Dynamic Metric Learning with Cross-Level Concept Distillation Supplementary Material

Wenzhao Zheng[1,2], Yuanhui Huang[1,2],
Borui Zhang[1,2], Jie Zhou[1,2], and Jiwen Lu[1,2,⋆]

[1] Department of Automation, Tsinghua University, China
[2] Beijing National Research Center for Information Science and Technology, China
{zhengwz18, huang-yh18, zhang-br21}@mails.tsinghua.edu.cn;
{jzhou, lujiwen}@tsinghua.edu.cn;

A Detailed Dataset Setting

We follow existing work to conduct experiments on the three dynamic metric learning datasets: DyML-Vehicle, DyML-Animal, and DyML-Product [5]. We detail the dataset setting as follows.

- The DyML-Vehicle dataset [5] is composed of 454.7K vehicle re-ID images collected from PKU VehicleID [3] and VERI-Wild [4]. For training, we use 343.1K images labeled with 5, 89, and 36,301 classes for the coarse, middle, and fine scale, respectively. For testing, we use 5.9K, 34.3K, and 63.5K images labeled with 6, 127, and 8,183 classes for the coarse, middle, and fine scale, respectively.
- The DyML-Animal dataset [5] is composed of 446.8K animal images collected from ImageNet-5K [1]. For training, we use 407.8K images labeled with 5, 28, and 495 classes for the coarse, middle, and fine scale, respectively. For testing, we use 12.5K, 23.1K, and 11.3K images labeled with 5, 17, and 162 classes for the coarse, middle, and fine scale, respectively.
- The DyML-Product dataset [5] is composed of 448.6K online product images selected from iMaterialist-2019 [2]. For training, we use 747.1K images labeled with 36, 169, and 1,609 classes for the coarse, middle, and fine scale, respectively. For testing, we use 1.5K images labeled with 6, 37, and 315 classes for the coarse, middle, and fine scale, respectively.

B Further Experimental Analysis

Ablation study about the refiner: We conducted an ablation study to analyze different designs of the refiner, as shown in Table 1. We see that using a constant dimension for multi-level concepts leads to slightly inferior performance degradation, yet using a parallel design results in much worse performance. This is because using a parallel refiner fails to exploit the hierarchical label structure, rendering the image representation less aware of the relations between hierarchical concepts.

[⋆] Corresponding author.

Table 1: Ablation study of the refiner on DyML-Product.

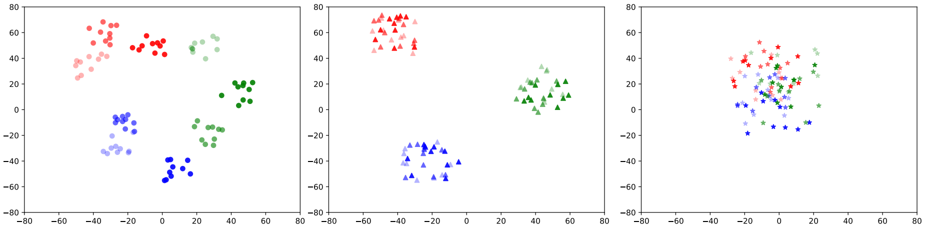
Dimension	Layer	Fine level		Middle level		Coarse level		Overall	
		mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1
Constant	Parallel	7.1	23.0	14.9	52.4	50.3	84.4	24.1	53.3
Decreasing	Parallel	7.3	23.5	16.3	53.3	51.3	85.1	25.0	54.0
Constant	Serial	12.1	28.2	21.0	57.8	52.7	87.8	28.6	57.9
Decreasing	Serial	**13.9**	**29.4**	**22.4**	**59.2**	**54.2**	**89.8**	**30.2**	**59.5**

Table 2: Effect of different $\gamma(k)$ on DyML-Product.

$\gamma(k)$	DyML-Vehicle				DyML-Animal				DyML-Product			
	mAP	R@1	R@10	R@20	mAP	R@1	R@10	R@20	mAP	R@1	R@10	R@20
$k-1$ (ACR)	16.0	42.9	74.0	84.1	**36.0**	**57.1**	**85.2**	**90.1**	29.4	58.8	86.2	90.7
$k-2$	**16.6**	43.1	74.5	85.2	35.4	56.3	84.8	89.9	30.0	59.1	86.4	91.4
0 (ICR)	**16.6**	**43.7**	**75.4**	**86.3**	35.7	56.0	84.8	89.7	**30.2**	**59.5**	**87.1**	**92.1**

Effect of different $\gamma(k)$: We proposed two strategies to distill concepts: ACR ($\gamma(k)=k-1$) and ICR ($\gamma(k)=0$), where different $\gamma(k)$ indicates using concepts from different levels to distill new concepts. As $K=3$ in all the datasets, we can only further evaluate $\gamma(k)=k-2$, as shown in Table 2.

Visualization of concept embeddings: We provide a t-SNE visualization of the multi-level concept embeddings in Figure 1. We randomly select 1 coarse-level label, 3 middle-level labels, and 9 fine-level labels on DyML-animal. We then sample 10 images for each fine-level label for visualization. We use different colors and transparencies to denote different middle-level and fine-level labels, respectively. We plot the concept embeddings of the fine, middle, and coarse levels in Figures 1a, 1b, and 1c, respectively. We observe that the concept embeddings from each level are able to cluster samples with the same label of the corresponding level. This verifies the effectiveness of the proposed concept distillation method.



(a) Fine-level (b) Middle-level (c) Coarse-level

Fig. 1: T-SNE visualization of the concept embeddings.

References

1. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: CVPR. pp. 248–255 (2009) 1
2. Han, X., Guo, S., Huang, W., Kong, L., Scott, M.: imaterialist challenge on product recognition (2019), https://github.com/MalongTech/imaterialist-product-2019 1
3. Liu, H., Tian, Y., Wang, Y., Pang, L., Huang, T.: Deep relative distance learning: Tell the difference between similar vehicles. In: CVPR. pp. 2167–2175 (2016) 1
4. Lou, Y., Bai, Y., Liu, J., Wang, S., Duan, L.: Veri-wild: A large dataset and a new method for vehicle re-identification in the wild. In: CVPR. pp. 3230–3238 (2019) 1
5. Sun, Y., Zhu, Y., Zhang, Y., Zheng, P., Qiu, X., Zhang, C., Wei, Y.: Dynamic metric learning: Towards a scalable metric space to accommodate multiple semantic scales. In: CVPR. pp. 5393–5402 (2021) 1