A. Appendix: CompCars Dataset Augmentation

We further augment the CompCars dataset by creating rotating hue and minor perspective jitter. In prior unpublished experiments, these changes seemed to improve accuracy. We rotate the hue by simply swapping color channels. We do this because we hypothesized that car models have varying color, but they seem to have color styles. For instance, family sedans seem to have conservative low saturated colors while sports cars tend to have hot intense and highly saturated colors. We obtain perspective jitter by randomly perturbing three Euler angles by +/- 0.00286 degrees, then we create a perspective transformation matrix from it. We create 24 augmentations per image, from which 14 have random perspective jitter and 12 have hue rotation. Six of the hue rotated images also have perspective jitter. Four images are just a repeat of the un-augmented original. An example of these alterations can be seen in figure 6. We create these permutations before training since perspective transformation is modestly expensive, but still can be a bottleneck.

B. Appendix: Further Ablation Details

B.1. Yoked Jitter Gets Better with Extra Patches

As mentioned, one of the goals of using hybrid patches was to reduce the ability of the network to learn trivial pattern completion between adjacent patches. As a test, we tried the usage of extra patch configurations (EPC) with yoked jitter and with random jitter. The results can be seen in Table 6. By getting a larger gain for yoked jitter, there is some evidence that EPC may have the effect of reducing trivial pattern completions.

Method	CUB	CCars	Mean	Improvement
CB	64.29	80.80	72.55	–
CB + YJ	65.17	80.95	73.06	0.51
CB + TP + EPC	65.21	80.17	72.69	–
CB + TP + EPC + YJ	67.07	80.50	73.79	1.10

Table 6. We get a general improvement from using a yoked jitter over a random jitter. When we then include the extra patch configurations, the improvement grows. The hybrid patches intrinsically may prevent low-level trivial boundary completion since they have mixed scales.

Method	CUB	CCars	Mean
CB + YJ + TP + EPC + UBT + RA	68.01	82.07	75.04
... + RWC 180 without classification	68.26	81.82	75.04
... + RWC 180 with classification	68.89	84.23	76.56

Table 7. By just rotating the patches but not classifying them, we obtain almost no gain. It appears critical that rotations should have their own class. Note we are using the full toolset except for RRM or WV.

Method	CUB	CCars	Mean	Impr.
YJ	65.04	80.21	72.62	–
... + CD	62.36	79.70	71.03	–
... + CB	65.17	80.95	73.06	0.44
YJ + TP + EPC + UBT + RA + RWC	68.23	83.70	75.96	–
... + CD	65.73	82.94	74.33	–
... + CB	68.42	83.58	76.00	0.04

Table 8. Here we compare color dropping (CD) and chroma blurring (CB) to no color processing. CUB birds is a pathological case for color dropping since it is very dependent on color patterns for classification. However, and somewhat perplexingly, color dropping does not appear to help CompCars either. Chroma blurring ceases to help once we add in the full set of tools (not including RRM or WV). We suspect this is because rotation with classification at least partially mitigates the effects of chromatic aberration.

B.2. Do Rotations Need Classification?

We tested to see if just rotating a patch without classification for that rotation was sufficient to improve performance. Table 7 shows that if we just rotate the patch, there is almost no difference than without rotation. The classification component might help to sharpen the features of objects by forcing the network to recognize the rotated objects uniquely. Also, as we have discussed, classification may help to mitigate chromatic aberration.

B.3. How much does Chroma Blurring Help?

As figure 7 shows, chroma burring definitely seems to remove any chromatic aberration effects while preserving at least some color feature processing. Looking at Table 8, we do get a moderate boost in classification by using chroma blurring compared with no color processing. However, improvement in classification from chroma blurring subsides when all tools are used. We believe that rotation with classification is probably responsible for this.

B.4. The Benefit of Adding Different Kinds of Patches

Extra patch configurations definitely seem to help, but their interaction with each other and the other tools is not deterministic. Table 9 parses out the contribution of the two types of new configurations we use.

B.5. Two v. Three Apertures

We chose to only apply the patch aperture to two patches and not all three in a set. The idea was to inhibit the highest levels of the network and instead focus learning on mid-levels. If we applied the aperture to all three patches, we reasoned that we would *always* inhibit the higher levels when we only want to inhibit them *some of the time*. As table 10 shows, two apertures are definitely better than three.

Figure 6. These show all 24 augmentations for a single image in CompCars. These variations are applied to all training images.

