

000
001
002
003
004
005
006
007
008
009
010
011
012
013
014
015
016
017
018
019
020
021
022
023
024
025
026
027
028
029
030
031
032
033
034
035
036
037
038
039
040
041
042
043
044

000
001
002
003
004
005
006
007
008
009
010
011
012
013
014
015
016
017
018
019
020
021
022
023
024
025
026
027
028
029
030
031
032
033
034
035
036
037
038
039
040
041
042
043
044

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.

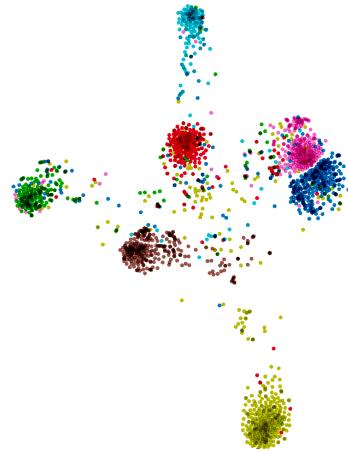
2 t-SNE Embedding

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.

PACS: Art painting domain



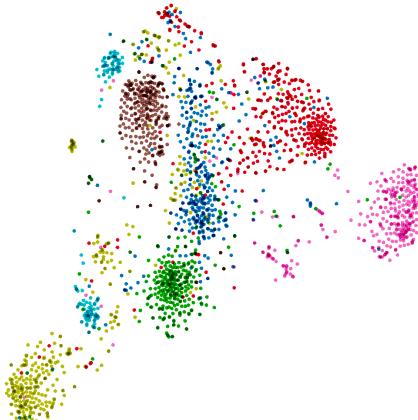
(a) Baseline



(b) Our method

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

PACS: Cartoon domain



(a) Baseline



(b) Our method

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

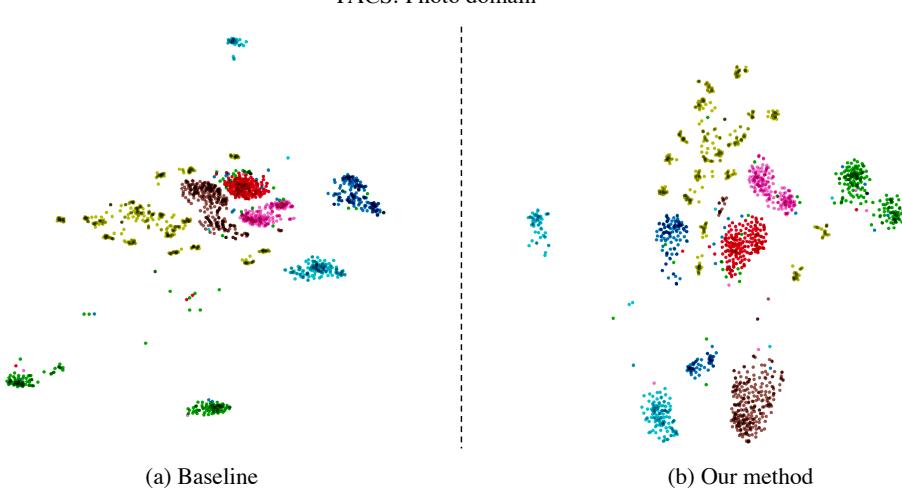


Fig. 3. t-SNE visualization in Photo domain of PACS dataset.

