Supplementary material:

Learning from Extrinsic and Intrinsic Supervisions for Domain Generalization

Anonymous ECCV submission

Paper ID 759

1 Overview

In this supplemental material, we provide the following contents.

- We show feature clustering results using extracted semantic features by employing t-SNE [3] in Section 2.
- We show activation map results to indicate class-specific image regions using Class Activation Mapping (CAM) [5] in Section 3.
- We present comparison results against another state-of-the-art domain generalization method on an additional dataset Office-Home [4] in Section 4.

2 t-SNE Embedding

We demonstrate t-SNE visualization results by comparing the baseline model with our proposed method in Figures 1, 2, 3, and 4 below on the PACS dataset [2]. From the comparison results, we can see that features extracted from our designed network are better clustered and more distinctive among the different categories.

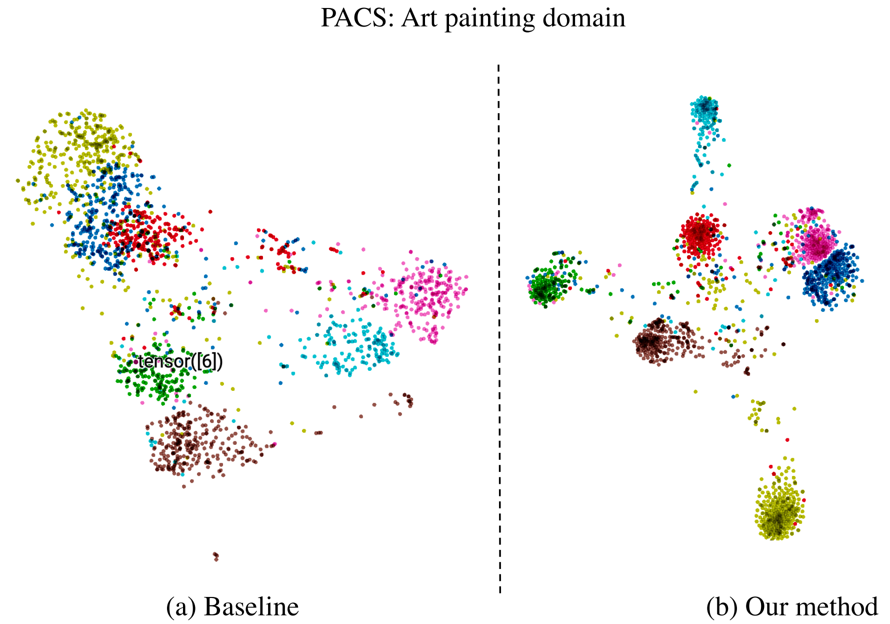

Fig. 1. t-SNE visualization in Art painting domain of PACS dataset.

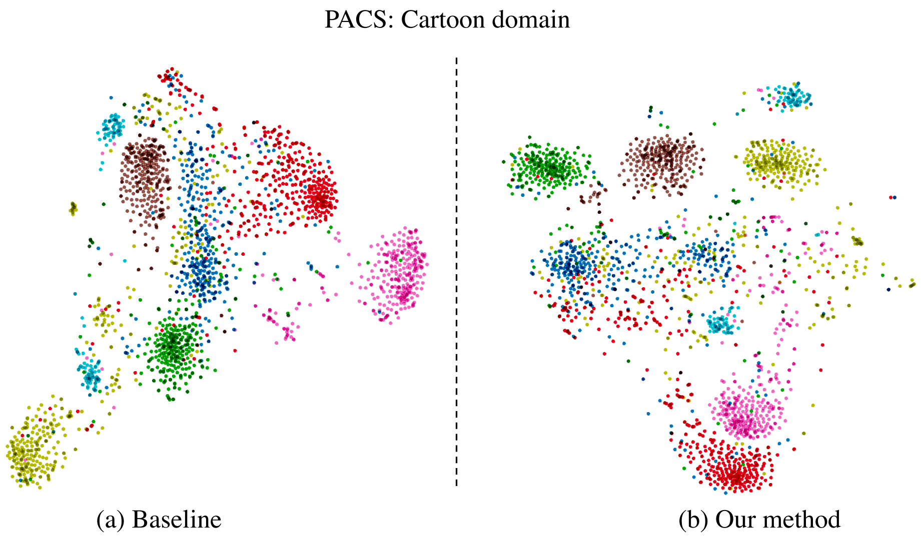

Fig. 2. t-SNE visualization in Cartoon domain of PACS dataset.

