

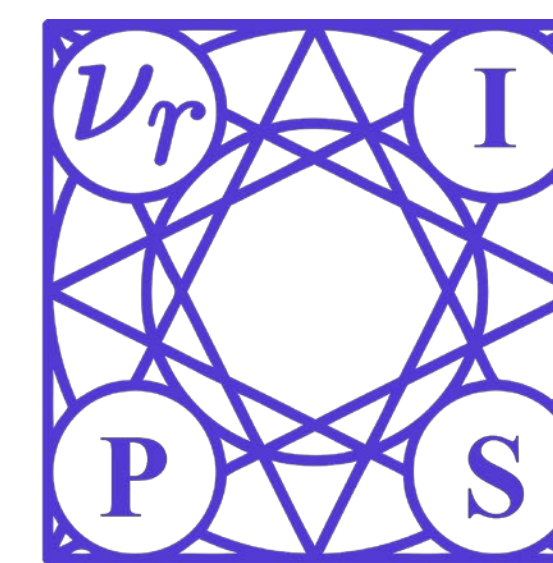


Cross-Modal Learning with Adversarial Samples

Chao Li^{1,2}, Cheng Deng¹, Shangqian Gao², De Xie¹, Wei Liu³

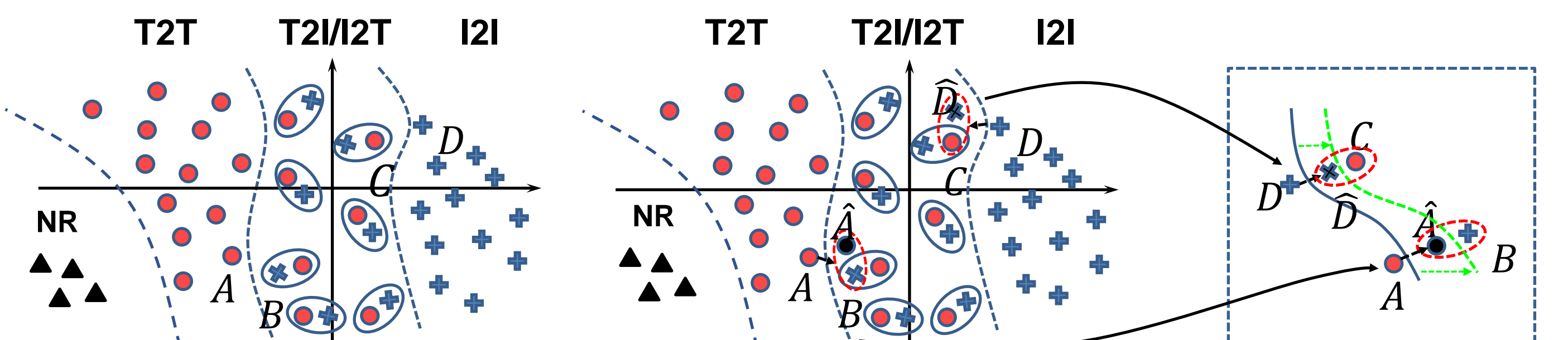
1 School of Electronic Engineering, Xidian University

2 Electrical and Computer Engineering, University of Pittsburgh 3 Tencent AI Lab



Introduction

□ Cross-Modal Learning



(a) Cross-modal Search Space (b) Adversarial Sample Learning (c) Adversarial Training