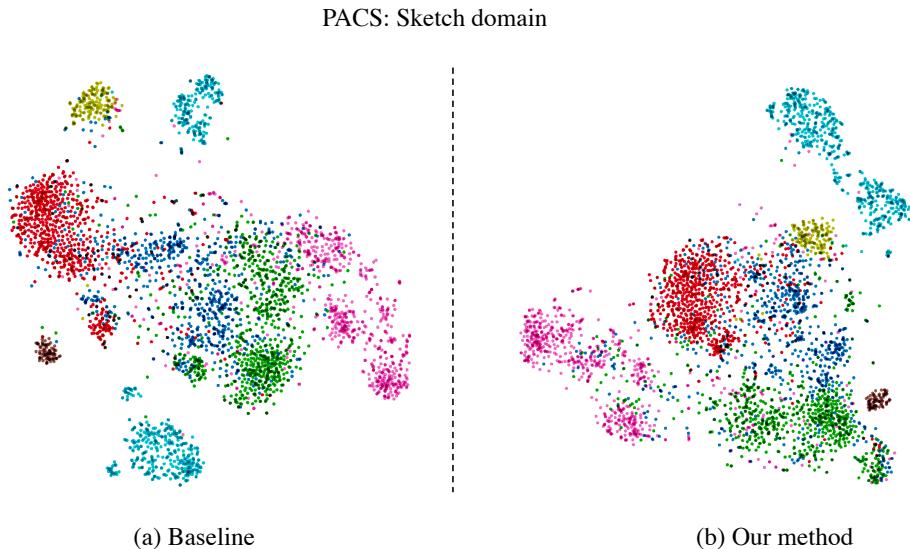


Fig. 4. t-SNE visualization in Sketch domain of PACS dataset.

180 3 Class Activation Maps

181
 182 We compare our method with the baseline model in the localiza-
 183 tions of class-specific regions on the PACS dataset [2] and results
 184 are shown in Figures 5, 6, 7, and 8 below. From the comparison re-
 185 sults, it is shown that our method could recognize object categories
 186 with more meaningful regions of high activation values.
 187

188 4 Results on the Office-Home Dataset

189
 190
 191
Table 1. Domain generalization results on the **Office-Home** dataset with object
 192 recognition accuracy (%) on the **ResNet15** backbone. The top results are highlighted
 193 in **bold**.
 194

195 196 197 Target	198 199 200 ResNet-18		
	201 202 DeepAll	203 204 JiGen [1]	205 206 Ours
207 Art	208 52.15	209 53.04	210 58.43
211 Clipart	212 45.86	213 47.51	214 47.48
215 Product	216 70.86	217 71.47	218 73.03
219 Real-World	220 73.15	221 72.79	222 73.47
223 Average	224 60.51	225 61.20	226 63.10

227
 228 The Office-Home dataset consists of four domains: Art, Clipart,
 229 Product, and Real-World. Each domain contains images from 65
 230 categories, which indicates the challenges of the Office-Home dataset
 231 under the unsupervised setting. The comparison results are shown in
 232 Table 1 above. **DeepAll** is the baseline method that uniformly trains
 233 a network with images from all the domains. **JiGen** [1] adopts an
 234 additional self-supervision task by solving puzzles to constrain the
 235 network. It is observed that our proposed method outperforms JiGen
 236 by 1.9% on average.
 237

