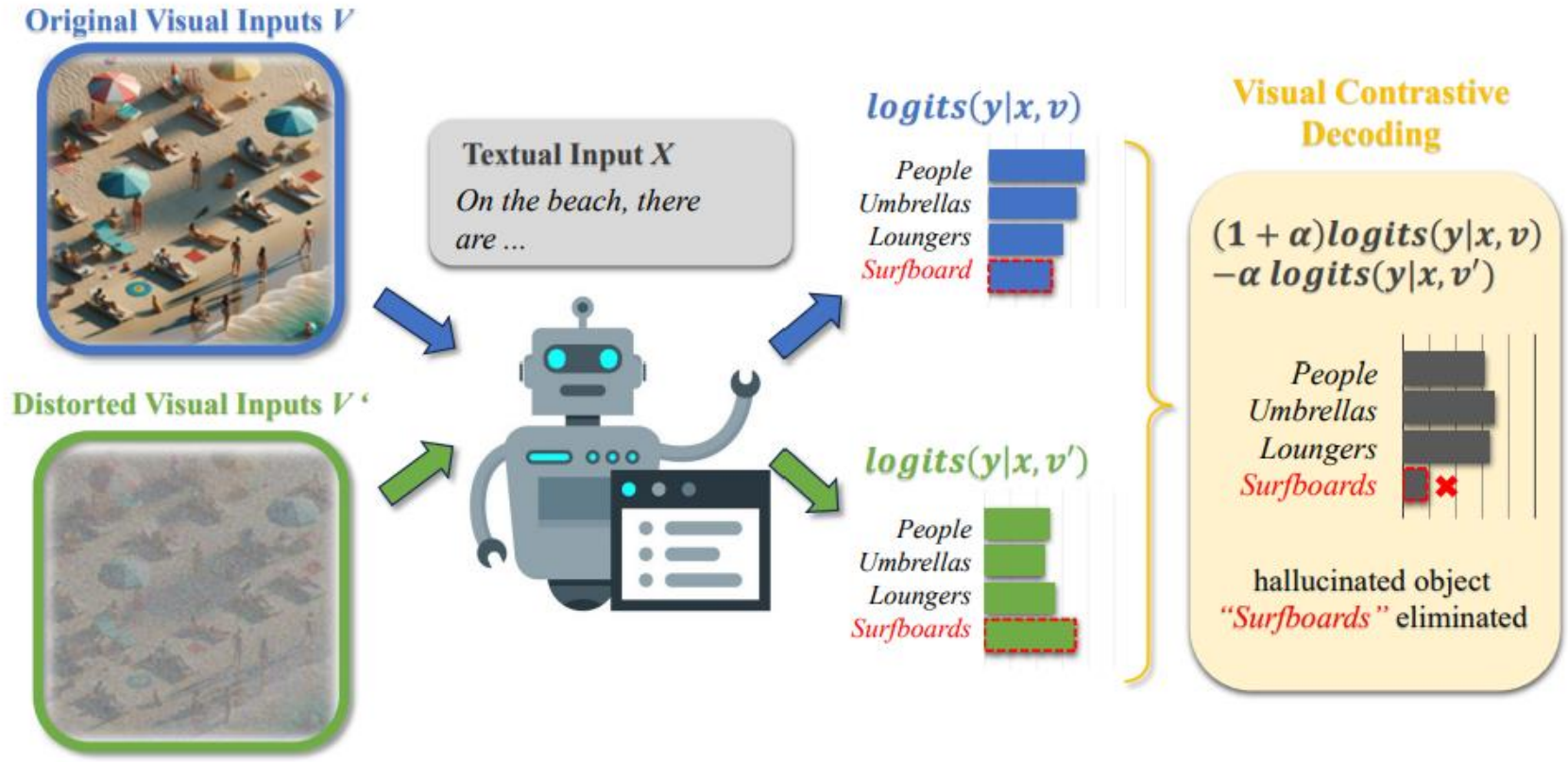
Why this happen?

