# A Appendix

## A.1 Visual Explanation of Invariance, Equivariance and Steerability

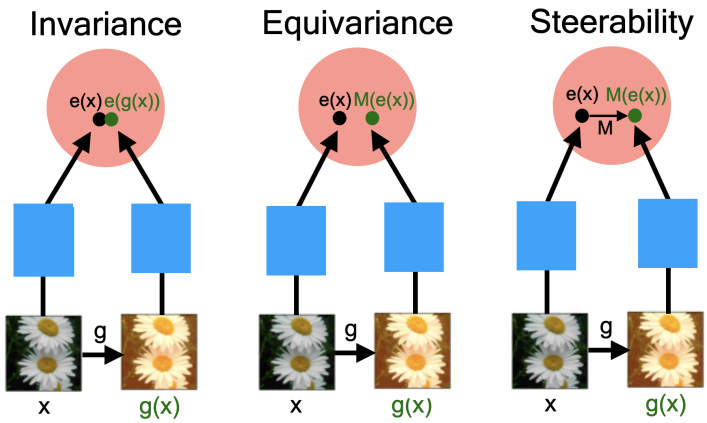


Figure 7: The concepts of invariance, equivariance and steerability of embeddings. Blue boxes represent the (shared) encoder that takes the input $\mathbf{x}$ to the embedding $e(\mathbf{x})$. $g(\mathbf{x})$ represents a transformation of $\mathbf{x}$ in input space; and $M(e(\mathbf{x}))$ is a mapping in embedding space. Equivariance is a necessary but not sufficient condition for steerability.

## A.2 Training Details

Both the invariant and equivariant models use a ResNet-50 encoder trained for 250 epochs with a batch size of 4096, on 224x224 sized ImageNet images with AutoAugment (Cubuk et al., 2018) applied on the input to the cross-entropy loss. The remaining optimization details are as follows: SGD optimizer with 0.9 nesterov momentum, 0.1 learning rate with cosine decay warmed up over 12 epochs, $\alpha$=0.1 and $\beta$=0.1.

## A.3 Robustness

In Table 4, we show the accuracy of both the models on ImageNet-C dataset (Hendrycks & Dietterich, 2019), on all 15 corruptions and 3 different severity levels. We can see that the equivariant model outperforms the invariant across the board, with better accuracy on all but 5 out of 45 data points.

## A.4 Out-of-Distribution Detection

Test-time Augmentation Details: Here we give details of how we perform test time augmentation. We use $M_{geo}$ / $M_{photo}$ to generate multiple augmentations in latent space for a given input image. We compute the geometric mean across the set of logits generated in this manner (for a given number of augmentations), and then use this average logit to compute softmax probabilities. The maximum softmax value is the confidence for this sample. We use these confidences across a set of ImageNet and ImageNet-C samples and probability threshold values to compute a PR curve, and measure the AUC of this PR curve. We repeat this for upto 60 augmentations, and plot the AUC values across the number of augmentations. In Figure 8, we provide more examples of OOD detection for all 15 corruptions from ImageNet-C (Hendrycks & Dietterich, 2019), with severity level 3. In 12 of the 15 cases, the equivariant latent outperforms invariant latent space in AUC on both photometric and geometric augmentations. We display similar plots on different severity levels with the geomentric augmentation in Figures 9, 10 and 11. Adding more number of augmentations may help to further improve performance on the equivariant model.

| | Equivariant | | | Invariant | | |
|---|---|---|---|---|---|---|
| Corruption Severity | 1 | 3 | 5 | 1 | 3 | 5 |
| Zoom Blur | **62.61** | **39.88** | **8.44** | 61.62 | 34.57 | 4.50 |
| Gaussian Noise | **62.15** | **38.95** | **11.11** | 61.27 | 33.85 | 6.84 |
| JPEG Compression | 50.23 | **34.93** | **7.95** | **50.45** | 31.38 | 4.15 |
| Fog | **55.42** | **31.07** | **11.59** | 54.93 | 30.17 | 10.78 |
| Shot Noise | **57.10** | **13.46** | **5.78** | 56.73 | 12.91 | 5.25 |
| Impulse Noise | **63.58** | **27.92** | **7.25** | 61.82 | 24.00 | 5.65 |
| Defocus Blur | **49.43** | **32.80** | **19.80** | 48.88 | 30.59 | 18.86 |
| Glass Blur | **54.77** | **37.03** | **23.42** | 54.36 | 35.89 | 21.36 |
| Motion Blur | **60.66** | **34.55** | **26.53** | 60.12 | 32.66 | 24.05 |
| Snow | 67.60 | 59.41 | 46.28 | **68.07** | **60.63** | **48.00** |
| Frost | **73.62** | **70.06** | **62.73** | 73.36 | 69.89 | 62.21 |
| Brightness | **69.15** | **58.49** | **11.02** | 68.58 | 55.17 | 7.68 |
| Contrast | **68.41** | **51.20** | **9.87** | 68.01 | 48.11 | 8.14 |
| Elastic Transform | **68.30** | **59.93** | **46.66** | 67.77 | 59.84 | 45.80 |
| Pixelate | 62.42 | **57.54** | **43.91** | **62.87** | 57.44 | 41.07 |