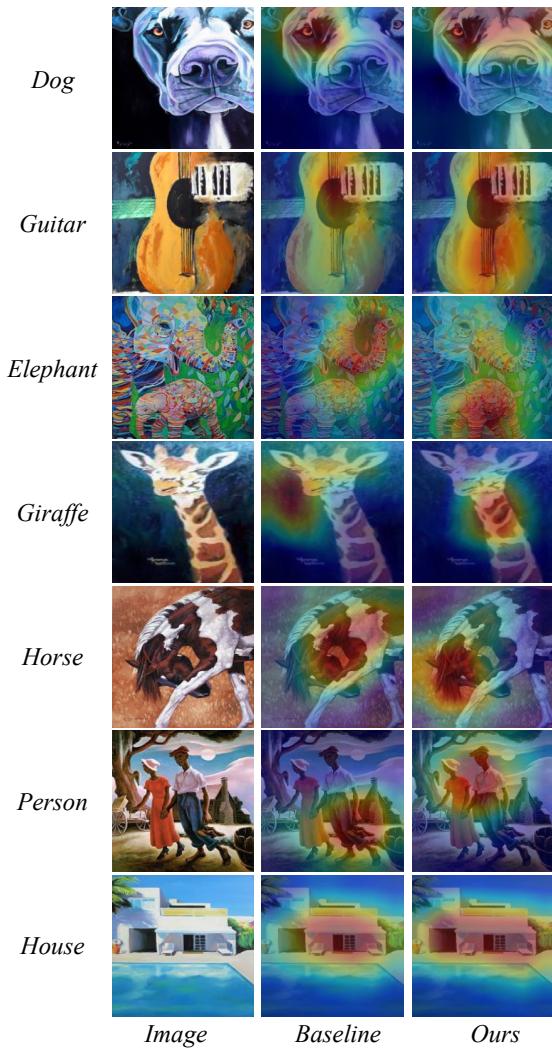


Fig. 5. More result comparison in Art domain of PACS dataset.

270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314

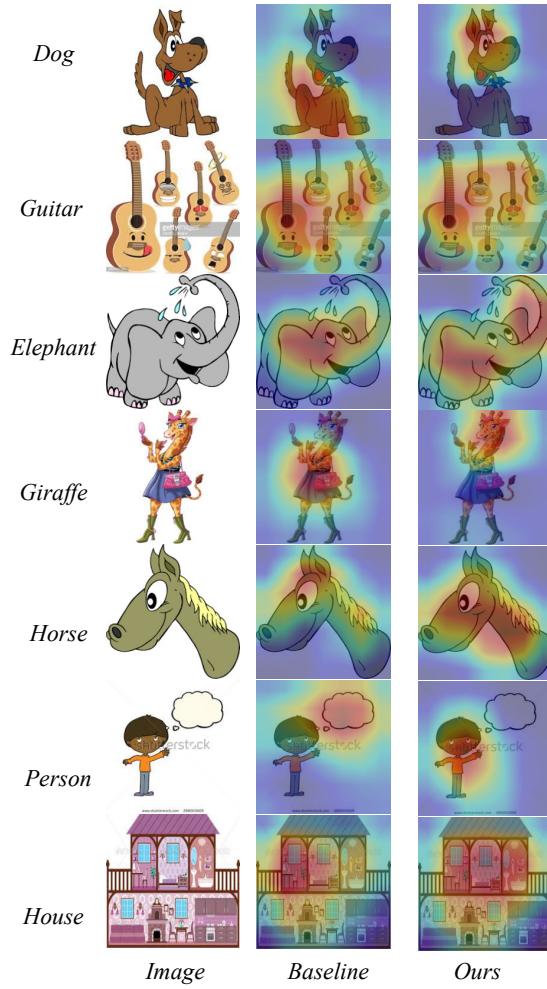


Fig. 6. More result comparison in Cartoon domain of PACS dataset.

