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

Reply to Reviewer 1:
“the main argument was a little counter-intuitive” We agree that your intuition is right: local patches are often
predictive (in-sample), and a large body of evidence suggests that ConvNets naturally exploit this property. Our main
argument does not refute this. Instead, we argue that these patches, while predictive in-sample, may be less reliable
out-of-domain as compared to larger-scale patterns. We will improve the writing to make this point clearer.

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

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

Reply to Reviewer 2:
“missing recent baselines” Thanks for pointing out our these references to recent domain generalization papers. We will
add these comparisons to the camera-ready version if accepted.

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

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

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

Reply to Reviewer 3:
Thanks for recognizing the fundamental nature and significance of our contribution.

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

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