FedCAFE: Federated Cross-Modal Hashing with Adaptive Feature Enhancement

Ting Fu, Yu-Wei Zhan, Chong-Yu Zhang, Xin Luo, Zhen-Duo Chen, Yongxin Wang, Xun Yang, Xin-Shun Xu

Shandong University, Shandong Jianzhu University, University of Science and Technology of China

BACKGROUND

Deep Cross-modal hashing has attracted much attention in the large-scale multimedia search area. Existing methods need to first collect data and then be trained with these accumulated data. In many real applications, data may be generated and possessed by **different owners** like devices or institutions. Collecting and centralizing a large amount of distributed data is not only expensive but also impractical. Besides, considering the legal restrictions and growing concerns about data privacy protection, such training schemes may be forbidden as transmitting data may face potential security risks and threats.

Federated learning, as a distributed machine learning framework, can ensure that a global model is trained collaboratively by uniting a series of clients without sharing local raw data, which solves the problem that local data cannot be shared under privacy and security requirements. Therefore, using federated learning to accomplish distributed training for deep cross-modal hashing methods is a feasible solution.

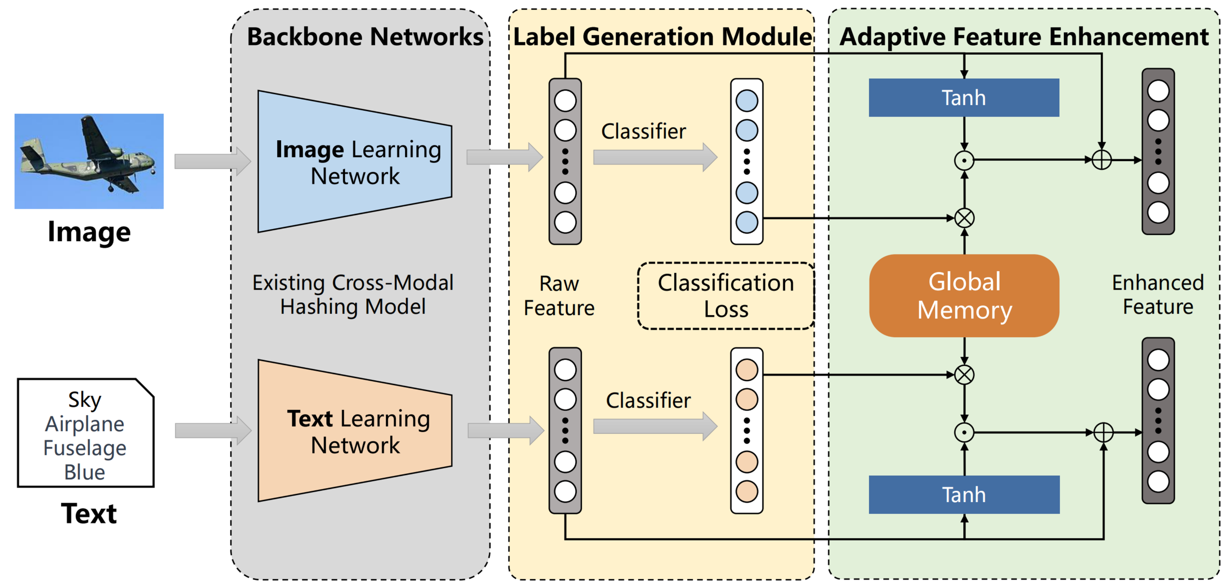
CONTRIBUTIONS

➢A novel framework **Federated Cross-modal Hashing with Adaptive Feature Enhancement (FedCAFE)** is proposed . FedCAFE is endowed with a novel adaptive feature enhancement module and a new weighted aggregation strategy. Besides, it could fully utilize the rich global information carried in the global model to constrain the model during the local training process.

➢Extensive experiments are conducted on four widely-used datasets with both IID and non-IID settings. The proposed FedCAFE achieves better retrieval accuracy than all chosen baselines.

METHOD

➢The framework of our method



➢Memory in global and local

➢Local memory: the centers of each class (acquiring by K-means) $P_k^t \in \mathbb{R}^{C\mathrm{x}r}$

➢Global memory: the centers of each class from local clients (acquiring by K-means) $P_g^t \in \mathbb{R}^{C\mathrm{x}r}$

➢Adaptive feature enhancement module

➢Constraints imposed on the raw features

$$V_M^* = \hat{L}^* \otimes P_g,\ V_E^* = O^* + A \odot V_M^*,\ \mathcal{L}_E = 1 - \cos(V_E^*, \boldsymbol{O}^*),$$

➢Global guidance in cross modal correlations

$$\mathcal{L}_C = \frac{1}{2}(1 - \cos(V_E^{*-k}, V_E^x)) + \frac{1}{2}(1 - \cos(V_E^{*-k}, V_E^y)),$$

➢KL divergence for classification ability

$$\mathcal{L}_P = KL\left(\hat{L}^* | L\right),$$

➢Overall Loss Function

$$\mathcal{L}_{all} = \mathcal{L}_{hash} + \alpha\mathcal{L}_E + \eta\mathcal{L}_C + \gamma\mathcal{L}_P,$$

$\boldsymbol{O}^*$, V_E^* denotes the raw features and enhanced features.

