Towards Robust Uncertainty Calibration for Composed Image Retrieval

Yifan Wang[1], Wuliang Huang[2], Yufan Wen[1], Shunning Liu[1], Chun Yuan[1,*]
1 Tsinghua University, 2 Institute of Computing Technology, CAS

NEURAL INFORMATION PROCESSING SYSTEMS

Motivation

- **Background**: Composed image retrieval **distinguishes positive from negative** samples by **comprehensively understanding semantics** in bimodal queries.
- **Challenges**: Unreliable retrieval results due to *aleatoric uncertainty* and *epistemic uncertainty*.
  - *Aleatoric uncertainty* roots in **implicit semantic overlap in candidate images.**
  - *Epistemic uncertainty* originates from **semantic concept imbalance & over-concentration on salient features.**

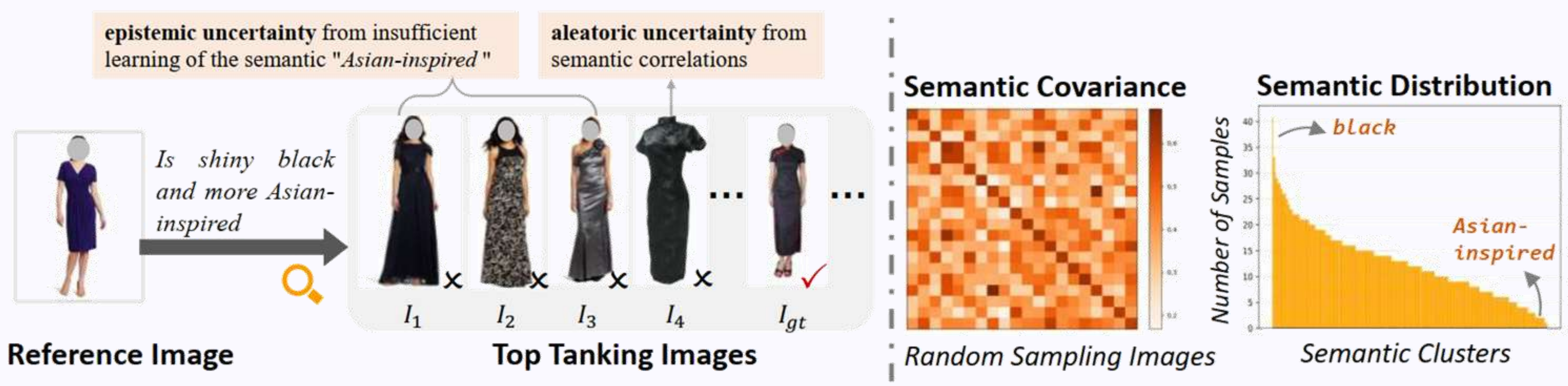


Figure 1: Illustration of uncertain matching results. (*left*) Unreliable top-ranking results disrupted by insufficient learning of "*Asign-inspired* " and partial semantic correlations. (*right*) Strong correlations within the semantic matrix and imbalanced semantic distributions underlying visual candidates.

Methodology

- Employ evidential priors through high-order Normal Inverse Gamma (NIG) distributions to **fit the semantic covariance matrix:**

$$\mathcal{L}_i^{NLL} = -\log(p(\boldsymbol{c}_{ij}|\gamma_i,\nu_i,\alpha_i,\beta_i)) = \frac{1}{2}\log(\frac{\pi}{\nu_i}) - \alpha_i\log\Omega_i + (\alpha_i+\frac{1}{2})\log((\boldsymbol{c}_{ij}-\mu_i)^2\nu_i+\Omega_i) + \log(\frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i+\frac{1}{2})})$$

- Infer sample-level uncertainty to **penalize uncertainty sample** in contrastive learning

$$u_i = \mathbb{E}[\delta_i^2] + \mathrm{Var}[\mu_i] = \frac{\beta_i}{\alpha_i-1} + \frac{\beta_i}{\nu_i(\alpha_i-1)} = \frac{\beta_i(\nu_i+1)}{\nu_i(\alpha_i-1)}$$

- Construct virtual guidance $\boldsymbol{p}^*$ to **differentiate the retention & modification characteristics**:

$$\mathcal{L}_{ort} = \sum_{i\neq j}(\mathrm{Cov}(\boldsymbol{p}^*,\boldsymbol{m})_{ij})^2 = \sum_{i\neq j}(\frac{1}{B}((\boldsymbol{p}^*)^\top\boldsymbol{m})_{ij})^2$$