- The language prior becomes dominant over vision input



Conventional Approaches

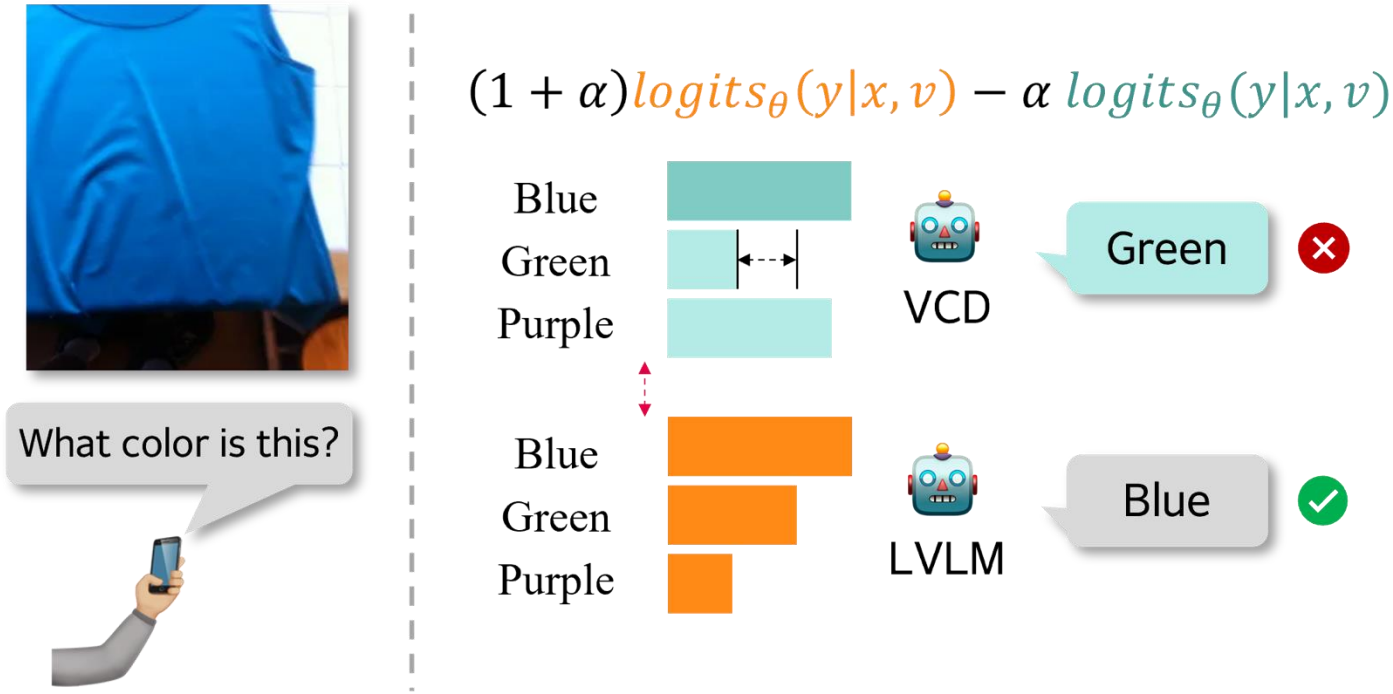
- Visual Constructive Decoding [1]



[1] Leng, Sicong, Hang Zhang, Guanzheng Chen, Xin Li, Shijian Lu, Chunyan Miao, and Lidong Bing. "Mitigating object hallucinations in large vision-language models through visual contrastive decoding." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13872-13882. 2024.

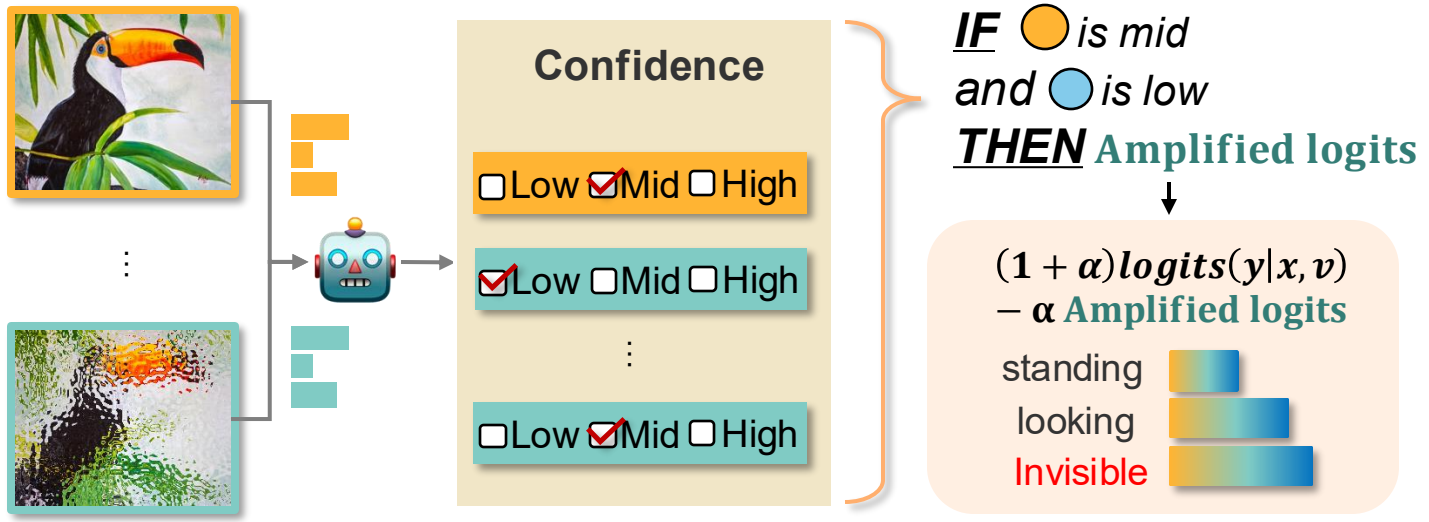
Flaws of Conventional Approaches

- Adjust LLM prior by subtracting hallucination logits
- If the prior is not properly reflected, visual cues are underutilized
- Lead to wrong answers or new hallucinations



