

UGGRH:

Unsupervised Generative Completion and Graph-attention Refinement for Incomplete Cross-modal Hashing

Yunfei Chen¹ · Yuchen Zhang² · Hongyu Lin³ · Peng Liu² · Zhan Yang^{2*}

¹ Information Support Force Engineering University · ² Central South University · ³ Xiamen University

Paper · Code





1. Motivation: Why Incomplete Cross-modal Hashing?

1

Cross-modal Hashing

Maps image and text into compact binary codes for fast Hamming retrieval. It is highly efficient for large-scale cross-modal search tasks in real-world applications.

2

The Incompleteness Problem

Real-world data often suffers from missing modalities (images or texts). Paired signal breaks make unsupervised hashing unreliable, and performance degrades sharply with high missing rates.

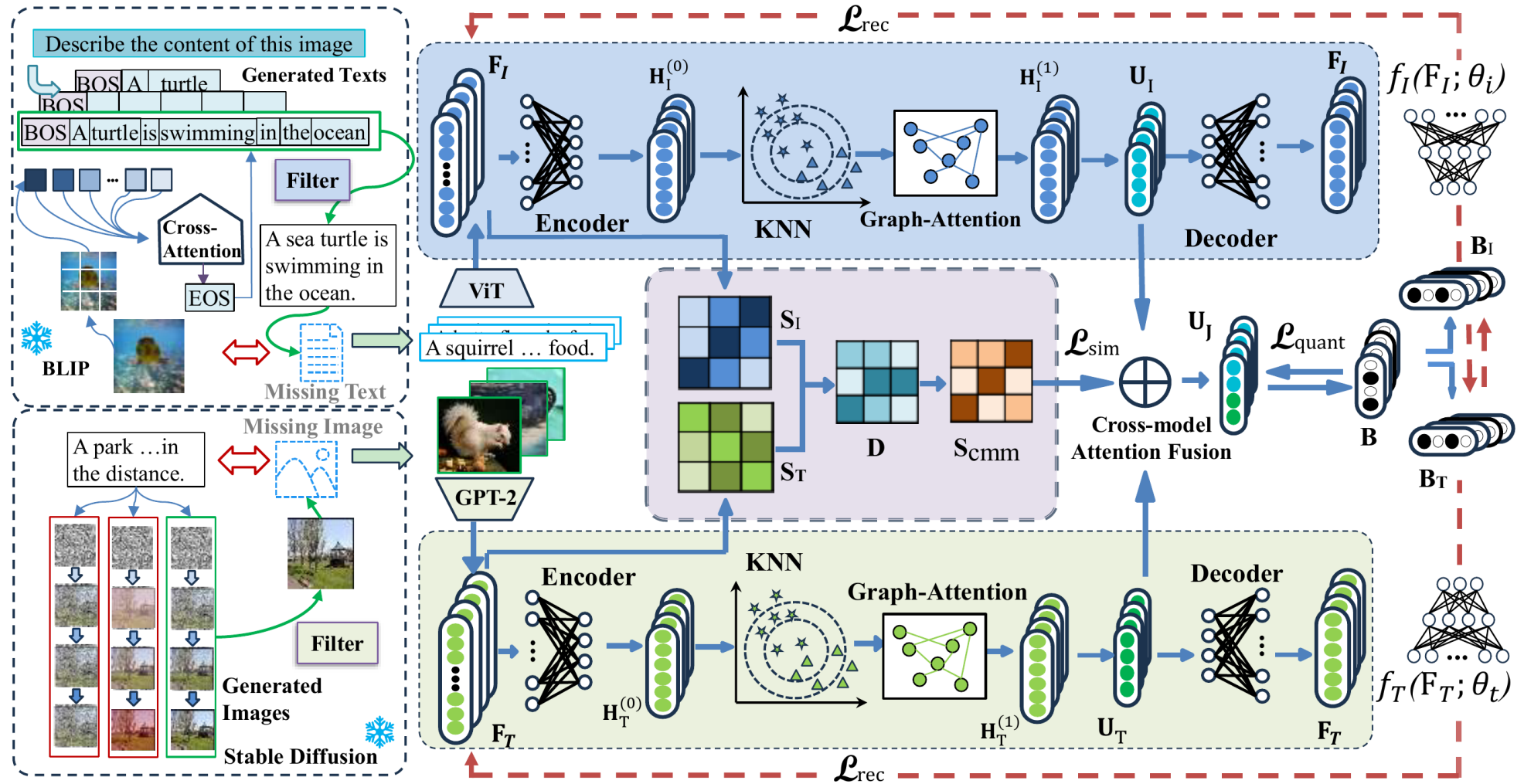
3

Existing Weaknesses

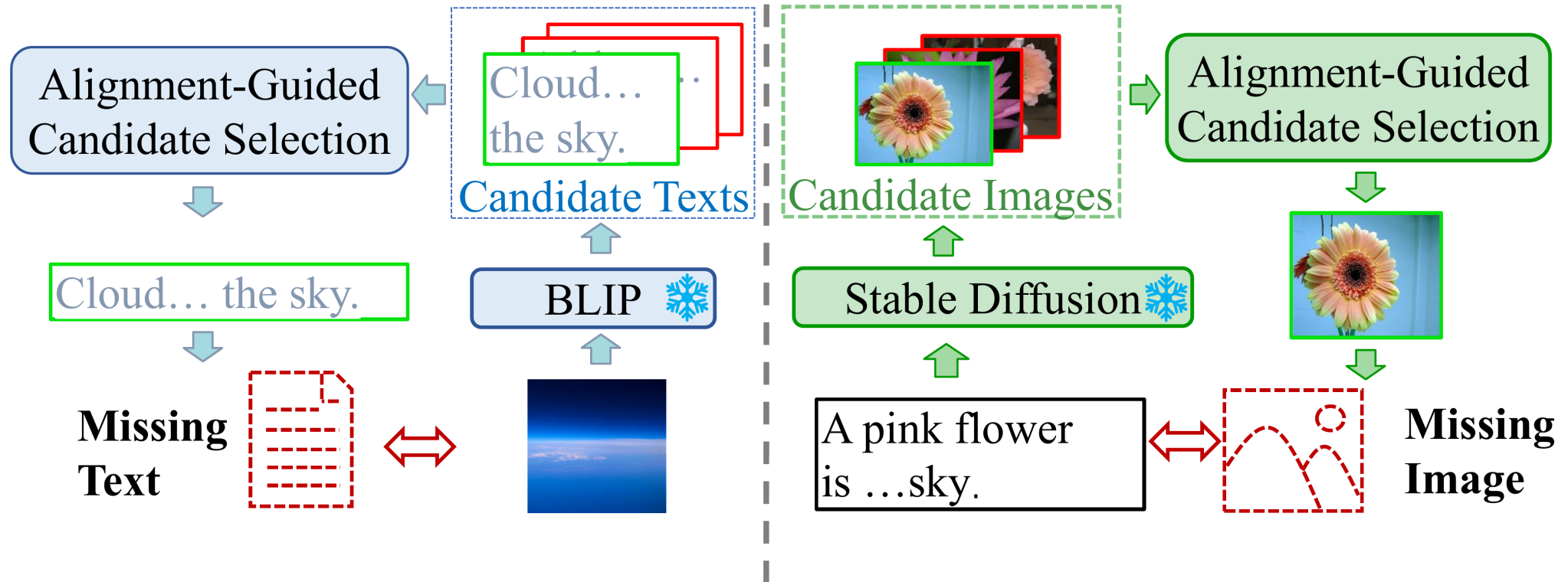
- Naive completion leads to semantic drift
- Similarity matrix polluted by noisy imputation
- Lack of hash-aware message passing mechanism