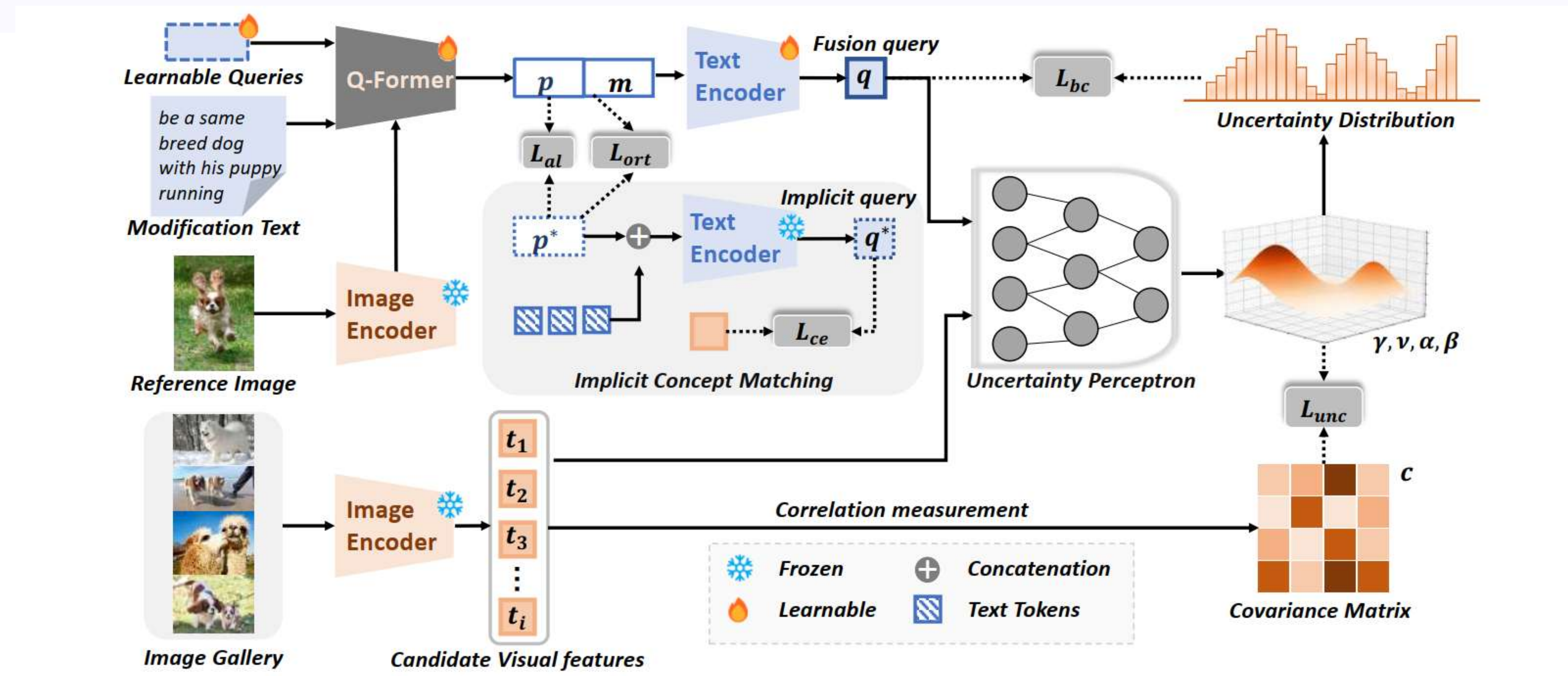
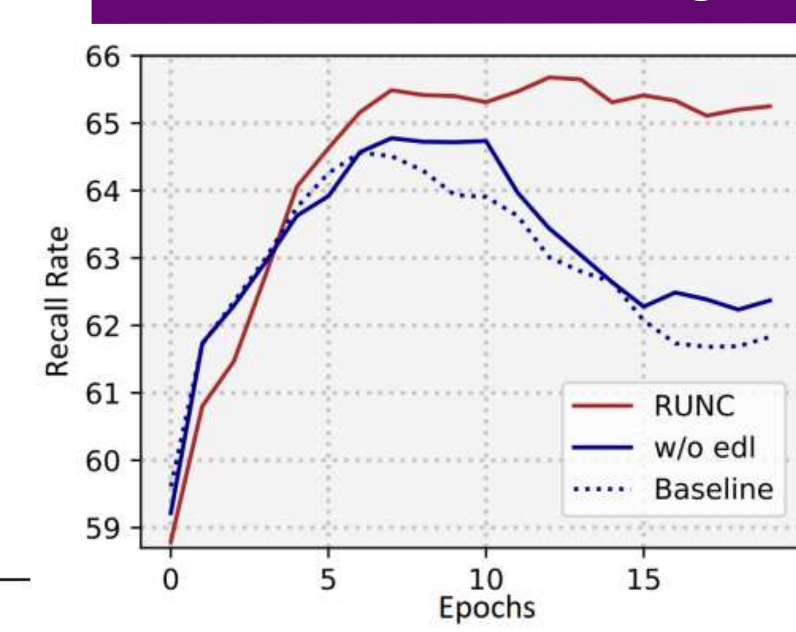


Figure 2: **The Framework of RUNC**. Uncertainty perceptron introduces evidential priors to fit the semantic covariance and yield uncertainty distribution to calibrate the supervision on the fusion query. Implicit guidances p^* are incorporated to distill effective features for retention and modification.