Table 4: Accuracy of models on all the corruptions from the ImageNet-C with multiple severities.

### A.5 Nearest Neighbor Retrieval

In Figures 12, 13, 14, 15, we show more examples of image retrieval with color, crop/zoom, composed-color-zoom, and brightness queries respectively across different classes. Qualitatively, this displays the range of transformations and their parameters that our steerable equivariance respresentations generalize to.

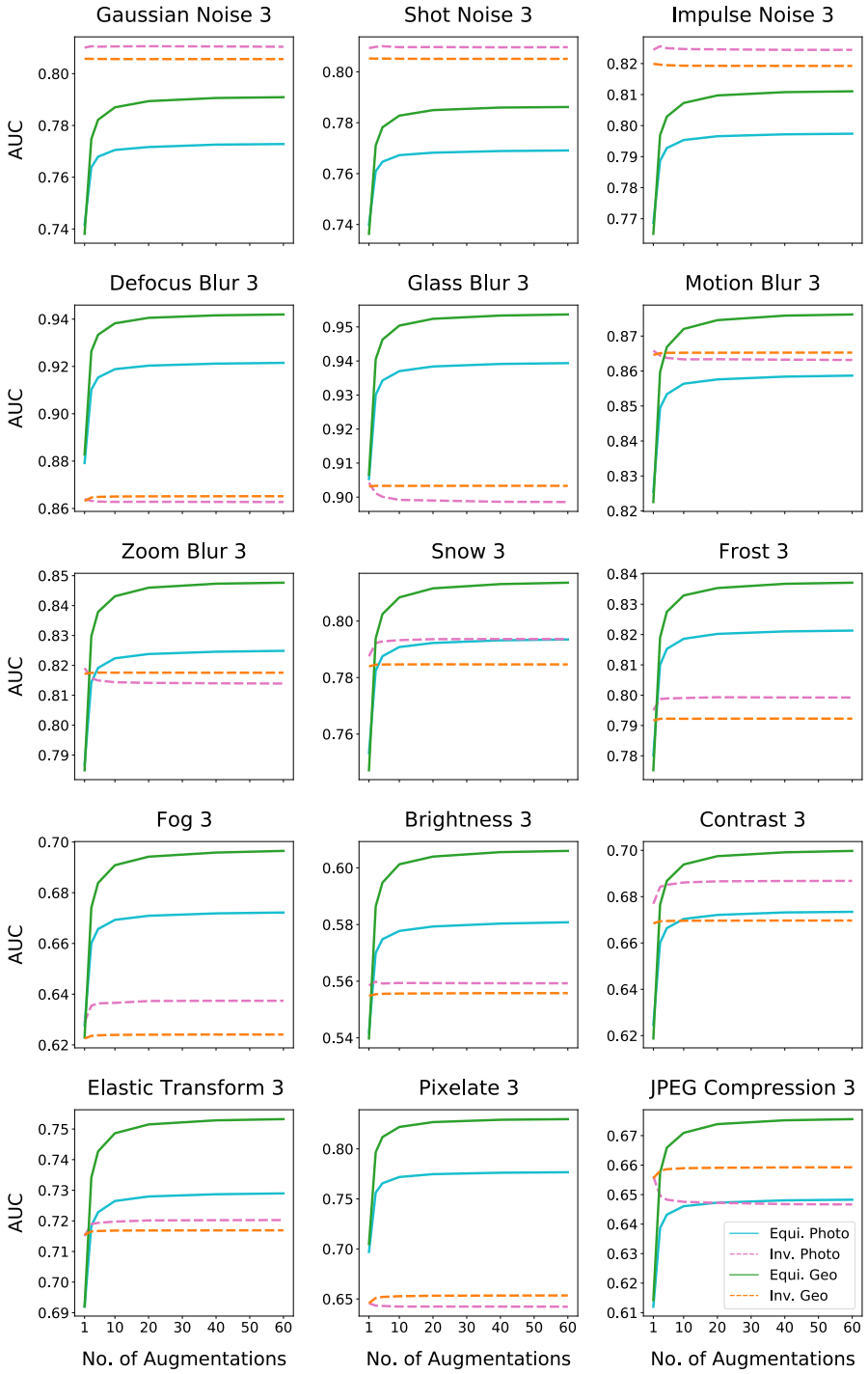


Figure 8: OOD Detection for ImageNet-C when both photometric and geometric augmentations are applied. We see that both augmentations lead to improved OOD performance.

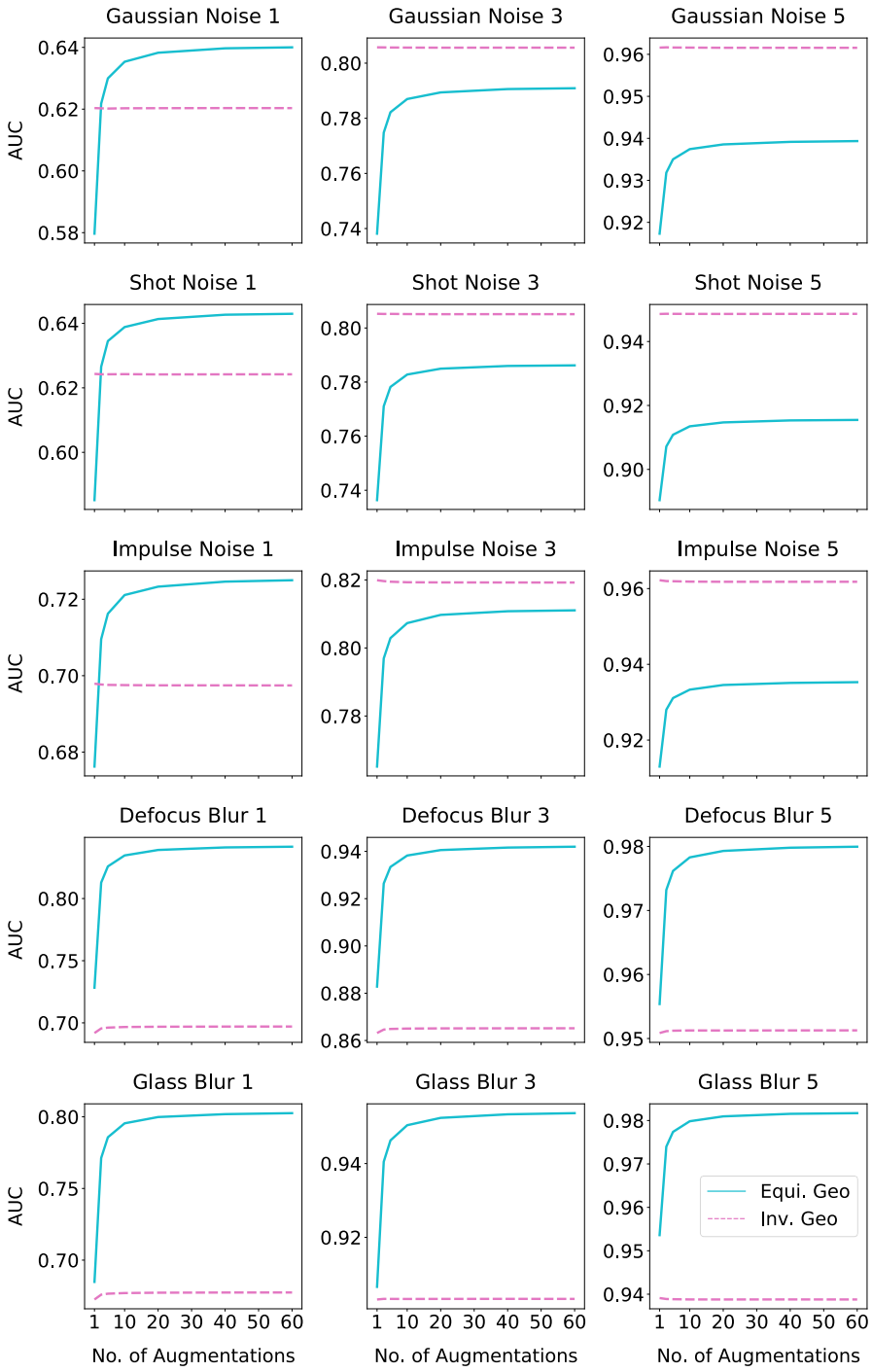


Figure 9: OOD Detection for ImageNet-C with geometric augmentations and multiple severity levels.