Key Idea: Offline generative completion with alignment filtering noise-tolerant similarity target hash-aware graph attention mechanism.

2. Overall Framework of UGGRH



3. Module 1 — Offline Generative Completion



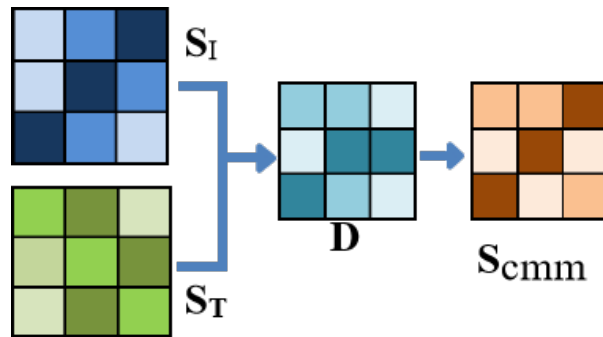
Why Alignment Filtering Matters

Naive generation often suffers from semantic drift, producing candidates that deviate from the original intent. We re-rank these candidates using a frozen CLIP model via cosine similarity alignment, retaining only the best match. The resulting completed feature is cached as a robust input for hashing, not treated as pseudo ground-truth.

4. Similarity Refinement + Hash-aware GAT

Module 2 · Similarity Refinement

Fuse → Top-K sparsify → Smooth → Symmetrize



S_{cmm} : noise-tolerant similarity target

Cross-modal signal is learned by a lightweight MLP on observed pairs (no labels). No external supervision needed for qcalibration.

Module 3 · HashGAT

① Dynamic KNN Graph

Recompute neighborhoods each epoch as embeddings evolve. Captures dynamic semantic shifts in the latent space, avoiding static graph limitations.