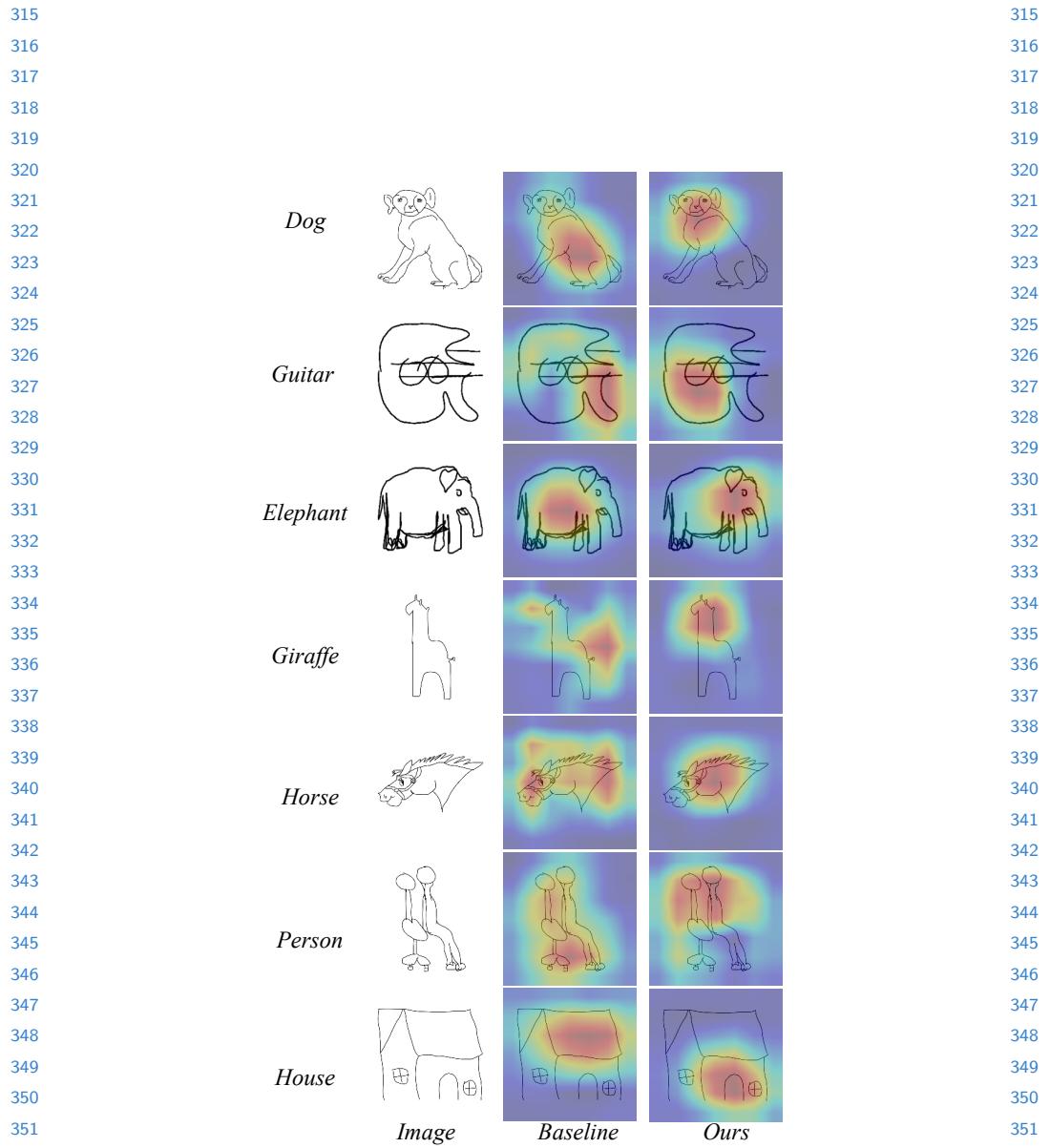


Fig. 7. More result comparison in Sketch domain of PACS dataset.