Problem Definition

- Generate hallucination-amplified logits to represent the LLM prior
- Use them to contrast and reduce hallucinated outputs

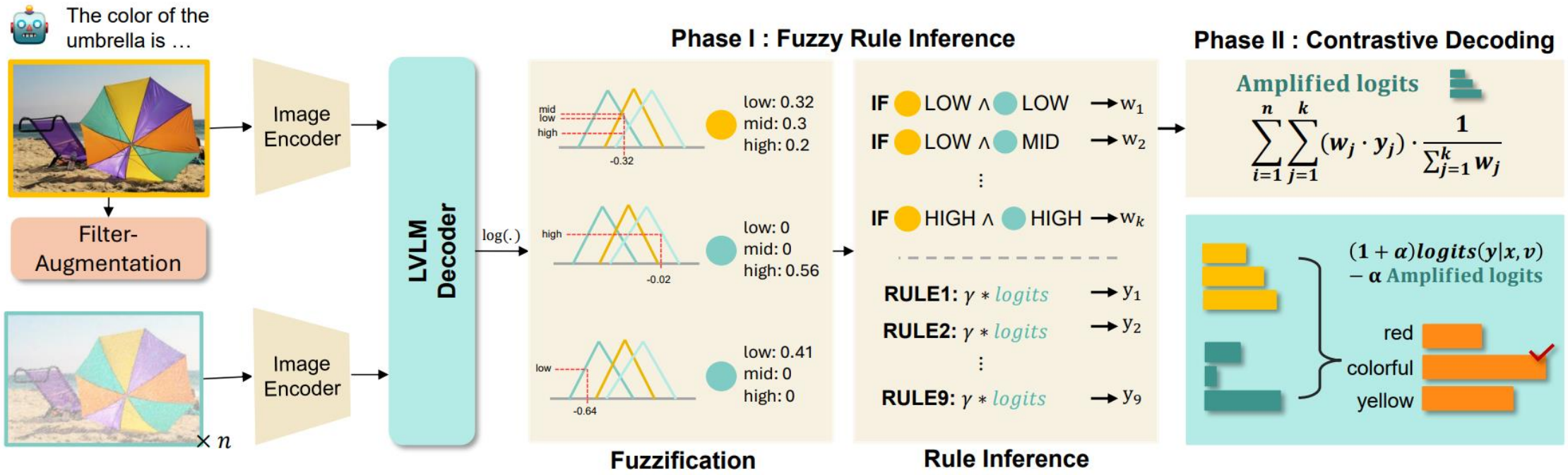


$$y_t \sim P(y_t|x, v, y_{<t})$$

$$\acute{y}_t \sim P(\acute{y}_t|\acute{x}, y_{<t})$$

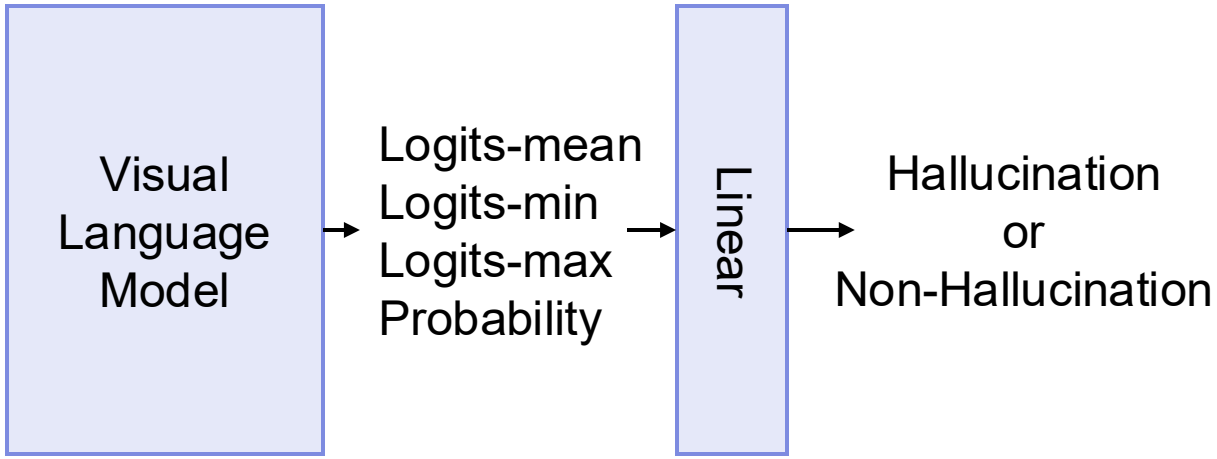
$$y_t \sim (1+\alpha) \cdot y_t - \alpha \cdot \acute{y}_t$$

