We have conducted supplementary experiments as requested by reviewer 3E4C and present the results here.

1. The influence of the number of sub-proxies

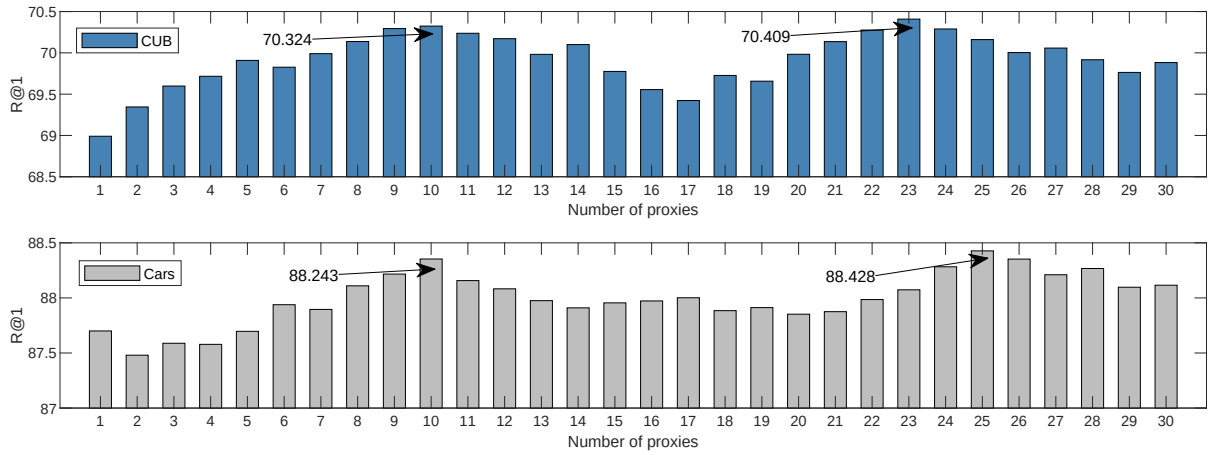

Figure 1: Recall@1 (%) performance of the proposed DMA method with different numbers of sub-proxies (controlled by the parameter P) on the CUB and Cars datasets.

Figure 1 illustrates the impact of the number of the sub-proxies of the proposed DMA on the image retrieval problem of the CUB and Cars data. From the figure, it can be seen that, for the CUB data, when the number of the sub-proxies P is larger than 1, the proposed DMA achieves higher Recall@1 scores than that with $P = 1$, which indicates that multiple sub-proxies fit the data distribution of the CUB images better than a single proxy. Particularly, when the number of the sub-proxies was increased, the Recall@1 score was improved generally, with its first local peak appearing at $P = 10$ and second local peak at $P = 23$. The experimental phenomenon on the Cars data is quite similar to that on the CUB data where the highest local peak appears at $P = 25$, except that the performance of the proposed method is inferior to the single proxy solution when $P < 6$.

From the above phenomena, we see that choosing a proper P is important for the optimal performance of DMA. In order to balance computational complexity and performance, we set P to 10 for both datasets.

2. The influence of parameter δ

Table 1: Recall@1 (%) performance of the proposed DMA method with different margin (controlled by the parameter δ) on the CUB and Cars datasets.

Datasets	δ								
	0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
CUB	69.278	**70.324**	69.801	69.193	69.024	68.923	54.828	54.828	54.828
Cars	87.935	**88.243**	87.615	87.308	45.542	45.542	45.542	45.542	45.542

Table 1 illustrates the impact of different margins on the performance in the CUB and Cars dataset. It can be observed that the proposed DMA achieves best performance on both datasets when $\delta = 0.1$.

Moreover, it can also be seen that when δ is too large, the recall@1 score drops sharply. For the CUB dataset, the turning point is 0.6, while for the Cars dataset, it is 0.4.