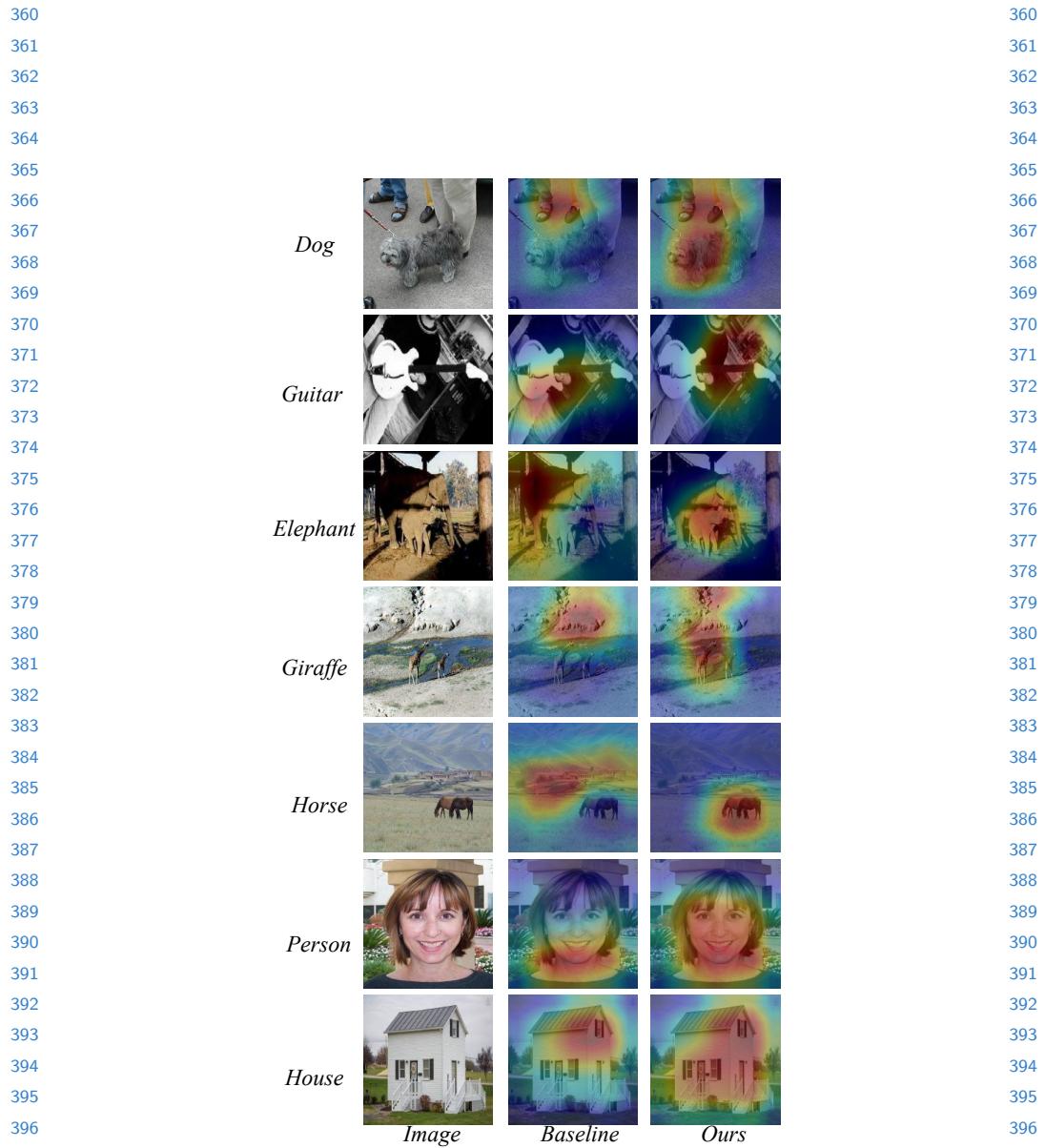


Fig. 8. More result comparison in Photo domain of PACS dataset.

405 References

- 406
- 407 1. Carlucci, F.M., D’Innocente, A., Bucci, S., Caputo, B., Tommasi, T.: Domain gen-
408 eralization by solving jigsaw puzzles. In: Proceedings of the IEEE Conference on
409 Computer Vision and Pattern Recognition. pp. 2229–2238 (2019)
- 410 2. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.M.: Deeper, broader and artier domain
411 generalization. In: Proceedings of the IEEE international conference on computer
412 vision. pp. 5542–5550 (2017)
- 413 3. Maaten, L.v.d., Hinton, G.: Visualizing data using t-SNE. Journal of machine learn-
414 ing research **9**(Nov), 2579–2605 (2008)
- 415 4. Venkateswara, H., Eusebio, J., Chakraborty, S., Panchanathan, S.: Deep hashing
416 network for unsupervised domain adaptation. In: Proceedings of the IEEE Confer-
417 ence on Computer Vision and Pattern Recognition. pp. 5018–5027 (2017)
- 418 5. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features
419 for discriminative localization. In: Proceedings of the IEEE conference on computer
420 vision and pattern recognition. pp. 2921–2929 (2016)
- 421
- 422
- 423
- 424
- 425
- 426
- 427
- 428
- 429
- 430
- 431
- 432
- 433
- 434
- 435
- 436
- 437
- 438
- 439
- 440
- 441
- 442
- 443
- 444
- 445
- 446
- 447
- 448
- 449