Overview of Fuzzy Constructive Decoding

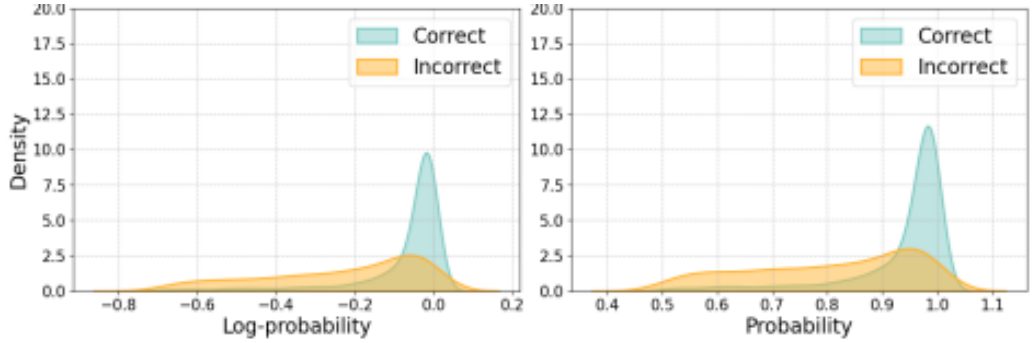


How to get model confidence?

- Logits can serve as confidence scores for hallucination detection.
- Probability & log-probability achieve the best performance

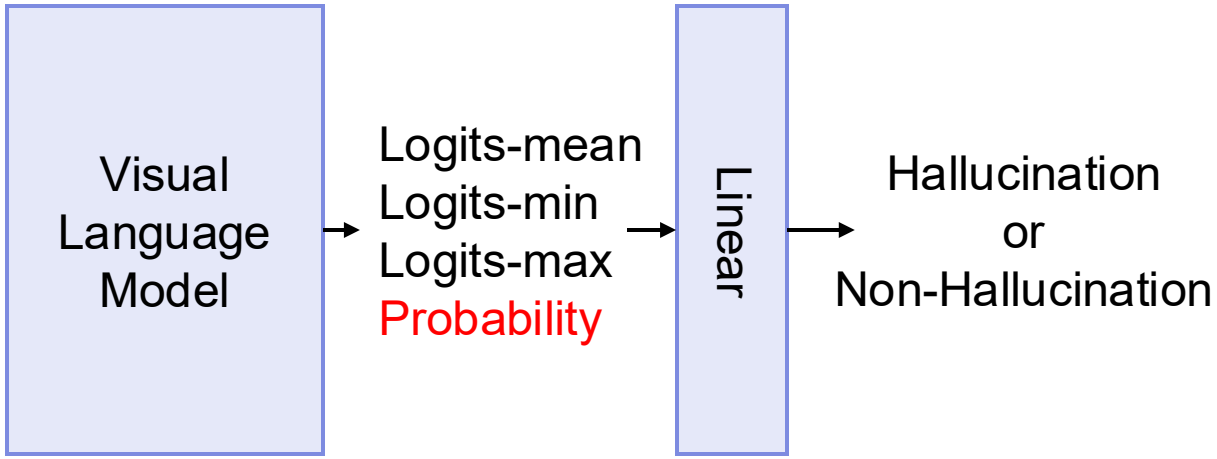


	VizWiz	POPE		
	val	random	popular	adversarial
Logits-mean	0.58	0.24	0.26	0.31
Logits-min	0.56	0.25	0.27	0.31
Logits-max	0.70	0.69	0.72	0.72
Probability	**0.75**	**0.79**	**0.78**	**0.76**
Log-probability	**0.75**	**0.79**	**0.78**	**0.76**

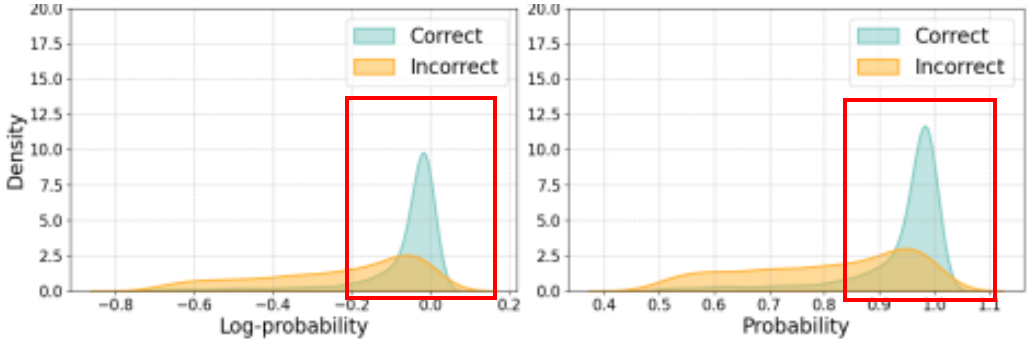


How to get model confidence?