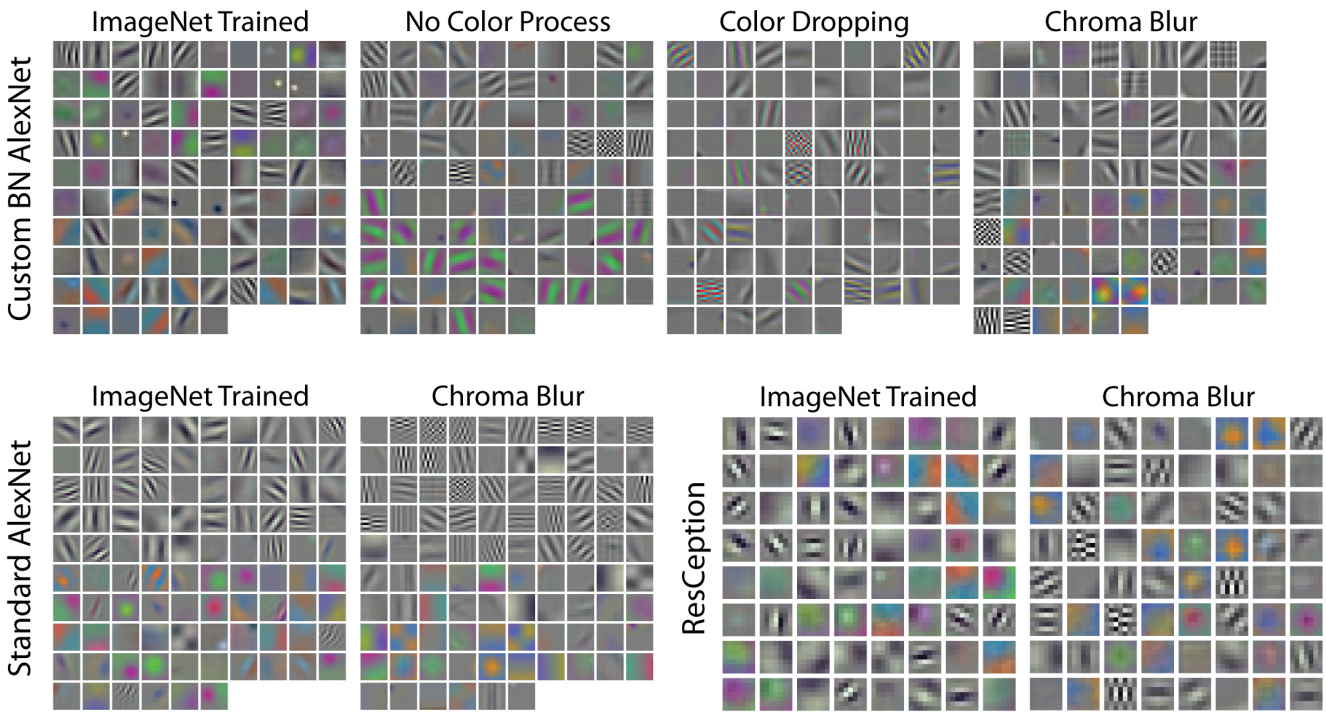

Figure 7. These are the first layer filters from several networks which have either been supervise trained on ImageNet or self-supervise trained. All self-supervised networks are using the full set of methods (but not RRM or WV). Even though rotation with classification may mitigate chromatic aberration, the network still forms filters sensitive to it. Thus, we believe it is only a partial solution. The chroma blur networks are all free of chromatic aberration effects and show formation of healthy color filters (especially when compared to color dropping). As an observation, we can zoom to an effective size of 171x171 during self-supervised training; as a result, we see the presence of finer wavelets compared with ImageNet training.

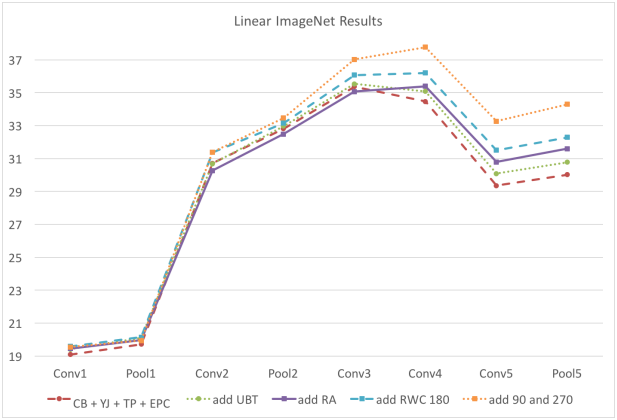

Figure 8. These are linear results for CaffeNet/AlextNet on ImageNet. We can see that when we add Random Aperture (RA), the results seem to switch over to improving layers four and five at the expense of layers one, two and three. Note that we did not use padding during self-supervised training for this experiment.