➢Weighted aggregation strategy

➢the similarity between the local memory of k-th client and the global memory

$$sim_k = -\sum_{i,j=1}^{C}\left(S_{ij}\theta_{ij} - log\left(1 + e^{\theta_{ij}}\right)\right),\quad \theta_{ij} = \frac{1}{2}\left(P_{k_{i\cdot}}^t\right)\left(P_{g_{j\cdot}}^t\right)^T$$

➢the degree of difference between global memory and local memory as weight

$$\alpha_k = \frac{\log(sim_k)}{\sum_k^K \log(sim_k)}$$

➢the global model parameters for different modalities in round $(t+1)$

$$W_x^{(t+1)} = \sum_{k=1}^{K} \alpha_k\, W_{x-k}^{(t)},\ W_y^{(t+1)} = \sum_{k=1}^{K} \alpha_k\, W_{y-k}^{(t)}$$

EXPERIMENTS

➢MAP results on MIRFlickr-25K and NUS-WIDE.

Table 2: The MAP results of different federated learning methods on MIRFlickr-25K and NUS-WIDE.

Methods	MIRFlickr-25K						NUS-WIDE					
	Image-to-Text			Text-to-Image			Image-to-Text			Text-to-Image		
	16bits	32bits	64bits	16bits	32bits	64bits	16bits	32bits	64bits	16bits	32bits	64bits
Centralized [10]	0.7383	0.7427	0.7527	0.7654	0.7669	0.7749	0.5903	0.6031	0.6093	0.6389	0.6511	0.6571
Local [10]	0.6405	0.6473	0.6544	0.6723	0.6821	0.6922	0.5042	0.5269	0.5418	0.4875	0.5025	0.5223
FedAvg [34]	0.6610	0.6734	0.6836	0.6907	0.7040	0.7143	0.5381	0.5581	0.5567	0.5186	0.5526	0.5609
FedProx [18]	0.6597	0.6724	0.6829	0.6899	0.7033	0.7133	0.5352	0.5565	0.5564	0.5176	0.5501	0.5599
FedCMR [55]	0.6729	0.6856	0.6956	0.7129	0.7229	0.7318	0.5572	0.5668	0.5735	0.5569	0.5792	0.5851
MOON [17]	0.6740	0.6862	0.6948	0.7081	0.7219	0.7317	0.5650	0.5840	0.5758	0.5664	0.5753	0.5867
FedProto [42]	0.6748	0.6904	0.7018	0.7173	0.7278	0.7364	0.5687	0.5880	0.5928	0.5731	0.5829	0.6066
PLFedCMH [24]	0.6531	0.6840	0.6992	0.6998	0.7221	0.7359	0.5738	0.5817	0.5852	0.5948	0.6028	0.6070
FedCAFE	**0.7113**	**0.7206**	**0.7294**	**0.7451**	**0.7554**	**0.7648**	**0.5987**	**0.6425**	**0.6567**	**0.6065**	**0.6582**	**0.6617**

➢MAP results on FashionVC and Ssense in Non-IID settings.

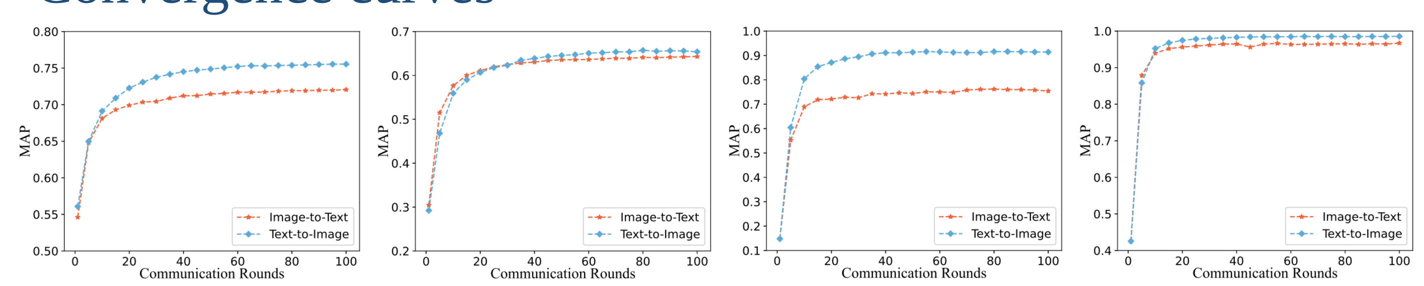
Table 3: The MAP results of different federated learning methods with varying degrees of data heterogeneity on FashionVC.