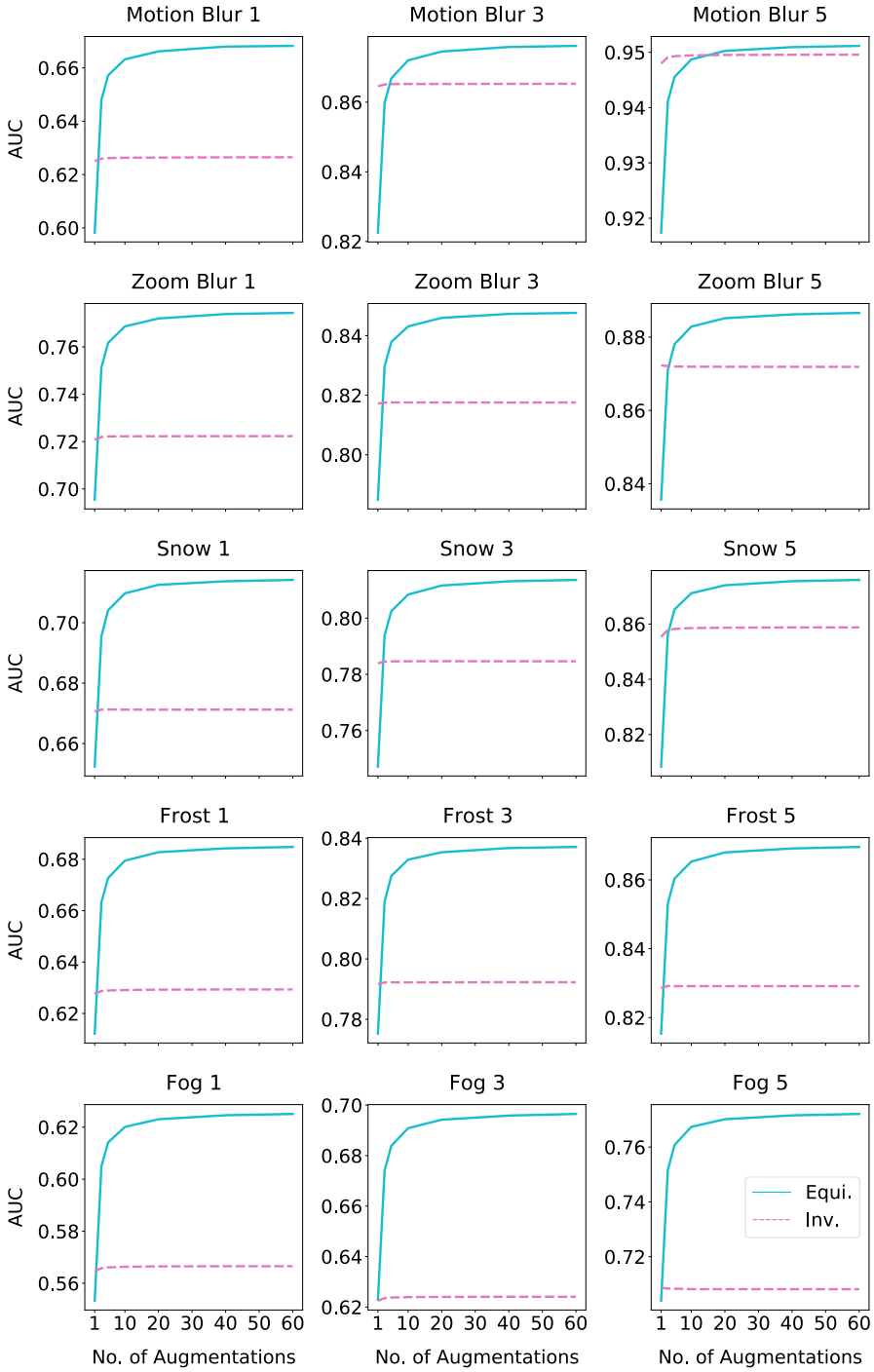


Figure 10: OOD Detection for ImageNet-C with geometric augmentations and multiple severity levels.