Method	**CUB**	**CCars**	**Mean**	**Impr.**
CB + YJ + TP	65.19	81.54	73.36	–
... + EPC (2x2 Patches)	66.80	81.80	74.30	0.93
... + EPC (Hybrid Patches)	67.32	81.69	74.50	1.14
... + EPC (2x2 *and* Hybrid Patches)	67.07	80.50	73.79	0.43
CB + YJ + TP + UBT + RA + RWC	68.63	82.87	75.75	–
... + EPC (2x2 Patches)	68.29	83.67	75.98	0.23
... + EPC (Hybrid Patches)	67.09	82.58	74.83	-0.92
... + EPC (2x2 *and* Hybrid Patches)	68.89	84.23	76.56	0.81

Table 9. Here we show the effect of the two types of extra patch configurations on their own. Improvement is not straight forward from the addition of the two types. Using both is always better than using just the 3x3 patches. However, when using the full tool set (not including RRM or WV), the hybrid patches by themselves are actually worse. Our hypothesis is that the 2x2 patches help with rotation classification (RWC) since they always include the top and bottom of the image. These are good locations for cues an image is upside-down (sky v. ground). Without that help, the hybrid patches somehow inhibit performance. It is not entirely clear why.

We ran some further testing to see if applying the aperture has the effect of improving middle layers. Figure 8 shows that it has a boost on layers four and five which are middle layers in AlexNet. Interestingly, it has somewhat degrading effects on layers one and two. This is a kind of behavior we would expect if mid-layers are being biased for.

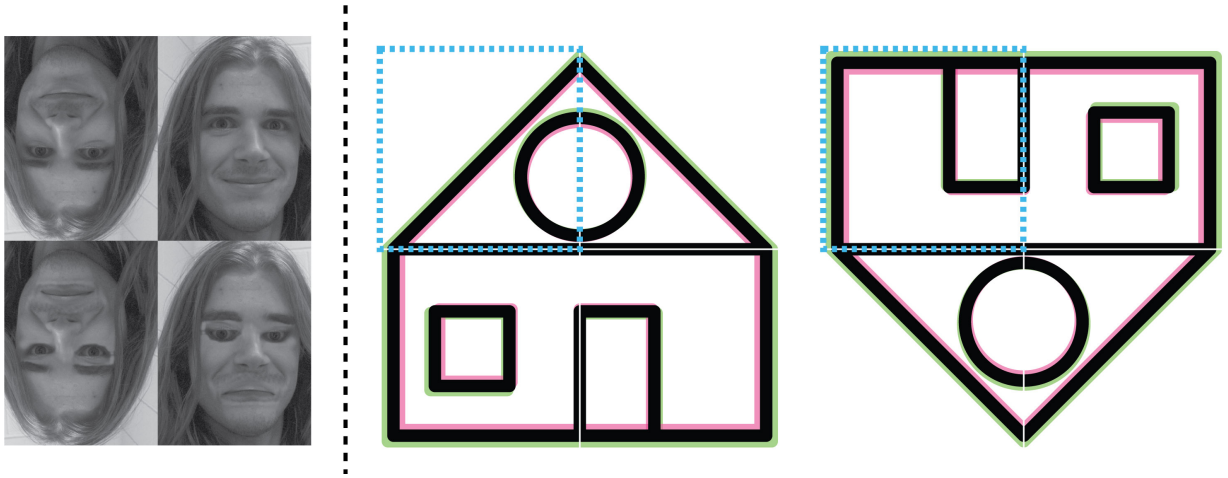

Figure 9. This is figure 4 enlarged. On the left is an example of the famous Thatcher illusion [41, 8]. It demonstrates conditional sensitivity to upside-down features in an image against the background. We used this mostly as inspiration. On the left house image [42], the network can tell that the blue bordered area comes from the upper left corner based on chromatic aberration alone. However, on the right image, rotation with classification makes it tell us if the patch is inverted and comes from the lower right corner. If it uses chromatic aberration as the only cue, it would be wrong 50% of the time.

Figure 10. These are examples of birds in the CUB birds dataset. Each one is a different species. They are a Bewick Wren, Carolina Wren, Anna Hummingbird, Ruby Throated Hummingbird, Vesper Sparrow, Henslow Sparrow, Tree Swallow and a Bank Swallow.

Method	CUB	CCars	Mean
With three apertures	67.76	83.22	75.49
With two apertures	69.04	83.46	76.25

Table 10. If we aperture all three patches in a set, we see a noticeable drop particularly in CUB Birds. We did not test the aperture of one patch.

B.6. RRM is a Tiny Bit Better

Randomization of rescaling methods (RRM) yielded slightly better results on the ImageNet linear and VOC tests. It is roughly even on the CSAIL places linear test. Earlier experiments showed a stronger pattern of gain, but it is now somewhat unclear how much it helps. However, it doesn't seem to hurt.