

1 **General Reply:**

2 First, we would like to thank all of the reviewers for their detailed and thoughtful reviews. We are glad to see that
3 reviewers generally appreciated the strengths of our paper, especially the significance of the methodological contribution
4 (R1, R3) including its potential impact in production (R1), the clarity of our writing (R2, R3), and the introduction of
5 the ImageNet-Sketch dataset. Additionally, we thank the reviewers for constructive suggestions, in particular requests
6 for additional analysis to support PAR’s effectiveness and modifications to the exposition. Below, we address each
7 reviewer’s comments in turn.

8 **Reply to Reviewer 1:**

9 **“the main argument was a little counter-intuitive”** We agree that your intuition is right: local patches are often
10 predictive (in-sample), and a large body of evidence suggests that ConvNets naturally exploit this property. Our main
11 argument does not refute this. Instead, we argue that these patches, while predictive in-sample, may be less reliable
12 out-of-domain as compared to larger-scale patterns. We will improve the writing to make this point clearer.

13 **“how well this approach would work on models focusing on loss functions across multiple layers?”** This is a great
14 question. While we didn’t address object detection and indeed, adapting our approach to the SSD architecture might
15 require modifications (say, an additional convolutional layer prepended before the standard SSD layers). However, we
16 see no obstacles to integrating our approach with other object detection architectures, such as Faster R-CNN and YOLO.
17 We are grateful for the suggestions and will explore these directions in future work.

18 **“comparison to the related work that used domain knowledge”** Thanks for this suggestion. We are working to
19 implement this and other requested baselines and will add the results to the camera-ready version if accepted.

20 **Reply to Reviewer 2:**

21 **“missing recent baselines”** Thanks for pointing out these references to recent domain generalization papers. We will
22 add these comparisons to the camera-ready version if accepted.

23 **“how should we select which PAR to use?”** This is a great question. None of the variants of PAR outperform the
24 vanilla PAR consistently. **However, the vanilla PAR outperforms nearly all other baselines in the vast majority of**
25 **our experiments.** Our draft includes results for these variants for the sake of thoroughness.

26 **“clarifications about Section 4.2”** In some experiments, some of our baselines access domain labels (and thus treat
27 the problem as domain generalization). However, our model is blind to domain labels. Moreover, unlike unsupervised
28 domain adaptation approaches, our model does not incorporate the unlabeled target data into training. **Surprisingly,**
29 **our methods often perform better despite the unfair comparison.** Our experiments employ ResNet-50. We will
30 revised the draft to make these facts clearer and release our code publicly for full transparency.

31 **“validation of the patch-wise classifier”** Per your suggestion, we ran experiments to validate the patch-wise classifier.
32 Without (PAR) regularization, the patch-wise classifier can achieve roughly 20% accuracy on in-domain test data
33 (Figure 1(a), orange, before epoch 250). It achieves 12% accuracy on texture-altered out-of-domain data (Figure 1(a),
34 magenta and green, before epoch 250) and 5% accuracy color-altered out-of-domain data (Figure 1(a), maroon, before
35 epoch 250). With PAR, the patch-wise classifier achieves 15% in-domain prediction accuracy (5% drop) (Figure 1(a),
36 orange, after epoch 250), and 10% on texture-altered out-of-domain data (Figure 1(a), magenta and green, after epoch
37 250) and 8% on color-altered out-of-domain data (Figure 1(a), maroon, after epoch 250).

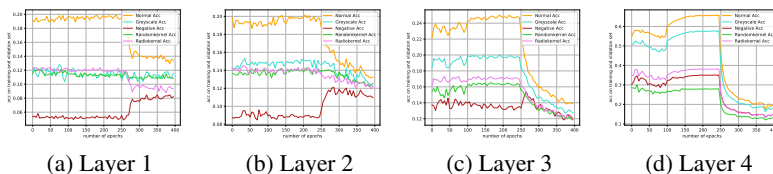


Figure 1: Prediction accuracy of patch-wise classifier. The regularization is introduced at Epoch 250.

38 **Reply to Reviewer 3:**

39 Thanks for recognizing the fundamental nature and significance of our contribution.

40 **“more related theoretical analysis if possible”** We share your enthusiasm for these foundational theoretical results
41 (Ben-David 2010), which have influenced theoretical inquiry into domain adaptation. However, we note that papers
42 claiming theoretical support for deep domain adaptation techniques have misinterpreted the theory. Two recent papers
43 [1,2] independently identified these flaws and construct simple counter-examples (e.g., when label distributions shift)
44 where these techniques are guaranteed to fail (if the optimization succeeds). We agree that theoretical support would be
45 a great asset, noting only that comparable methods, lack such theoretical support.

46 1. Johansson et al. Support and Invertibility in Domain-Invariant Representations (AISTATS 2019)

47 2. Wu et al. Domain Adaptation with Asymmetrically-Relaxed Distribution Alignment (ICML 2019)