Experiment

Methods	Recall@K				$\mathrm{R_{subset}}$@K			Avg(R@5, $\mathrm{R_{subset}}$@1)
	K=1	K=5	K=10	K=50	K=1	K=2	K=3	
TIRG [1]	11.04	35.08	51.27	83.29	23.82	45.65	64.55	29.45
CIRPLANT [43]	19.55	52.55	68.39	92.38	39.20	63.03	79.49	45.88
ARTEMIS [23]	16.96	46.10	61.31	87.73	39.99	62.20	75.67	43.05
CLIP4Cir [42]	33.59	65.35	77.35	95.21	62.39	81.81	92.02	63.87
CompoDiff [49]	22.35	54.36	73.41	91.77	35.84	56.11	76.60	45.10
BLIP4CIR [20]	40.17	71.81	83.18	95.69	72.34	88.70	95.23	72.07
SSN [50]	43.91	77.25	86.48	97.45	71.76	88.63	95.54	74.51
SPN [32]	45.33	78.07	87.61	98.17	73.93	89.28	95.61	76.00
CaLa [21]	49.11	81.21	89.59	98.00	76.27	91.04	96.46	78.74
SPRC [29]	51.96	82.12	89.74	97.69	80.65	92.31	96.60	81.39
ENCODER [19]	46.10	77.98	87.16	94.64	76.92	90.41	95.95	77.45
RUNC (Ours)	**53.81**	**83.47**	**91.11**	**98.22**	**80.87**	**92.36**	**96.94**	**82.17**

Further Analysis



Analysis of Evidential Learning

Table 4: Analysis of uncertainty estimation on R@50 metric.

Models	Dress	Shirt	Toptee	Mean
GMM	71.69	73.26	77.41	74.12
BMM	71.83	73.99	77.91	74.58
EDC	72.83	75.22	78.74	75.60
MGUR	72.19	74.09	78.73	75.00
MPC	72.29	74.53	78.84	75.22
Ours	**73.53**	**75.32**	**79.86**	**76.23**

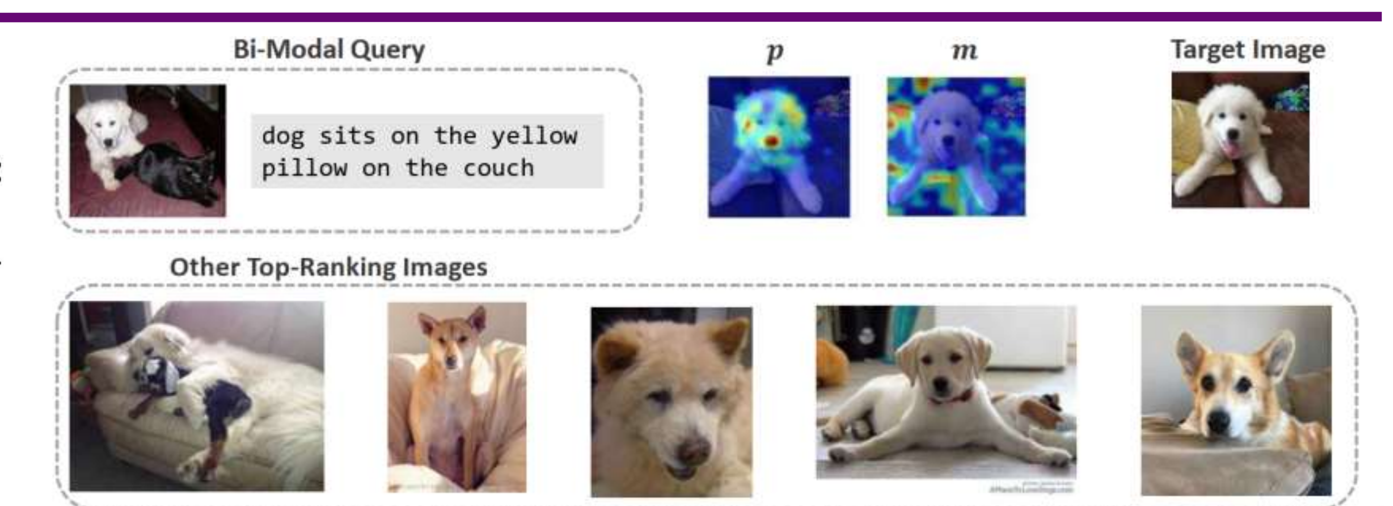


Figure 5: Heatmaps on Target Images with Other Top ranking results.