- Logits can serve as confidence scores for hallucination detection.
- Probability & log-probability achieve the best performance



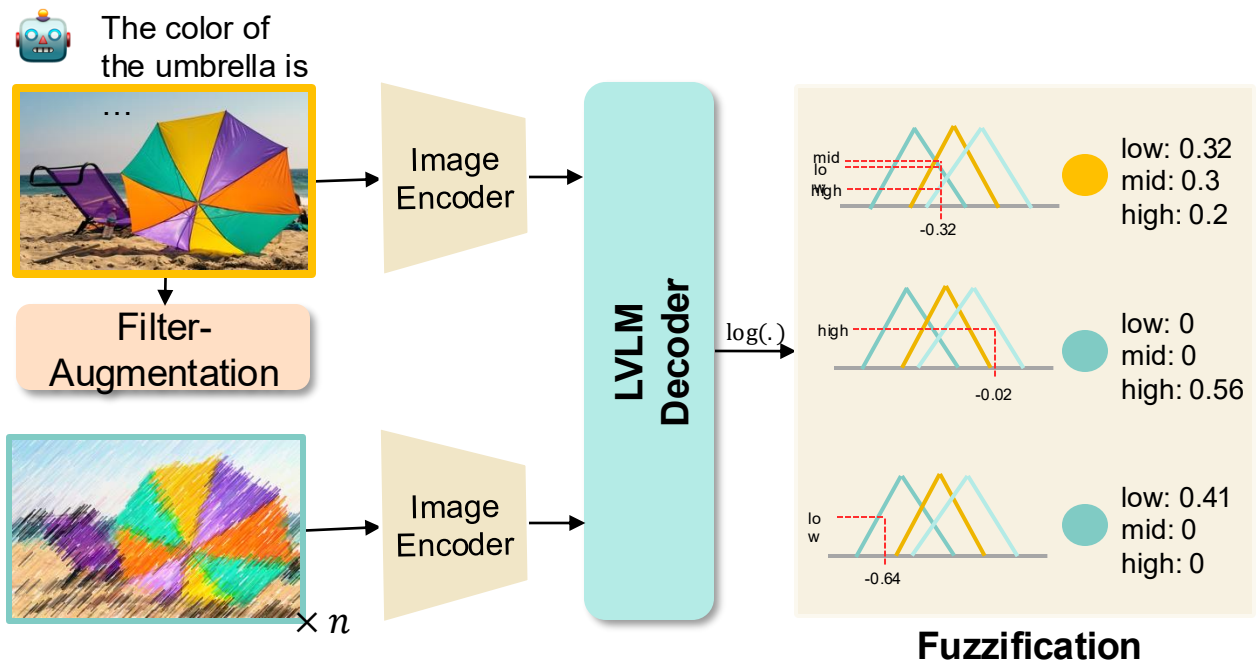
	VizWiz	POPE		
	val	random	popular	adversarial
Logits-mean	0.58	0.24	0.26	0.31
Logits-min	0.56	0.25	0.27	0.31
Logits-max	0.70	0.69	0.72	0.72
Probability	**0.75**	**0.79**	**0.78**	**0.76**
Log-probability	**0.75**	**0.79**	**0.78**	**0.76**



Fuzzy Rule Inference: Fuzzification

- Use log-probability as the model's confidence score
- Sample 10% of the dataset to compute mean ($\bar{x}$) and standard deviation (σ)
- Define Gaussian-based membership functions for three fuzzy sets: low, medium, high
- Enables fuzzy representation of model confidence for later rule inference

$$\mu_k(x) = \exp\left(-\frac{(x-(x_{mean}+k\cdot\sigma)^2}{2\sigma^2}\right),$$
$$k \in \{-1, 0, 1\}$$



Fuzzy Rule Inference: Rule Inference

- Combine confidence memberships from fuzzification to form fuzzy rules
- Apply Takagi–Sugeno fuzzy model to compute rule weights
- Each rule defines a combination of (low, mid, high) confidence levels