② Hash-conditioned Attention

Inject continuous hash proxy z_m to edge weights — fixes the non-differentiability of $\text{sign}(\cdot)$. Stabilizes gradient flow for discrete hash code optimization.

③ Adaptive Cross-modal Fusion

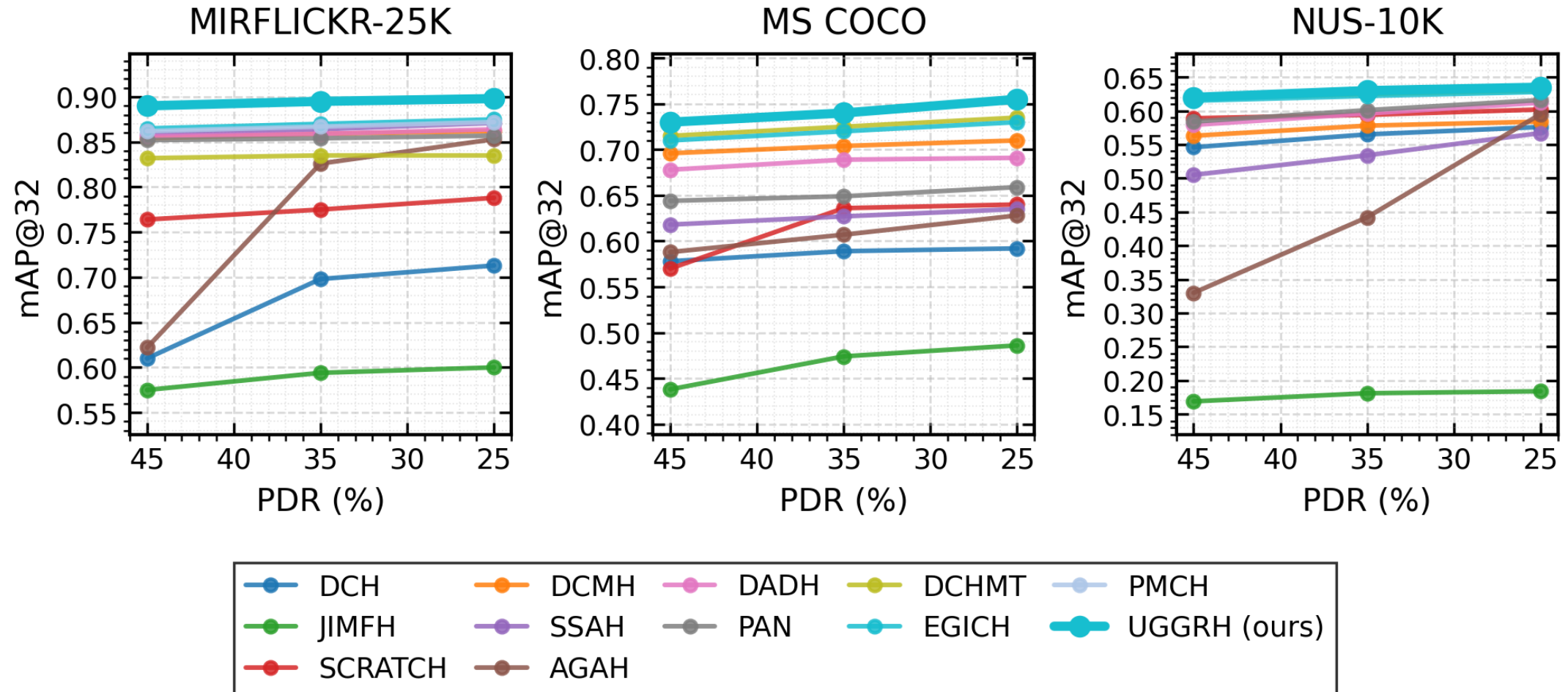
Softmax gate π dynamically weights image, text, joint logits into unified $\mathbf{B}$. Balances multi-modal contributions based on input characteristics.

Training objective: $L = \gamma L_{\text{rec}} + \eta L_{\text{sim}} + \beta L_{\text{quant}}$ (reconstruction + similarity matching + quantization)



5. Experiments — UGGRH Wins Under High Missing Rates

mAP across datasets and missing-data settings





6. Contributions & Takeaways

1

Offline Generative Completion

Frozen BLIP + Stable Diffusion with CLIP alignment filtering — decouples expensive generation from hashing training and reduces semantic drift.

2

Similarity Refinement

Fuse intra-modal cosine with a cross-modal MLP discriminator; top-K sparsification and smoothing yield a noise-tolerant self-supervised target.

3

Hash-aware Graph Attention

Dynamic KNN graphs with hash-conditioned edge attention and adaptive fusion produce unified, discriminative binary codes.