Methods	$\beta = 0.5$						$\beta = 0.2$					
	Image-to-Text			Text-to-Image			Image-to-Text			Text-to-Image		
	16bits	32bits	64bits	16bits	32bits	64bits	16bits	32bits	64bits	16bits	32bits	64bits
Centralized [49]	0.7620	0.7635	0.7616	0.9414	0.9527	0.9515	0.7620	0.7635	0.7616	0.9414	0.9527	0.9515
Local [49]	0.4006	0.4552	0.4585	0.3465	0.3928	0.3925	0.3128	0.3556	0.3639	0.2586	0.2837	0.2897
FedAvg [34]	0.5852	0.6753	0.7220	0.7368	0.8582	0.8976	0.4681	0.5502	0.6573	0.6321	0.7425	0.8204
FedProx [18]	0.5680	0.6615	0.7089	0.6837	0.8245	0.8623	0.4731	0.5777	0.6592	0.5171	0.6579	0.7580
FedCMR [55]	0.6119	0.6982	0.7303	0.7579	0.8660	0.8943	0.5200	0.5933	0.6611	0.6396	0.7521	0.8230
MOON [17]	0.5820	0.6808	0.7285	0.7024	0.7933	0.8638	0.4739	0.5722	0.6606	0.6404	0.6419	0.7311
FedProto [42]	0.6978	0.7099	0.7248	0.8808	0.9158	0.9032	0.5630	0.5785	0.5823	0.7663	**0.8381**	0.8109
PLFedCMH [24]	0.6027	0.6840	0.7198	0.7196	0.7726	0.8374	0.4974	0.5823	0.5808	0.5370	0.7583	0.7617
FedCAFE	**0.7257**	**0.7603**	**0.7527**	**0.8935**	**0.9169**	**0.9093**	**0.6576**	**0.6840**	**0.7007**	**0.7903**	0.8234	**0.8310**

Table 4: The MAP results of different federated learning methods with varying degrees of data heterogeneity on Ssense.

Methods	$\beta = 0.5$						$\beta = 0.2$					
	Image-to-Text			Text-to-Image			Image-to-Text			Text-to-Image		
	16bits	32bits	64bits	16bits	32bits	64bits	16bits	32bits	64bits	16bits	32bits	64bits
Centralized [49]	0.9692	0.9697	0.9720	0.9820	0.9881	0.9880	0.9692	0.9697	0.9720	0.9820	0.9881	0.9880
Local [49]	0.7161	0.7776	0.7895	0.6763	0.7436	0.7546	0.4888	0.5322	0.5518	0.4388	0.4752	0.4917
FedAvg [34]	0.8936	0.9517	0.9620	0.9386	0.9734	0.9840	0.7653	0.9059	0.9528	0.8028	0.9447	0.9792
FedProx [18]	0.8567	0.9424	0.9590	0.8789	0.9626	0.9773	0.7952	0.8964	0.9390	0.8152	0.9226	0.9715
FedCMR [55]	0.9033	0.9592	0.9669	0.9214	0.9781	0.9848	0.8414	0.9364	0.9552	0.8690	0.9632	0.9827
MOON [17]	0.9446	0.9628	**0.9683**	0.9549	0.9782	0.9845	0.7871	0.9205	0.9510	0.8028	0.9238	0.9806
FedProto [42]	0.9387	0.9620	0.9676	0.9531	0.9826	0.9845	0.9579	0.9585	**0.9613**	0.9789	0.9836	0.9825
PLFedCMH [24]	0.9545	0.9595	0.9557	0.9752	0.9827	0.9824	0.8832	0.8983	0.9014	0.8851	0.9366	0.9437
FedCAFE	**0.9665**	**0.9671**	0.9658	**0.9856**	**0.9856**	**0.9860**	**0.9582**	**0.9595**	0.9570	**0.9801**	**0.9847**	**0.9832**

➢Convergence curves



(a) MIRFlickr-25K (b) NUS-WIDE (c) FashionVC, $\beta = 0.5$ (d) Ssense, $\beta = 0.5$

Figure 2: Convergence curves of FedCAFE in 100 communication rounds on four datasets.

➢For more experimental results, please refer to our paper.

CONTACT

Ting Fu
futingxha@gmail.com
School of Software
Shandong University, China

MIMA Group
http://mima.sdu.edu.cn/