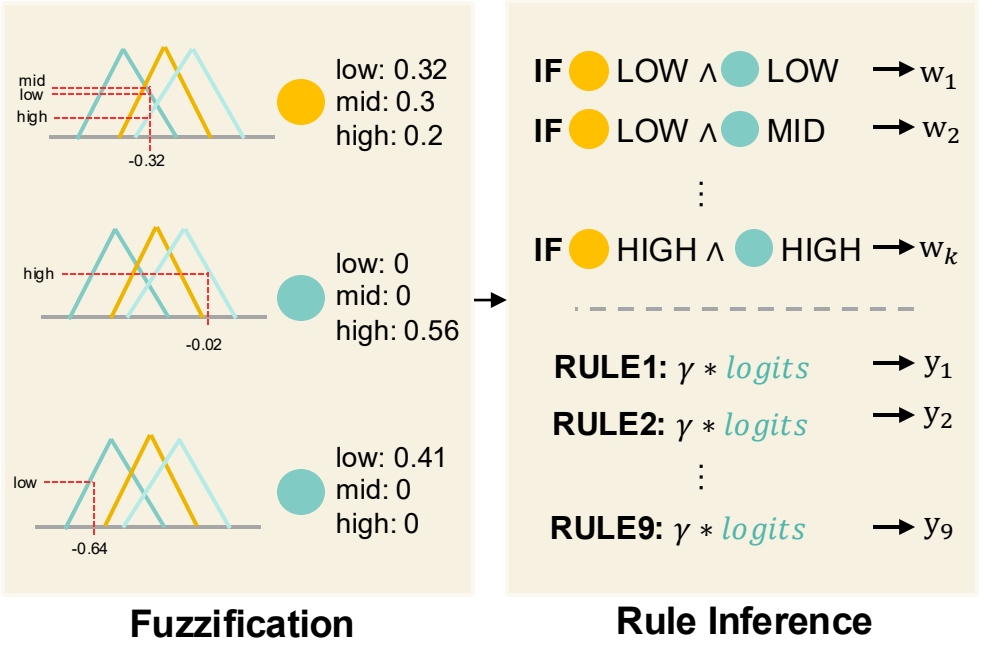
$$w_{rule_i} = \mu(x_1) \cdot \mu_{k`}(x_2) \quad for\ i = 1,2,\dots,9$$