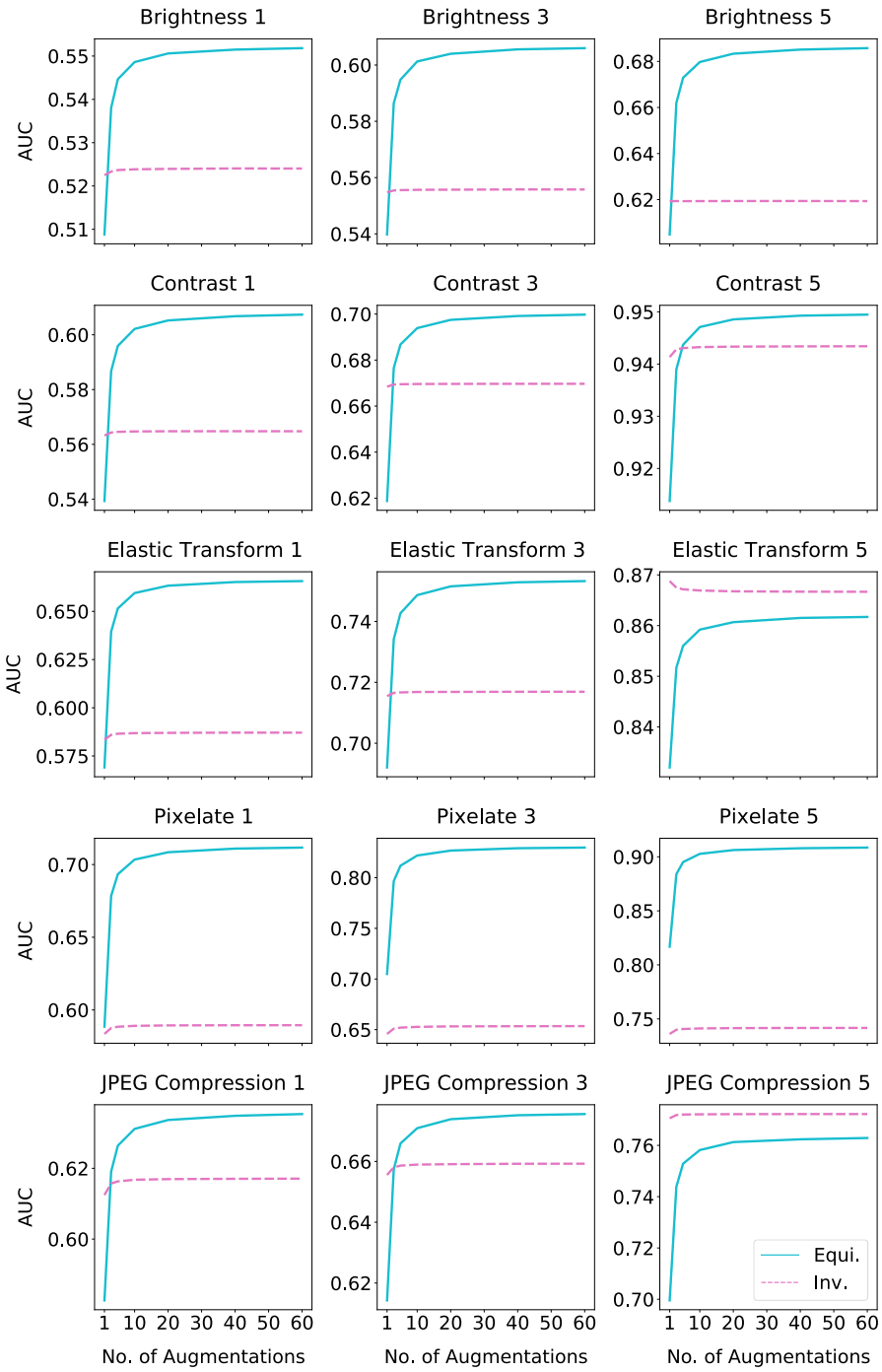


Figure 11: OOD Detection for ImageNet-C with geometric augmentations and multiple severity levels.


Figure 12: Image Color Retrieval Examples.