□ Cross-Modal Hashing Attack

$$\Delta(o^*, H^*) := \min_{\delta^*} \|\delta^*\|_p$$

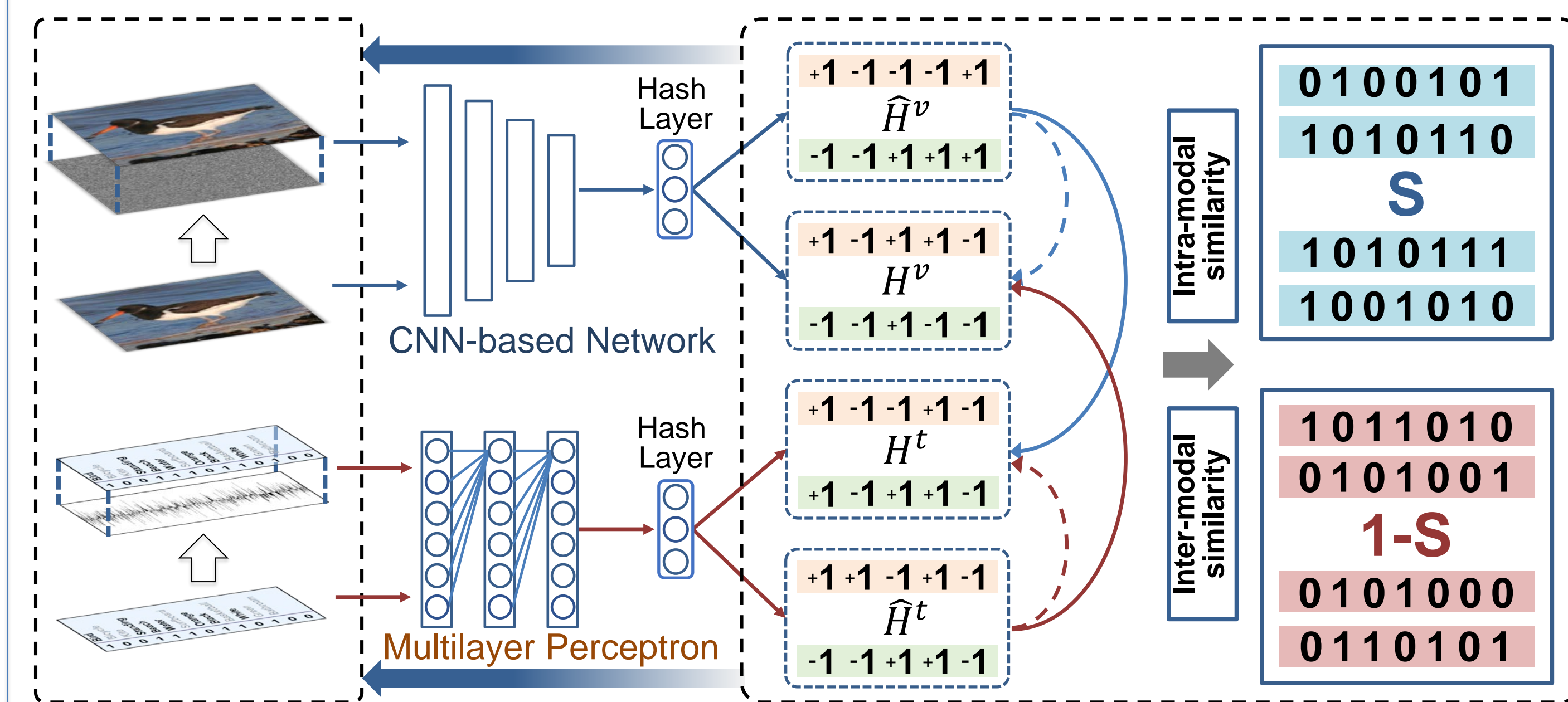
$$s.t. \max_{\delta^*} D(H^*(o^* + \delta^*; \theta^*), H^*(o^*; \theta^*)), \|\delta^*\|_p \leq \varepsilon$$

$$B^* = \text{sign}(H^*), * \in \{v, t\}$$

□ Contribution

- We propose a simple yet effective cross-modal learning method by exploring adversarial samples, where adversarial sample is defined in two aspects: learned perturbation is designed to fool a deep cross-modal network while the distortion on an individual modality will not impact the performance within its own modality.
- A novel cross-modal adversarial sample learning method is presented. We decrease the inter-modality similarity and simultaneously keep intra-modality similarity in one optimization, with which more deceptive samples can be learned.
- We additionally apply the proposed CMLA into cross-modal hashing learning. Experiments on two widely used cross-modal retrieval benchmarks show the effectiveness of our CMLA in attacking a target retrieval system and further improving its robustness.