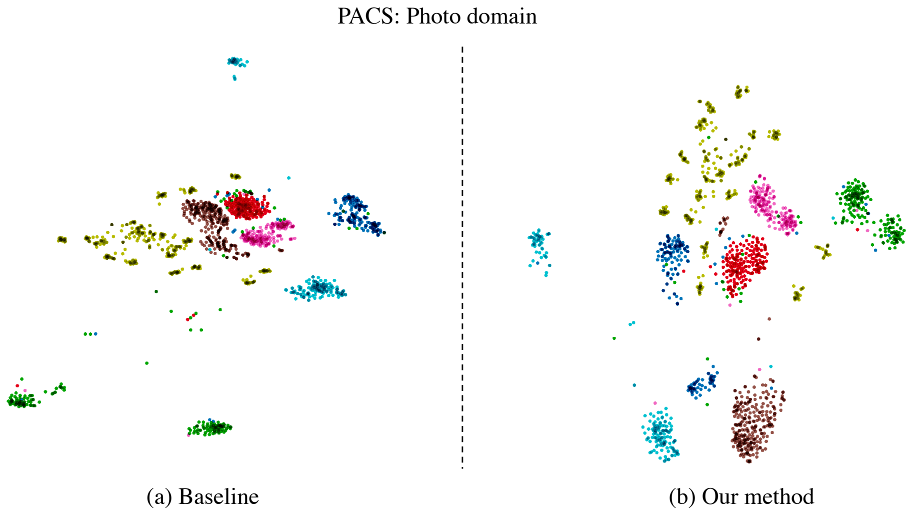

Fig. 3. t-SNE visualization in Photo domain of PACS dataset.

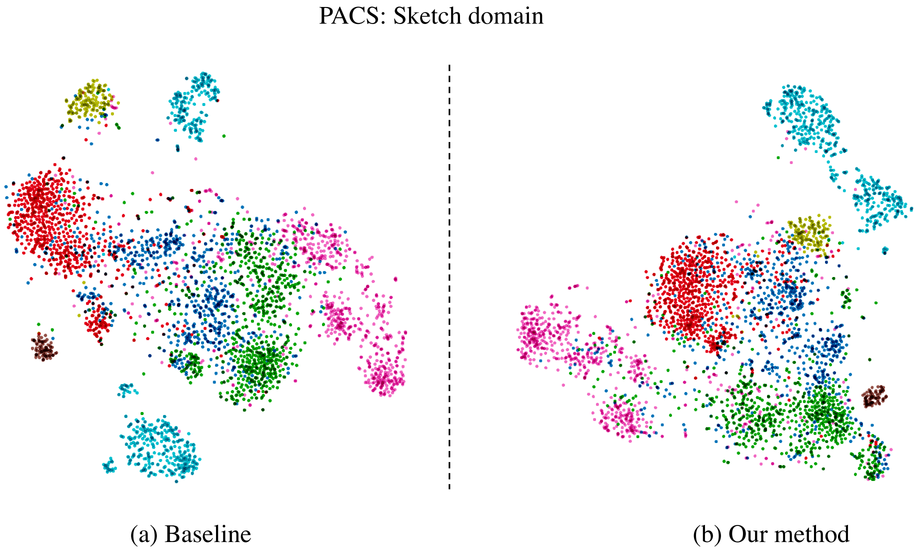

Fig. 4. t-SNE visualization in Sketch domain of PACS dataset.

3 Class Activation Maps

We compare our method with the baseline model in the localizations of class-specific regions on the PACS dataset [2] and results are shown in Figures 5, 6, 7, and 8 below. From the comparison results, it is shown that our method could recognize object categories with more meaningful regions of high activation values.

4 Results on the Office-Home Dataset

Table 1. Domain generalization results on the **Office-Home** dataset with object recognition accuracy (%) on the **ResNet15** backbone. The top results are highlighted in **bold**.

Target	**ResNet-18**		
	DeepAll	**JiGen** [1]	**Ours**
Art	52.15	53.04	**58.43**
Clipart	45.86	**47.51**	47.48
Product	70.86	71.47	**73.03**
Real-World	73.15	72.79	**73.47**
Average	60.51	61.20	**63.10**

The Office-Home dataset consists of four domains: Art, Clipart, Product, and Real-World. Each domain contains images from 65 categories, which indicates the challenges of the Office-Home dataset under the unsupervised setting. The comparison results are shown in Table 1 above. **DeepAll** is the baseline method that uniformly trains a network with images from all the domains. **JiGen** [1] adopts an additional self-supervision task by solving puzzles to constrain the network. It is observed that our proposed method outperforms JiGen by 1.9% on average.