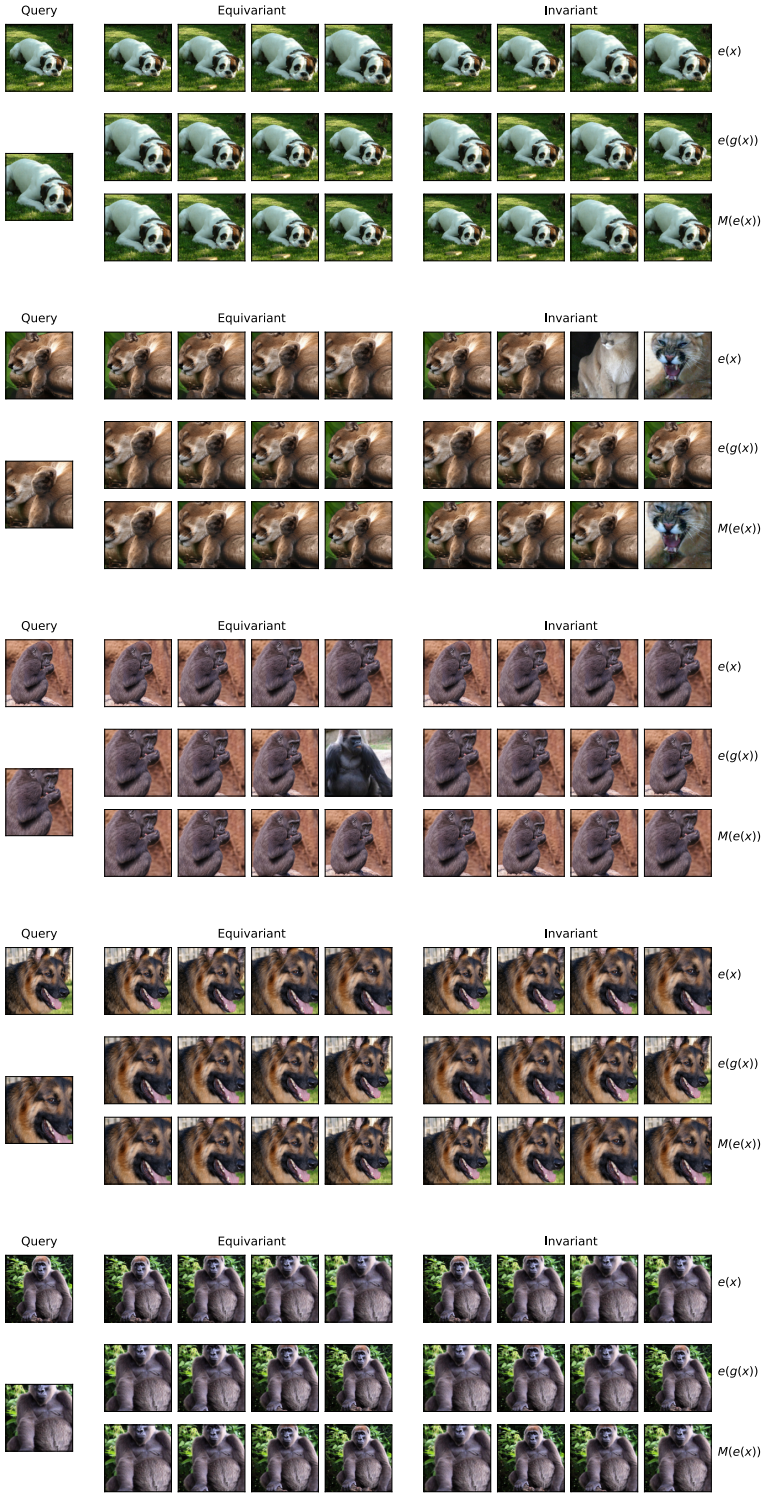


Figure 13: Image Crop/Zoom Retrieval Examples.

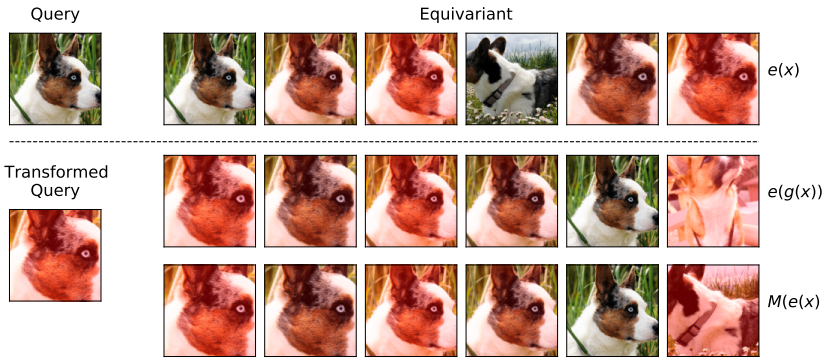


Figure 14: Image Color-Crop Composition Retrieval Examples.


Figure 15: Image Brightness Retrieval Examples.