IF LOW ∧ LOW

IF LOW ∧ MID

⋮

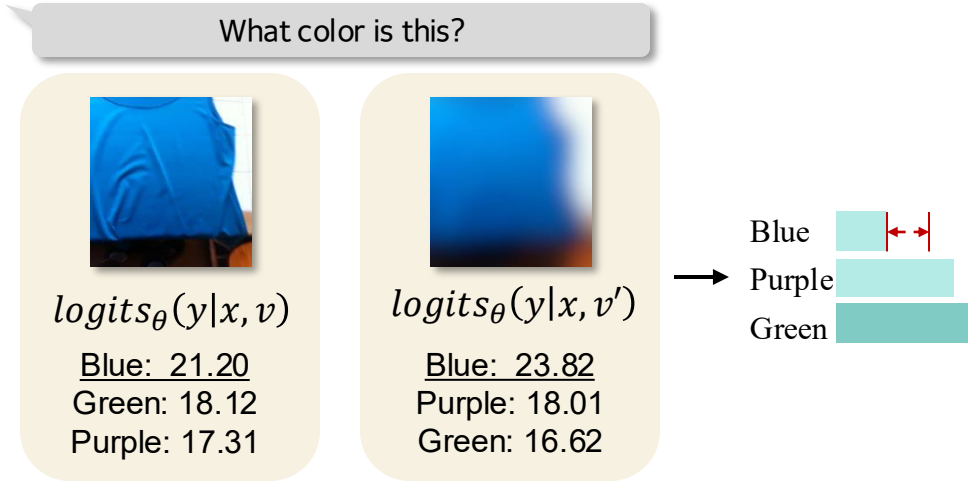
IF HIGH ∧ HIGH



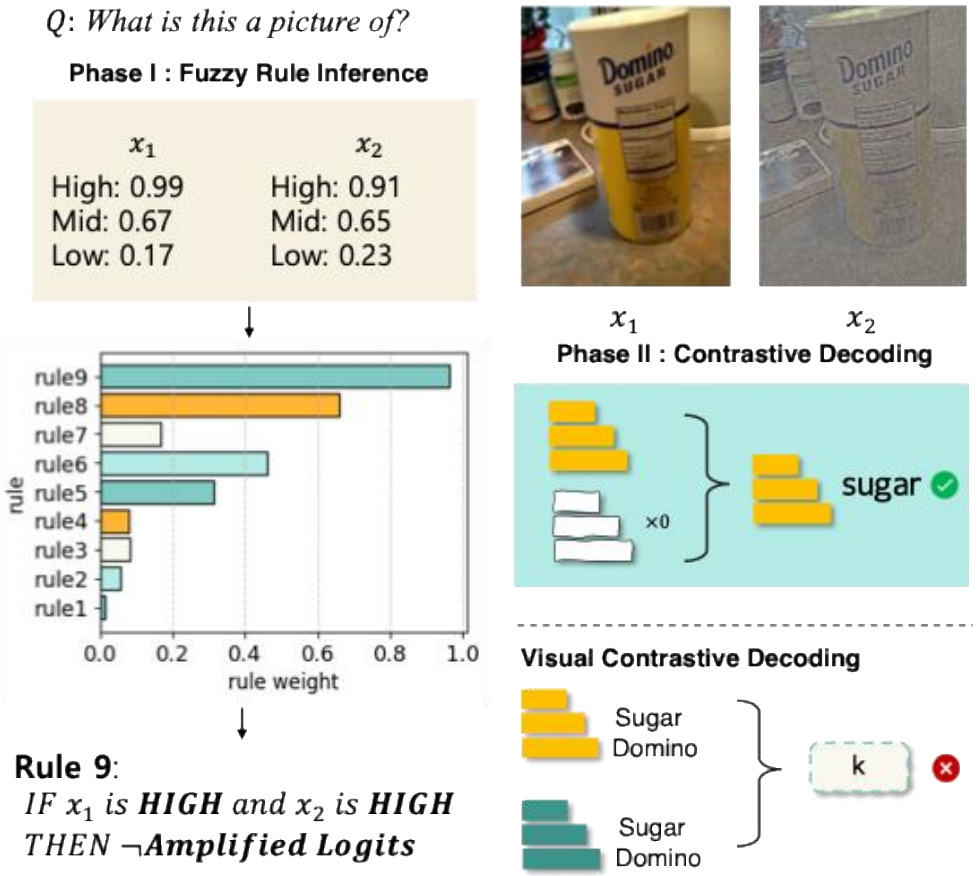
Fuzzification **Rule Inference**

Fuzzy Rule Inference: Rule Inference

- 9 rules: all combinations of *low / mid / high* confidence
- a_{high} , a_{medium} , a_{low} : control amplification strength
- a_{reduce} : prevent over-suppression of correct tokens



$$y_{rule_i} = \begin{cases} a_{high} \cdot q\,, & i \in \{1,4\} \\ a_{medium} \cdot q, & i \in \{2,5\} \\ a_{reduce} \cdot q, & i \in 3 \\ a_{low} \cdot q, & i \in \{6,7,8,9\} \end{cases}$$



Phase II : Constructive Decoding

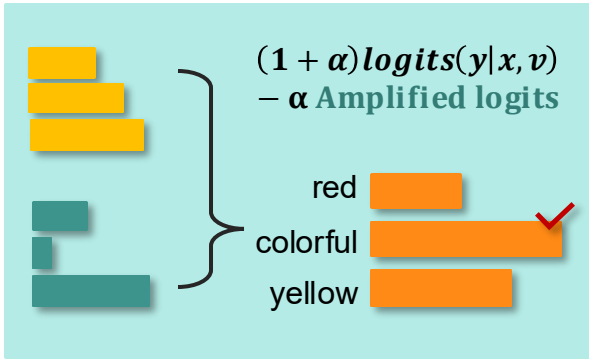
- Combine amplified logits from fuzzy rules using weighted averaging
- Suppress hallucinated tokens while preserving correct visual ones
- Achieves stable and balanced decoding

$$Y_{amplified} = \sum_{i=1}^{n} \sum_{j=1}^{k} \left(\frac{w_j \cdot y_j}{\sum_{j=1}^{k} w_j} \right)$$

Phase II : Contrastive Decoding

Amplified logits

$$\sum_{i=1}^{n} \sum_{j=1}^{k} (w_j \cdot y_j) \cdot \frac{1}{\sum_{j=1}^{k} w_j}$$