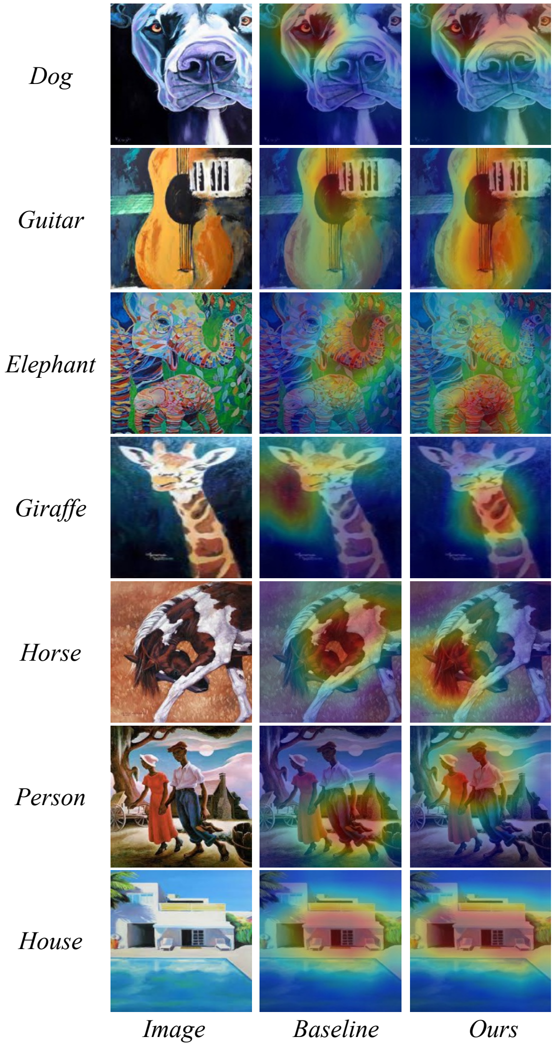

Fig. 5. More result comparison in Art domain of PACS dataset.

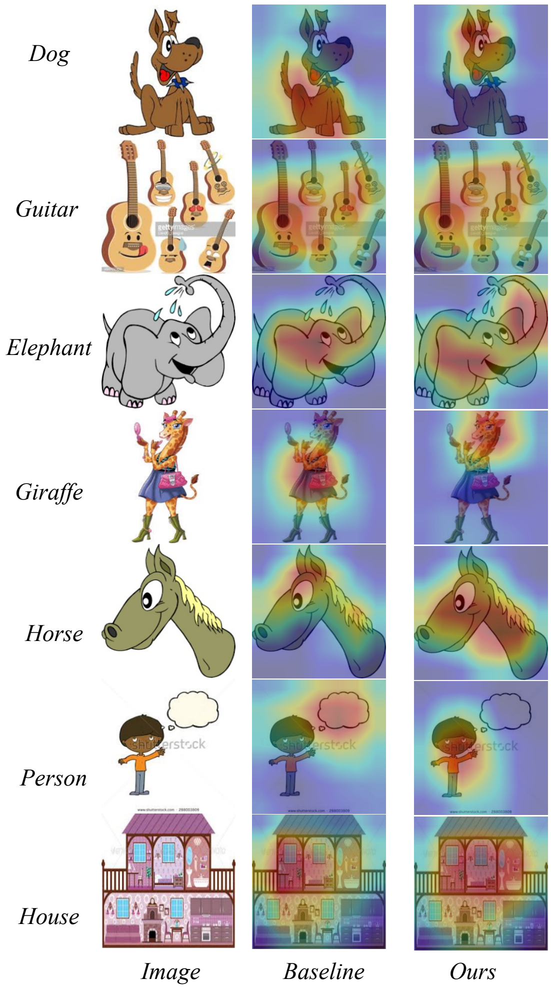

Fig. 6. More result comparison in Cartoon domain of PACS dataset.

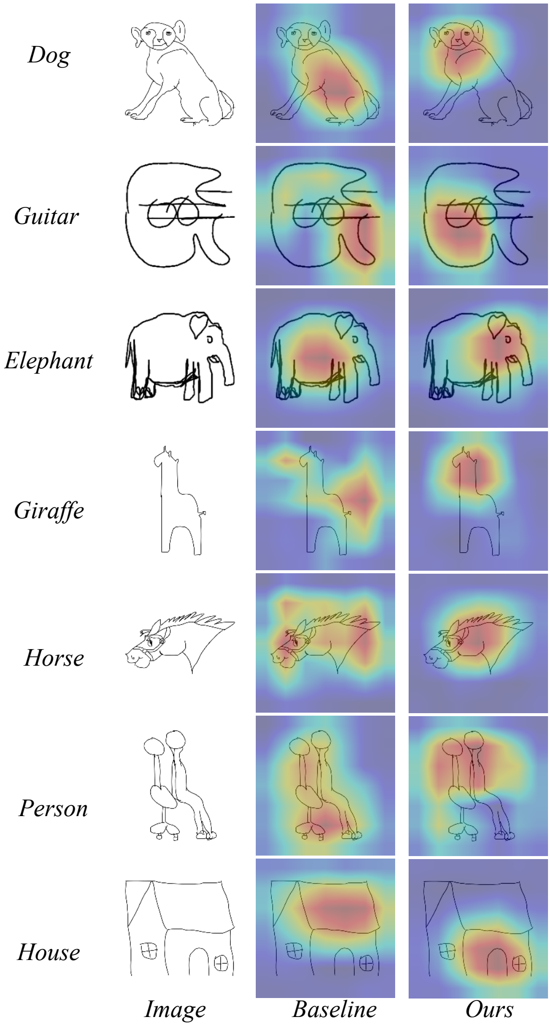

Fig. 7. More result comparison in Sketch domain of PACS dataset.

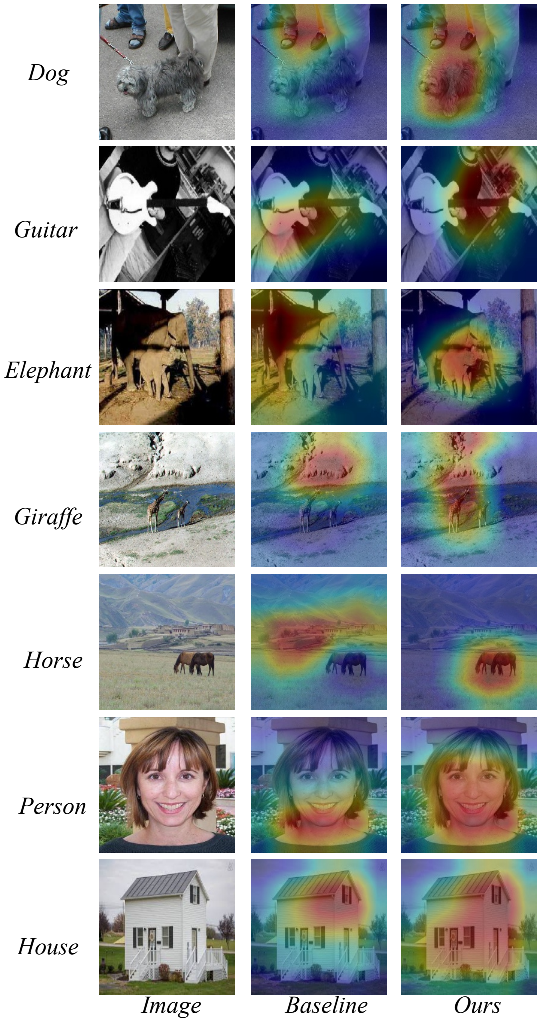

Fig. 8. More result comparison in Photo domain of PACS dataset.

References

1. Carlucci, F.M., D'Innocente, A., Bucci, S., Caputo, B., Tommasi, T.: Domain generalization by solving jigsaw puzzles. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2229–2238 (2019)
2. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.M.: Deeper, broader and artier domain generalization. In: Proceedings of the IEEE international conference on computer vision. pp. 5542–5550 (2017)
3. Maaten, L.v.d., Hinton, G.: Visualizing data using t-SNE. Journal of machine learning research **9**(Nov), 2579–2605 (2008)
4. Venkateswara, H., Eusebio, J., Chakraborty, S., Panchanathan, S.: Deep hashing network for unsupervised domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5018–5027 (2017)
5. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2921–2929 (2016)