Generation helps most when paired with alignment filtering and noise-tolerant similarity — not as a direct replacement.

Paper · Code





■ Ablation Study (MIRFlickr-25K, 50% PDR)

Component	Variant	I→T				T→I			
		16bits	32bits	64bits	128bits	16bits	32bits	64bits	128bits
Completion	w/o Completion	0.783	0.802	0.813	0.823	0.718	0.736	0.740	0.751
Completion	+GCAE (GCIMH)	0.806	0.815	0.827	0.829	0.744	0.759	0.764	0.768
Completion	+MLP (NCH)	0.801	0.817	0.827	0.828	0.744	0.760	0.769	0.772
Completion	+PMCH	0.826	0.839	0.845	0.845	0.761	0.773	0.778	0.782
Discriminator	w/o Discriminator	0.880	0.894	0.906	0.912	0.881	0.895	0.907	0.911
HashGAT	w/o HAF	0.884	0.888	0.896	0.882	0.882	0.883	0.896	0.894
HashGAT	w/o DGC	0.883	0.893	0.894	0.889	0.882	0.893	0.893	0.889
HashGAT	Standard GAT	0.750	0.717	0.735	0.711	0.744	0.715	0.733	0.710
Generator	OFA + GLIDE	0.879	0.893	0.904	0.908	0.878	0.894	0.903	0.907
Full Model	UGGRH (ours)	0.892	0.907	0.916	0.919	0.890	0.906	0.919	0.918

Every component contributes. Removing discriminator, hash-aware fusion, or dynamic graph construction all degrade performance. Standard GAT collapses.



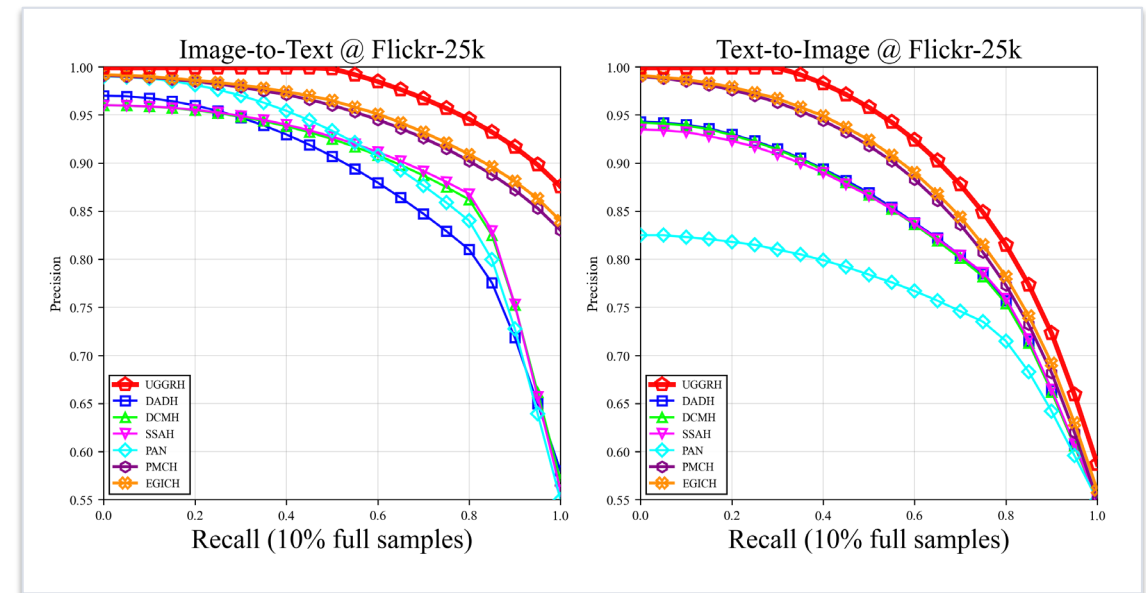
PDR Protocol & Hard-setting PR Curves

Partial Data Ratio (PDR) Protocol

Setting	Complete	Img-Missing	Txt-Missing
Easy	50%	25%	25%
Medium	30%	35%	35%
Hard	10%	45%	45%

Symmetric missing. All methods use identical splits and frozen CLIP features to ensure a fair comparison across different missing data scenarios.

Hard-setting precision–recall



MIRFlickr-25K · I→T and T→I

Same evaluation protocol across easy, medium, and hard settings.



■ Limitations & Future Work

Limitations

- Completion depends on pretrained generators and a frozen alignment model.
- Precomputed similarity has quadratic memory cost.
- Fixed hyperparameters may require cross-dataset tuning.

Future Work

- Use uncertainty-aware filtering for generated candidates.
- Adapt encoders to target domains.
- Explore sub-quadratic similarity construction.