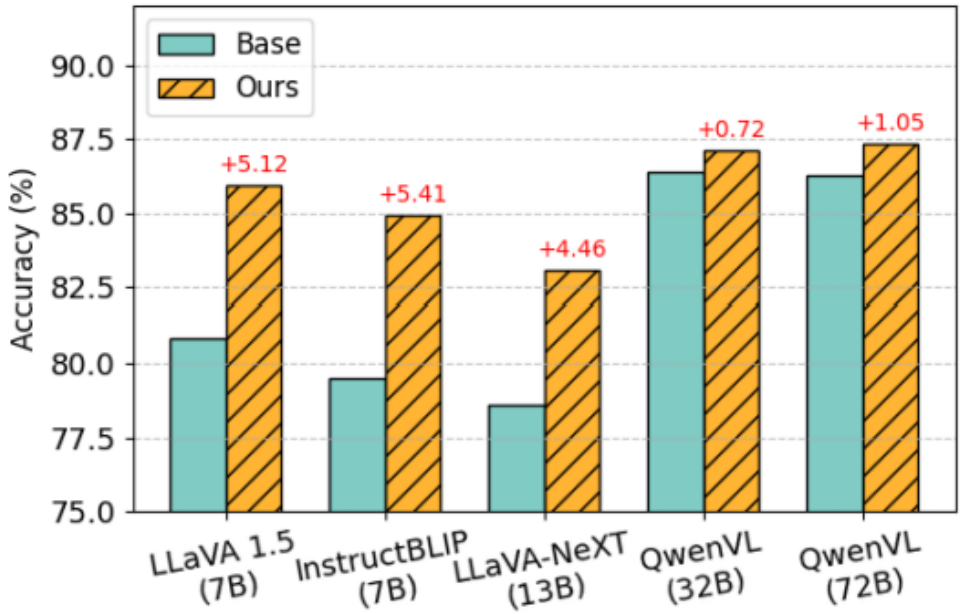
POPE & ROPE

Dataset	Setting	Method	LLaVA-1.5				InstructBLIP			
			Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1
MSCOCO	Random	*default*	82.90	92.00	72.06	80.82	80.63	81.19	79.73	80.46
		+vcd	87.73	91.42	72.80	87.16	84.53	88.55	79.32	83.68
		+icd	83.66	80.66	80.73	**88.77**	86.43	92.01	80.73	85.61
		+vdd	83.52	89.25	76.22	82.22	80.96	82.05	72.26	80.63
		+OPERA	86.26	97.14	74.73	84.47	86.56	90.72	**81.46**	85.84
		+fuzzycd	**88.97**	93.20	**84.07**	88.40	**86.66**	91.16	81.20	**85.89**
	Popular	*default*	81.00	87.88	72.06	79.19	80.50	80.80	80.00	80.40
		+vcd	85.38	86.92	83.28	85.06	81.47	82.89	79.32	81.07
		+icd	80.46	76.39	**88.83**	82.14	82.93	84.45	80.73	82.55
		+vdd	85.87	94.32	76.33	84.38	78.90	78.35	79.86	79.10
		+OPERA	85.26	94.64	74.73	83.53	80.93	75.46	**91.66**	82.78
		+fuzzycd	**86.10**	91.81	79.26	**85.08**	**85.20**	88.26	81.20	**84.58**
	Adversarial	*default*	78.60	82.89	72.06	77.10	77.40	76.11	79.87	77.94
		+vcd	80.88	79.45	83.29	81.33	79.56	79.67	79.39	79.52
		+icd	76.07	70.77	**89.47**	79.03	80.87	80.95	80.73	80.84
		+vdd	83.52	89.25	76.22	82.22	81.63	75.70	81.00	78.26
		+OPERA	83.23	89.84	74.93	81.71	76.33	70.15	**91.66**	79.47
		+fuzzycd	**83.93**	88.61	77.86	**82.89**	**81.90**	82.35	81.20	**81.77**

Model	Multi-Object			Single-Object		
	Wild	Hom	Het	Wild	Hom	Het
LLaVA1.5	13.96	31.88	3.98	13.96	31.88	3.98
+ vcd	13.57	28.68	6.37	24.76	48.33	9.84
+ icd	14.35	32.63	5.69	22.92	45.47	10.24
+ vdd	17.8	40.77	7.40	27.49	52.12	11.06
+ OPERA	13.20	37.14	3.82	13.20	37.14	3.82
+fuzzycd	**21.12**	**46.44**	**7.45**	**29.01**	**57.88**	**11.30**

Scalability

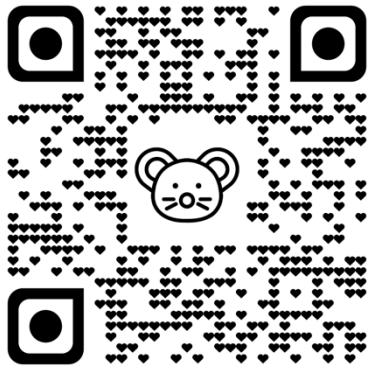
- Consistent accuracy improvement across all models
- Large gains for 7B–13B models (up to +5.4%)
- Stable improvement even in larger models (32B, 72B)
- Demonstrates strong scalability and generalizability



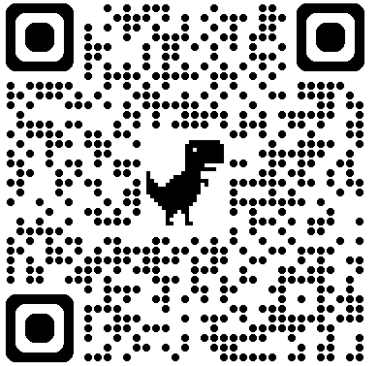
Takeaway and Conclusions

- Fuzzy logic enables the model to detect hallucination tendencies in logits
- Dynamically adjusts contrastive decoding strength based on fuzzy rules
- Balances language priors and visual evidence for stable decoding

Thank you



Contact



Paper

SOFT COMPUTING LABORATORY

YONSEI UNIVERSITY