Framework



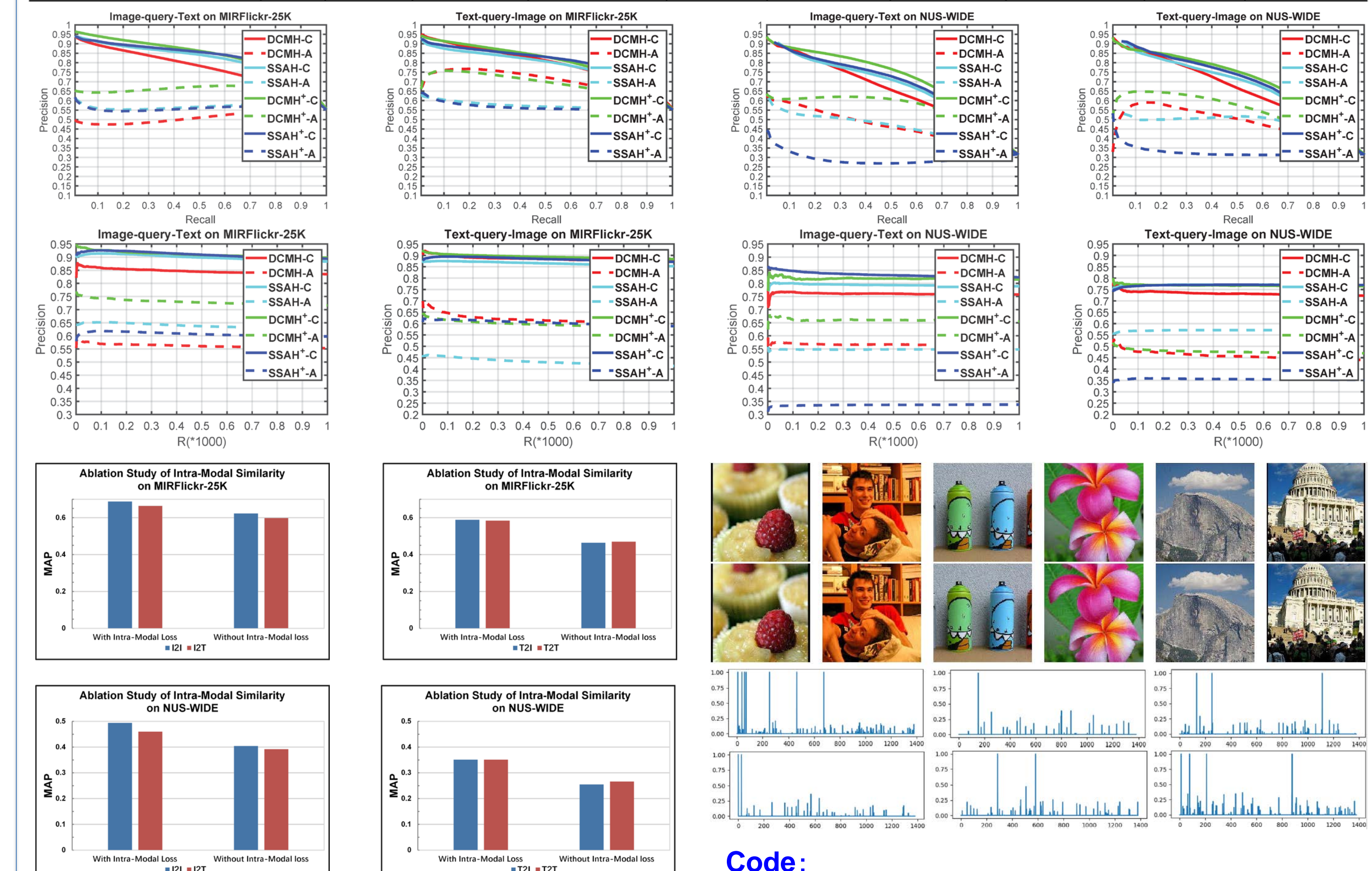
Propose CMLA

$$\min_{\delta^v} J^v = \sum_{i,j=1}^n (\log(1 + e^{-\Gamma_{ij}}) + S_{ij}\Gamma_{ij}) + \sum_{i,j=1}^n (\log(1 + e^{\Theta_{ij}}) - S_{ij}\Theta_{ij}) + \sum_{i=1}^n \|\hat{o}_i^v - o_i^v\|_p$$

$$s.t. \Gamma_{ij} = \frac{1}{2} (\hat{H}_i^v)(H_j^t)^T, \Theta_{ij} = \frac{1}{2} (\hat{H}_i^v)(H_j^v)^T$$

Experiments

Task	Iteration		MIRFlickr-25K				NUS-WIDE			
			DCMH	DCMH ⁺	SSAH	SSAH ⁺	DCMH	DCMH ⁺	SSAH	SSAH ⁺
Image Query v.s. Text Database	0	MAP	0.749	0.816	0.805	0.815	0.607	0.679	0.660	0.675
		D	0.039	0.041	0.034	0.038	0.031	0.033	0.032	0.025
	100	MAP	0.579	0.631	0.679	0.681	0.526	0.609	0.587	0.591
		D	0.023	0.038	0.028	0.032	0.026	0.031	0.029	0.026
	200	MAP	0.563	0.599	0.671	0.699	0.499	0.583	0.534	0.543
		D	0.019	0.029	0.020	0.023	0.025	0.028	0.026	0.024
Text Query v.s. Image Database	0	MAP	0.797	0.820	0.805	0.809	0.614	0.691	0.677	0.685
		D	0.048	0.037	0.031	0.021	0.037	0.035	0.042	0.025
	100	MAP	0.615	0.619	0.603	0.611	0.523	0.628	0.501	0.523
		D	0.027	0.033	0.025	0.019	0.035	0.031	0.035	0.023
	200	MAP	0.587	0.577	0.595	0.605	0.447	0.549	0.454	0.474
		D	0.019	0.021	0.023	0.017	0.030	0.027	0.017	0.019



Code: