Supplementary Material for Unsupervised Domain Generalization by Learning a Bridge Across Domains

Sivan Harary*[1], Eli Schwartz*[1,2], Assaf Arbelle*[1],
Peter Staar[1], Shady Abu-Hussein[1,2], Elad Amrani[1,3], Roei Herzig[1,2],
Amit Alfassy[1,3], Raja Giryes[2], Hilde Kuehne[6,7], Dina Katabi[5],
Kate Saenko[4,7], Rogerio Feris[7], Leonid Karlinsky*[1]

[1]IBM Research, [2]Tel-Aviv University, [3]Technion, [4]Boston University,
[5]MIT, [6]Goethe University, [7]MIT-IBM Watson AI Lab

1. Additional edge map examples

Additional comparison of typical examples of BrAD variants generated by Canny [2], pretrained HED [4], and our 'learned BrAD' are shown in Figures 2 and 3. The images are taken from PACS dataset [3]. Figure 2 presents images from the domains Real and Art, while Figure 3 presents images from Sketch and Cartoon. As can be seen, both 'learned BrAD' and 'HED BrAD' variants discard the background noise, but unlike HED, 'learned BrAD' learns to retain semantic details of shape and texture. These retained details are intuitively highly useful for making the representations, learned using the 'learned BrAD' as the bridge domain Ω, effective for the downstream tasks such as UDG or FUDA.

2. BrAD loss $\mathcal{L}_\Omega$ without pretrained HED

When training BrAD domain mappings Ψ_n we utilize the BrAD loss $\mathcal{L}_\Omega$ (Eq. (4) in the main paper, repeated in Eq. (1) below for convenience) for distilling from an edge-mapping $\mathcal{E}$ forcing the BrAD bridge domain Ω images to be similar to edge maps.

$$\mathcal{L}_\Omega(I_n) = ||\Psi_n(I_n^{a1}) - \mathcal{E}(I_n^{a1})||_2^2 \quad (1)$$

In the main variant of our approach the $\mathcal{E}$ is a HED [4] model pretrained on the BSDS500 dataset [1] and the Ψ_n models are initialized with the same pretrained model. To avoid this use of BSDS500 as additional data, we tested alternative loss functions based on Canny [2] instead of HED [4], while randomly initializing Ψ_n. Since the edges in Canny edge-maps are only one pixel wide we apply a Gaussian blurring before comparing to the current Ψ_n output. In our implementation we used a blur kernel of size 5 with $\sigma = 0.15$. We test both L1 and L2 norms, however, when using L1-norm we first stretch the Ψ_n output to the range [0,1] for stability. Eq. (2) and Eq. (3) present both these variants of the Canny-based $\mathcal{L}_\Omega$ loss functions.

$$\mathcal{L}_\Omega(I_n) = ||\Psi_n(I_n^{a1}) - Blur(\mathcal{E}_{Canny}(I_n^{a1}))||_2^2 \quad (2)$$

$$\mathcal{L}_\Omega(I_n) = ||\mathcal{S}(\Psi_n(I_n^{a1})) - Blur(\mathcal{E}_{Canny}(I_n^{a1}))||_1 \quad (3)$$

where $\mathcal{S}$ is a pixel-wise stretch function to the range [0,1].

Loss	1-shot	3-shots
L2 Hed (equation 1)	**48.64**	58.31
L2 Canny (equation 2)	47.40	**59.30**
L1 Canny (equation 3)	47.77	58.66

Table 1. FUDA accuracy (%) results on DomainNet using different L_Ω BrAD losses. **bold** = best, blue = second best.

Tab. 1 presents the average FUDA accuracy results on DomainNet using the above loss functions. As can be seen the differences in performance are quite small. Fig. 1 presents examples of BrAD images using the different losses. As can be clearly seen, all the learned BrAD variants (Fig. 1d-f) retain semantic details of shape and texture better than the fixed a-priori BrAD variants (Fig. 1b-c).

3. Demo

In in Figs. 4 to 19 we showcase the domain alignment capabilities of the feature representation learned without supervision using our BrAD approach. Each example shows top-5 nearest neighbors of a random query image (from the PACS dataset) searched in the entire set of images of each of the 4 different PACS domains: Photo, Art/Painting, Cartoon, and Sketch. All images are encoded using our self-supervised BrAD model trained on DomainNet dataset.

*Equal contribution

(a) Original image (b) Canny

(c) Hed (d) L2 Hed-based L_Ω (Eq. 1)

(e) L2 Canny-based L_Ω (Eq. 2) (f) L1 Canny-based L_Ω (Eq. 3)

Figure 1. Output images of BrAD mapping functions Ψ_n trained using different loss functions L_Ω also compared to Canny [2] and HED [4] edge maps. Canny produces noisy images with many irrelevant edges, while HED mostly outlines the object and discards important internal texture. All learned Ψ_n retain semantic details of shape and texture better than HED while discarding most of the noise. Please zoom.

Figure 2. Edge images for different choices of Ψ_n. Images are taken from Real and Art domains of PACS dataset [3].

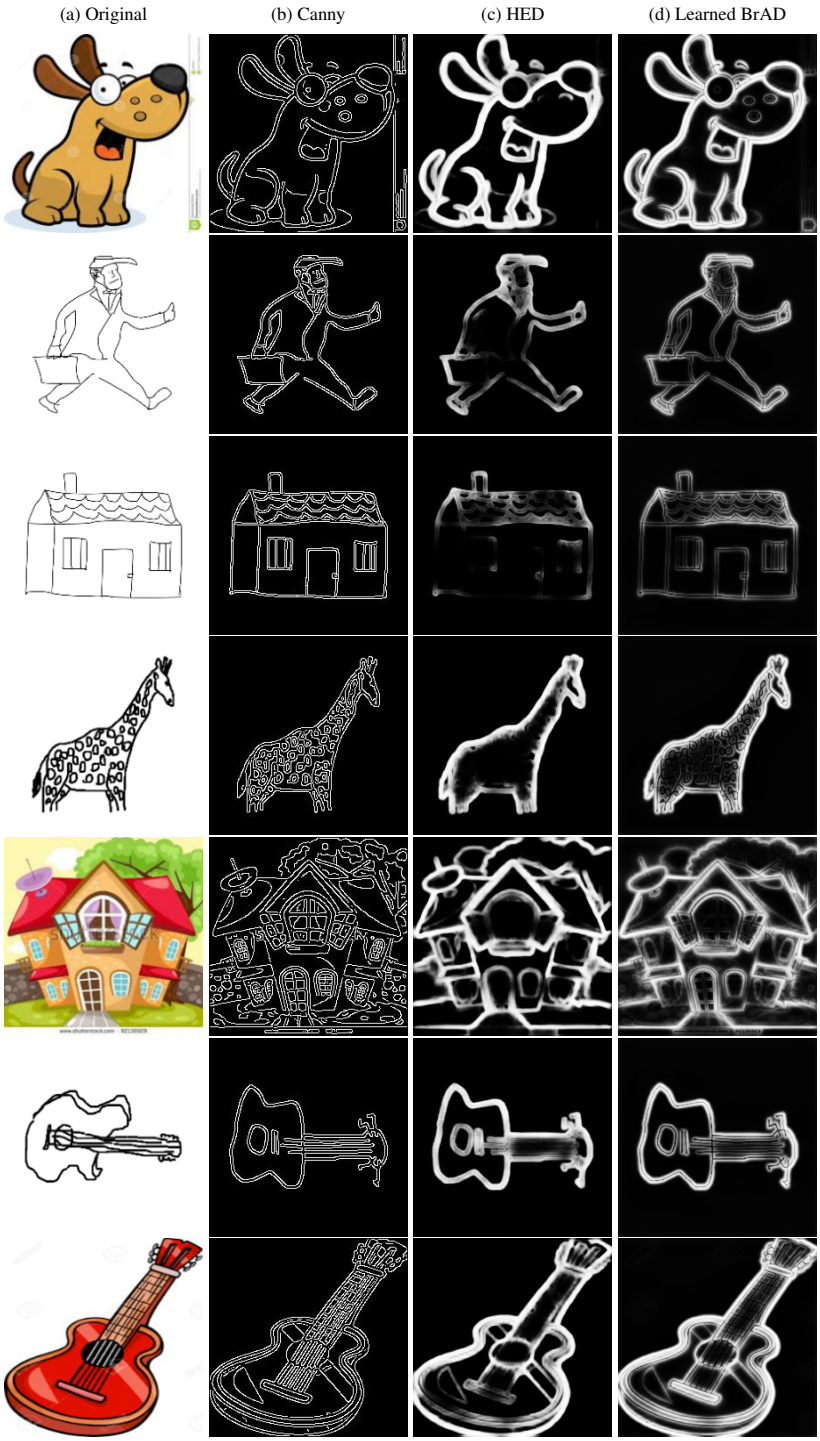

Figure 3. Edge images for different choices of Ψ_n. Images are taken from Cartoon and Sketch domains of PACS dataset [3].

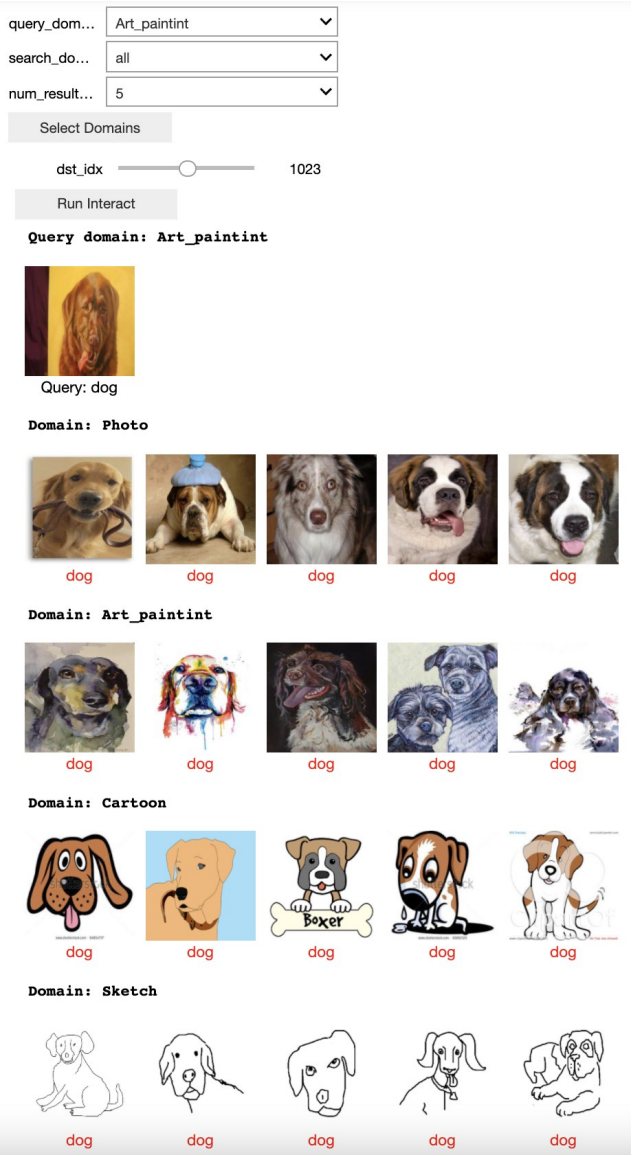

Figure 4. Random query example from our demo. For each query image (from PACS) we show the top 5 image matches among the entire set of images in each of the 4 PACS domains: Photo, Art/Painting, Cartoon, and Sketch. The matching is obtained using our self-supervised BrAD model trained using DomainNet data. The correct class is Dog. The text under each image is the ground truth class of that image in the PACS dataset.

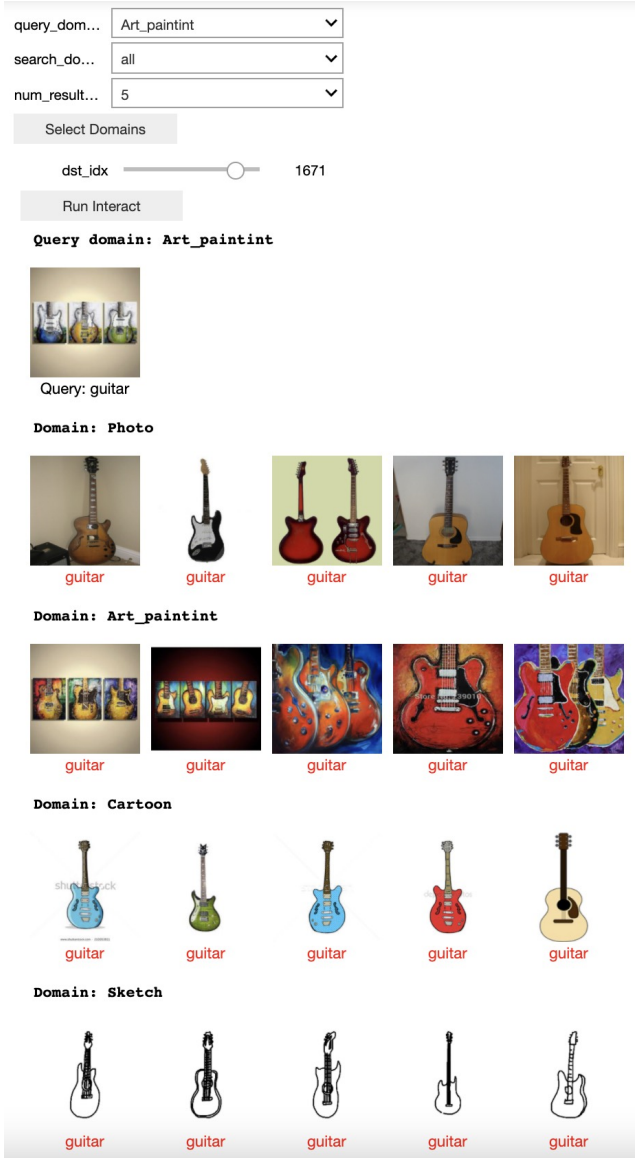

Figure 5. Random query example from our demo. For each query image (from PACS) we show the top 5 image matches among the entire set of images in each of the 4 PACS domains: Photo, Art/Painting, Cartoon, and Sketch. The matching is obtained using our self-supervised BrAD model trained using DomainNet data. The correct class is Guitar. The text under each image is the ground truth class of that image in the PACS dataset.

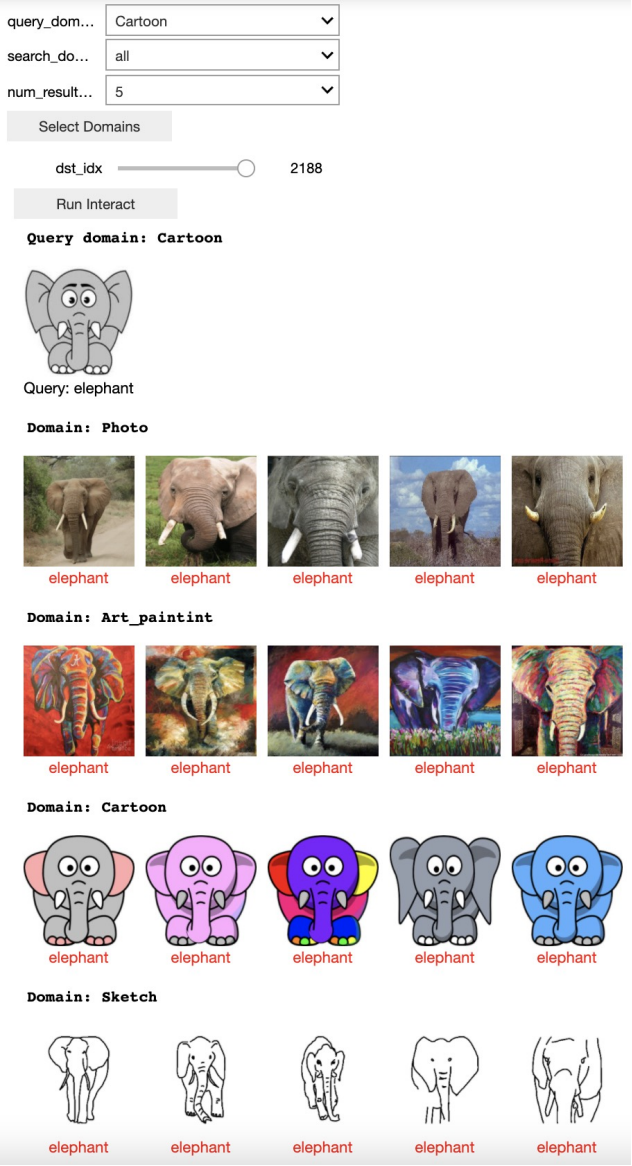

Figure 6. Random query example from our demo. For each query image (from PACS) we show the top 5 image matches among the entire set of images in each of the 4 PACS domains: Photo, Art/Painting, Cartoon, and Sketch. The matching is obtained using our self-supervised BrAD model trained using DomainNet data. The correct class is Elephant. The text under each image is the ground truth class of that image in the PACS dataset.

Figure 7. Random query example from our demo. For each query image (from PACS) we show the top 5 image matches among the entire set of images in each of the 4 PACS domains: Photo, Art/Painting, Cartoon, and Sketch. The matching is obtained using our self-supervised BrAD model trained using DomainNet data. The correct class is Giraffe. The text under each image is the ground truth class of that image in the PACS dataset.

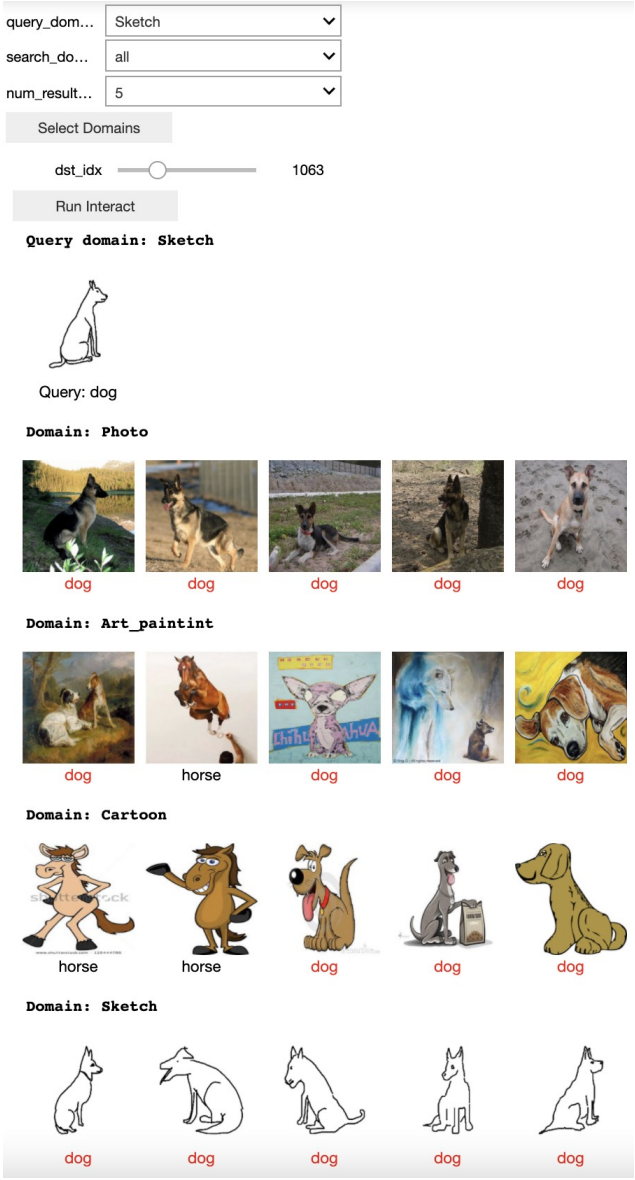

Figure 8. Random query example from our demo. For each query image (from PACS) we show the top 5 image matches among the entire set of images in each of the 4 PACS domains: Photo, Art/Painting, Cartoon, and Sketch. The matching is obtained using our self-supervised BrAD model trained using DomainNet data. The correct class is Dog. The text under each image is the ground truth class of that image in the PACS dataset.

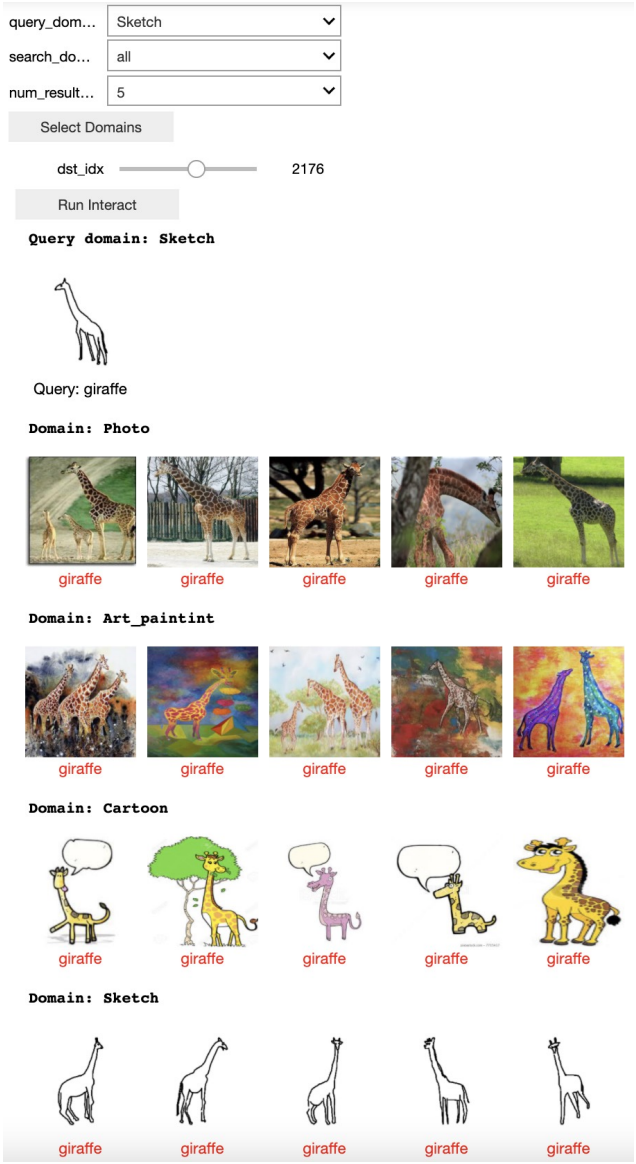

Figure 9. Random query example from our demo. For each query image (from PACS) we show the top 5 image matches among the entire set of images in each of the 4 PACS domains: Photo, Art/Painting, Cartoon, and Sketch. The matching is obtained using our self-supervised BrAD model trained using DomainNet data. The correct class is Giraffe. The text under each image is the ground truth class of that image in the PACS dataset.

Figure 10. Random query example from our demo. For each query image (from PACS) we show the top 5 image matches among the entire set of images in each of the 4 PACS domains: Photo, Art/Painting, Cartoon, and Sketch. The matching is obtained using our self-supervised BrAD model trained using DomainNet data. The correct class is House. The text under each image is the ground truth class of that image in the PACS dataset.

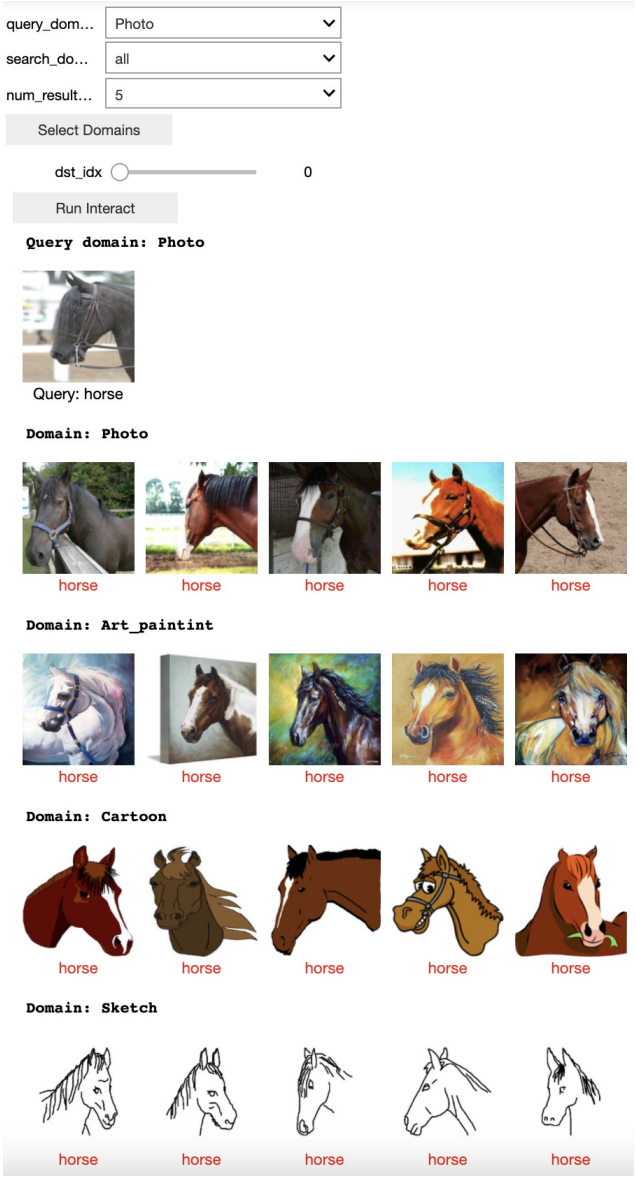

Figure 11. Random query example from our demo. For each query image (from PACS) we show the top 5 image matches among the entire set of images in each of the 4 PACS domains: Photo, Art/Painting, Cartoon, and Sketch. The matching is obtained using our self-supervised BrAD model trained using DomainNet data. The correct class is Horse. The text under each image is the ground truth class of that image in the PACS dataset.

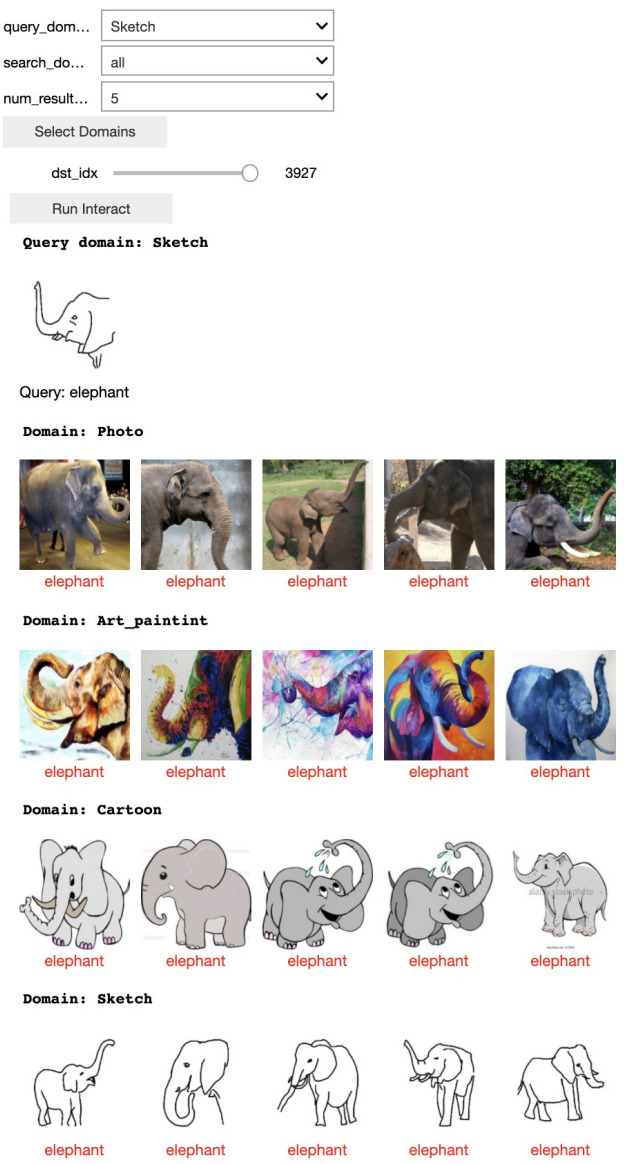

Figure 12. Random query example from our demo. For each query image (from PACS) we show the top 5 image matches among the entire set of images in each of the 4 PACS domains: Photo, Art/Painting, Cartoon, and Sketch. The matching is obtained using our self-supervised BrAD model trained using DomainNet data. The correct class is Elephant. The text under each image is the ground truth class of that image in the PACS dataset.

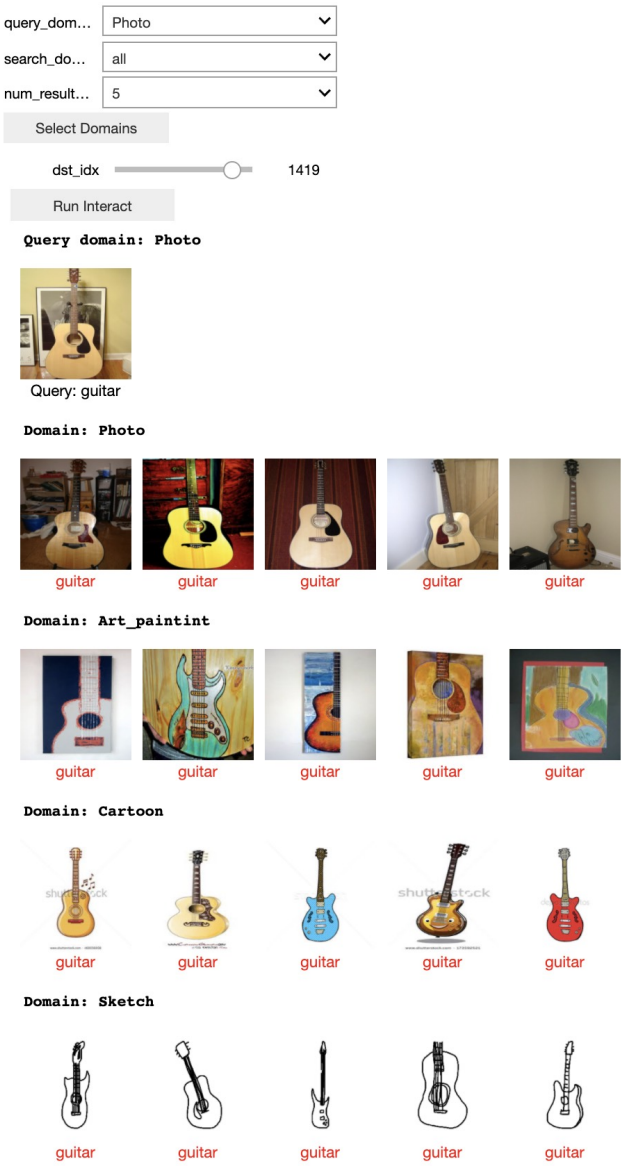

Figure 13. Random query example from our demo. For each query image (from PACS) we show the top 5 image matches among the entire set of images in each of the 4 PACS domains: Photo, Art/Painting, Cartoon, and Sketch. The matching is obtained using our self-supervised BrAD model trained using DomainNet data. The correct class is Guitar. The text under each image is the ground truth class of that image in the PACS dataset.

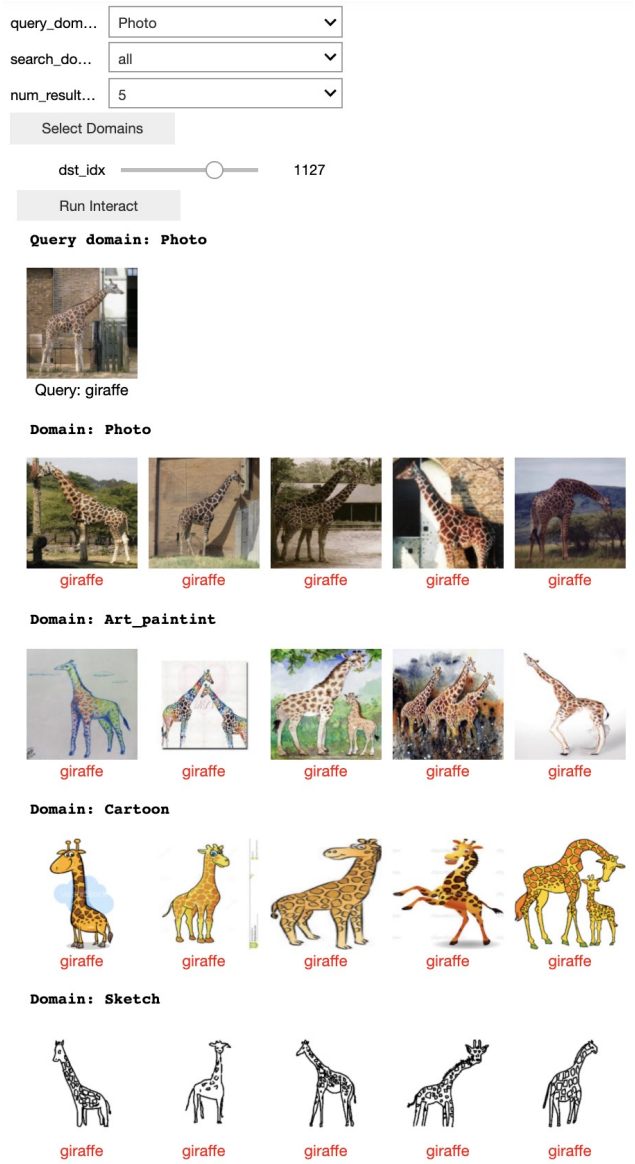

Figure 14. Random query example from our demo. For each query image (from PACS) we show the top 5 image matches among the entire set of images in each of the 4 PACS domains: Photo, Art/Painting, Cartoon, and Sketch. The matching is obtained using our self-supervised BrAD model trained using DomainNet data. The correct class is Giraffe. The text under each image is the ground truth class of that image in the PACS dataset.

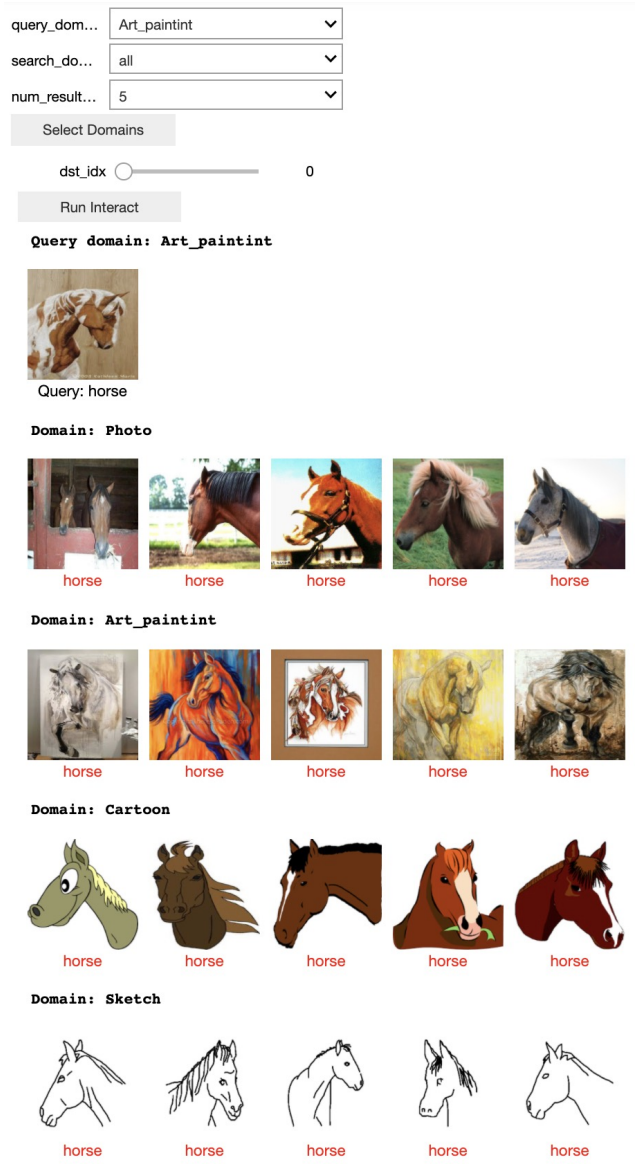

Figure 15. Random query example from our demo. For each query image (from PACS) we show the top 5 image matches among the entire set of images in each of the 4 PACS domains: Photo, Art/Painting, Cartoon, and Sketch. The matching is obtained using our self-supervised BrAD model trained using DomainNet data. The correct class is Horse. The text under each image is the ground truth class of that image in the PACS dataset.

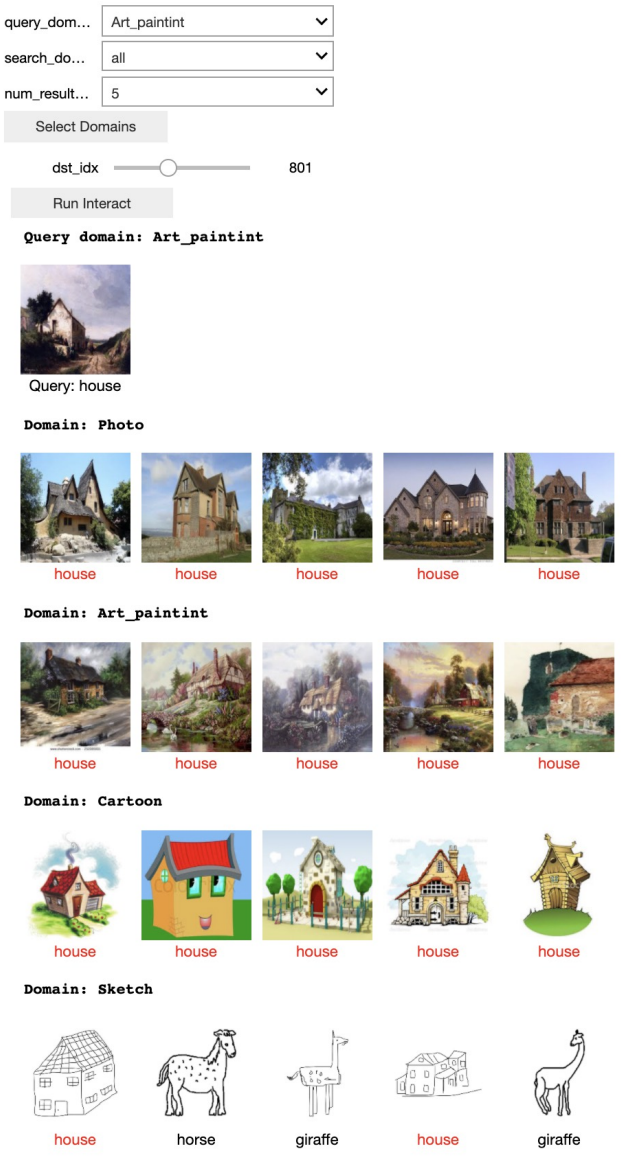

Figure 16. Random query example from our demo. For each query image (from PACS) we show the top 5 image matches among the entire set of images in each of the 4 PACS domains: Photo, Art/Painting, Cartoon, and Sketch. The matching is obtained using our self-supervised BrAD model trained using DomainNet data. The correct class is House. The text under each image is the ground truth class of that image in the PACS dataset.

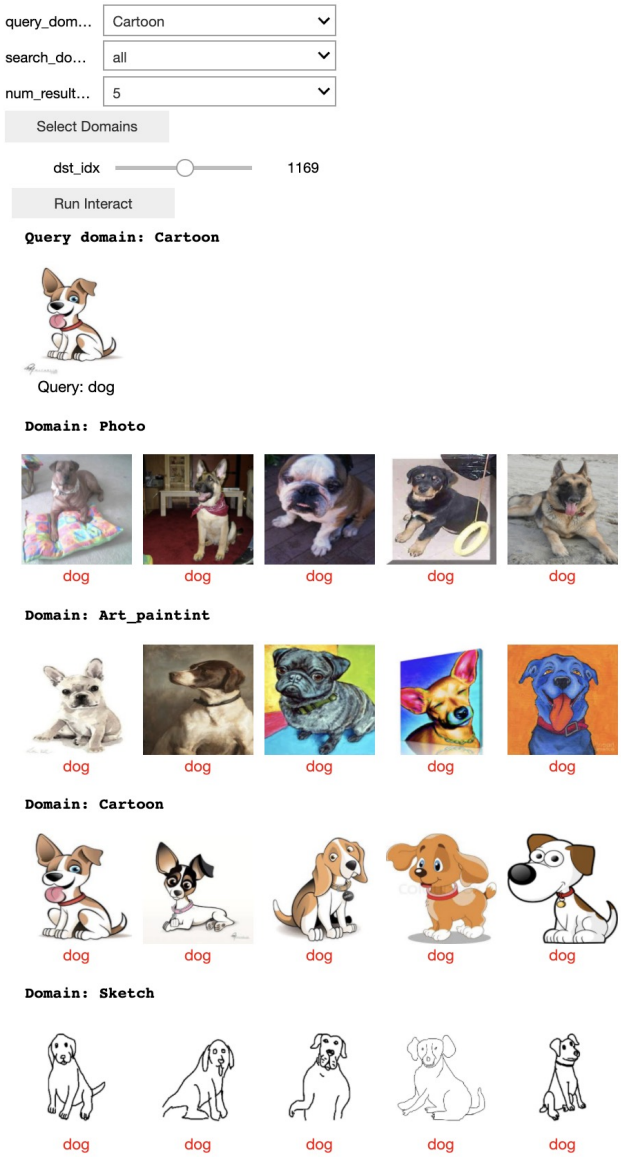

Figure 17. Random query example from our demo. For each query image (from PACS) we show the top 5 image matches among the entire set of images in each of the 4 PACS domains: Photo, Art/Painting, Cartoon, and Sketch. The matching is obtained using our self-supervised BrAD model trained using DomainNet data. The correct class is Dog. The text under each image is the ground truth class of that image in the PACS dataset.

Figure 18. Random query example from our demo. For each query image (from PACS) we show the top 5 image matches among the entire set of images in each of the 4 PACS domains: Photo, Art/Painting, Cartoon, and Sketch. The matching is obtained using our self-supervised BrAD model trained using DomainNet data. The correct class is Horse. The text under each image is the ground truth class of that image in the PACS dataset.

Figure 19. Random query example from our demo. For each query image (from PACS) we show the top 5 image matches among the entire set of images in each of the 4 PACS domains: Photo, Art/Painting, Cartoon, and Sketch. The matching is obtained using our self-supervised BrAD model trained using DomainNet data. The correct class is Guitar. The text under each image is the ground truth class of that image in the PACS dataset.

References

[1] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):898–916, May 2011. 1

[2] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986. 1, 2

[3] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M. Hospedales. Deeper, broader and artier domain generalization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. 1, 3, 4

[4] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015. 1, 2