Task	Method	Wiki				NUS-WIDE				MIRFlickr			
		16	32	64	128	16	32	64	128	16	32	64	128
Image query Text	CVH	0.179	0.162	0.153	0.149	0.272	0.262	0.406	0.39	0.606	0.599	0.596	0.598
	IMH	0.201	0.203	0.204	0.195	0.47	0.473	0.476	0.459	0.612	0.601	0.592	0.579
	LCMH	0.115	0.124	0.134	0.149	0.354	0.361	0.389	0.383	0.559	0.569	0.585	0.593
	CMFH	0.251	0.253	0.259	0.263	0.455	0.459	0.465	0.467	0.621	0.624	0.625	0.627
	LSSH	0.197	0.208	0.199	0.195	0.481	0.489	0.507	0.507	0.584	0.599	0.602	0.614
	DBRC	0.253	0.265	0.269	0.288	0.424	0.459	0.447	0.447	0.617	0.619	0.62	0.621
	RDFH	0.242	0.246	0.244	0.243	0.488	0.492	0.494	0.508	0.632	0.636	0.641	0.652
	UDCMH	**0.309**	**0.318**	**0.329**	**0.346**	**0.511**	**0.519**	**0.524**	**0.558**	**0.689**	**0.698**	**0.714**	**0.717**
Text query Image	CVH	0.252	0.235	0.171	0.154	0.401	0.384	0.442	0.432	0.591	0.583	0.576	0.576
	IMH	0.467	0.478	0.453	0.456	0.478	0.483	0.472	0.462	0.603	0.595	0.589	0.58
	LCMH	0.132	0.142	0.154	0.157	0.376	0.387	0.408	0.419	0.561	0.569	0.582	0.582
	CMFH	0.595	0.601	0.616	0.622	0.529	0.577	0.614	0.645	0.642	0.662	0.676	0.685
	LSSH	0.569	0.593	0.593	0.595	0.577	0.617	0.642	0.663	0.637	0.659	0.659	0.672
	DBRC	0.574	0.588	0.598	0.599	0.455	0.459	0.468	0.473	0.618	0.626	0.626	0.628
	RDFH	0.59	0.596	0.603	0.61	0.612	0.641	0.658	0.68	0.681	0.693	0.698	0.702
	UDCMH	**0.622**	**0.633**	**0.645**	**0.658**	**0.637**	**0.653**	**0.695**	**0.716**	**0.692**	**0.704**	**0.718**	**0.733**

Table 1: mAP results of three datasets at various code lengths (bits). The best performance is shown in boldface.

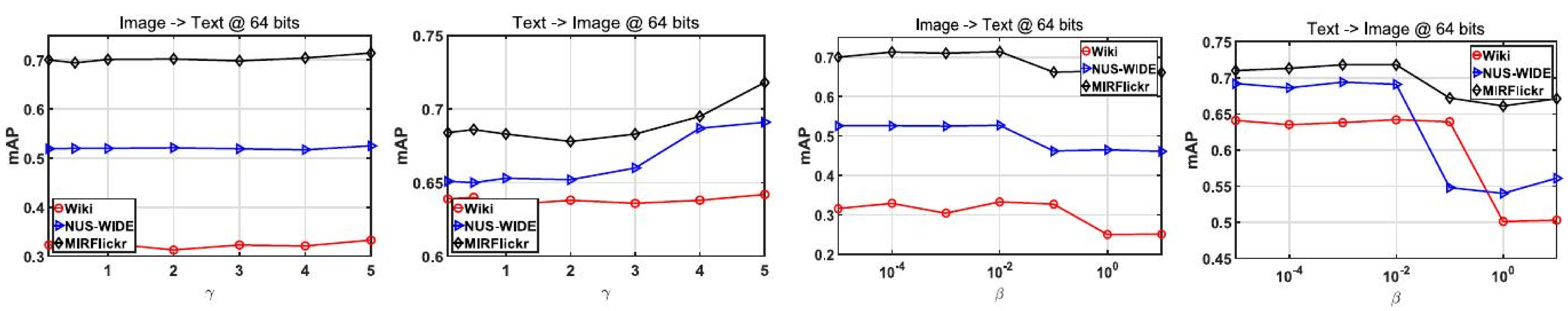

Figure 4: mAP versus γ and β at 64 bits on three datasets.

competitive. Compared to those baselines using matrix factorization such as CMFH and RFDH, the proposed framework integrates deep learning with collaborative binary latent representation model, where unified binary codes can be optimized directly without relaxation, thus improving the hash code quality significantly. For the purpose of comprehensive investigation, we also plot the topN-precision curves of various tasks on all datasets in Figure 3. As observed from those figures, the best performance is still achieved by the proposed UDCMH and the claimed superiority can be further validated. Moreover, we evaluate the system performance with the parameter variations in learning the binary latent representation, where γ is a positive number that controls the weight of each modality and β measures the impact of the Laplacian constraints. Only mAP@64 bits is reported for the limited space. Firstly, the mAP values remain stable when γ is varied from 0.1 to 5 for the image query text task, as shown in left two sub figures of Figure 4. As for another task, the mAP slightly increases when γ is larger than 3 and reaches the maximum at 5. Then we evaluate the retrieval performance when varying β in a wide range of [0.00001, 10]. As can be seen from Figure 4, the mAP values decrease when β is larger than 0.01 to some extent for different tasks.

Task	Training Size				
	2k	5k	10k	15k	20k
Image to Text	0.515	0.524	0.538	0.549	0.557
Text to Image	0.668	0.695	0.698	0.717	0.724

Table 2: Effect of training size on NUS-WIDE at 64 bits.

We further analyze the effects on mAP results when varying the training size, as shown in Table 2. For the limited space, only the results on NUS-WIDE at 64 bits are reported. The mAP values keep increasing when utilizing more training data. It is worth noting that competitive results still can be achieved when using only 5,000 data points by UDCMH, which indicates its powerful ability in producing effective hash codes with the limited data size.

4 Conclusion

In this paper, we presented a novel unsupervised deep multimodal hashing framework, UDCMH, which integrates the deep networks with the proposed binary latent representation learning. Particularly, the unified binary codes can be optimized in an alternating manner for different modalities by solving the discrete constrained objective function directly, thus avoiding the large quantization errors. Moreover, the Laplacian constraint is constructed for each modality and utilized in the binary code learning with the neighborhood structure of original data preserved. Besides, UDCMH adopts a dynamic strategy in assigning weights for different modalities during optimization. Extensive experiments on three datasets show the superiority over several state-of-the-art baselines.