
Cyclic Test Time Augmentation with Entropy Weight Method (Supplementary material)

Sewhan Chun¹

Jae Young Lee²

Junmo Kim²

¹NAVER CLOVA, Republic of Korea

²SIIT Lab, KAIST, Republic of Korea

1 APPENDIX A: DETAILS OF TTA POLICY

1.1 PREDEFINED TRANSFORMATIONS

In this study, we use predefined transformations, similar to the baseline. However, in order to leverage the effect from the iterative transformation of cyclic mechanism, we change the magnitudes and establish a modified set of transformations as Figure 1. The 12 transformations include: identity, rotation, zooming, contrast adjustment, sharpness adjustment, and color saturation change. The transformations were performed using functions from torchvision and PIL libraries. For identity transformation no actual transformation of image is conducted for the image, leaving no changes. Rotation refers to clock/counter clock wise rotational transformation with magnitude of 20 or -20 degrees. Zoom transformation refers to resizing the image and cropping the central part of the image. Auto contrast transformation calculates the color histogram and stretches it to achieve uniform contrast level. Sharpness transformation refers to adjustment in sharpness, less than 1.0 meaning edges will be blurred and larger than 1.0 refers to sharpening of the edges. Color saturation changes the the saturation values of each pixel color values. For all the padding-requiring transformations, symmetrical padding was used.

1.2 CONVENTIONAL TEST TIME AUGMENTATION

The typical conventional TTA includes center crop, horizontal flip, "5 crops", and "10 crops". Many performances on ImageNet competitions have used such schemes for performance improvement [Krizhevsky et al., 2012, Szegedy et al., 2015, Simonyan and Zisserman, 2015, He et al., 2016]. Center crop method basically utilizes the central cropped patch from an original image. Horizontal flip refers to utilizing a horizontally flipped image with addition to the center cropped image. "5 crops" takes the four corner image

patches and central cropped image patch to fully examine the entire area of the original image. "10 crops" refers to composing 5 additional cropped image patches from horizontally flipped version of the original image to the original 5 crops, in total of 10. The demonstration is presented in Figure 2.

1.3 IMAGENET-C CORRUPTIONS

In this study we have used ImageNet-C benchmark, corrupted version of ImageNet by Hendrycks et al. [Hendrycks and Dietterich., 2019]. ImageNet-C simulates 19 different types of corruptions: Gaussian Noise, Shot Noise, Impulse Noise, Defocus Blur, Glass Blur, Motion Blur, Zoom Blur, Snow, Frost, Fog, brightness, Contrast, Elastic Transform, Pixelate, JPEG, Speckle, Gaussian Blur, Spatter, and Saturate. Each of these corruptions is simulated in 5 different severities. The performance on this dataset is calculated with mean Corruption Error, *mCE*. The *mCE* is calculated as average error rate of a neural network to the error rate of Alexnet [Krizhevsky et al., 2012]. Thus, for *mCE*, 100% indicates a neural network performance of equal robustness to the Alexnet, and 0 being a perfectly robust neural network. Illustration of all the types of corruption on a single image is presented in Figure 3.

2 APPENDIX B: DIFFERENCE BETWEEN SINGLE ITERATION TO CYCLIC ITERATION

As we explained in the main material, single iteration and cyclic iteration differ in number of transformation on a single image. While single method selects a single transformation from the set of the predefined transformations, the cyclic method could select combination of multiple number of transformations for a single image. The illustration is presented in Figure 4.

Modified Predefined Transformations

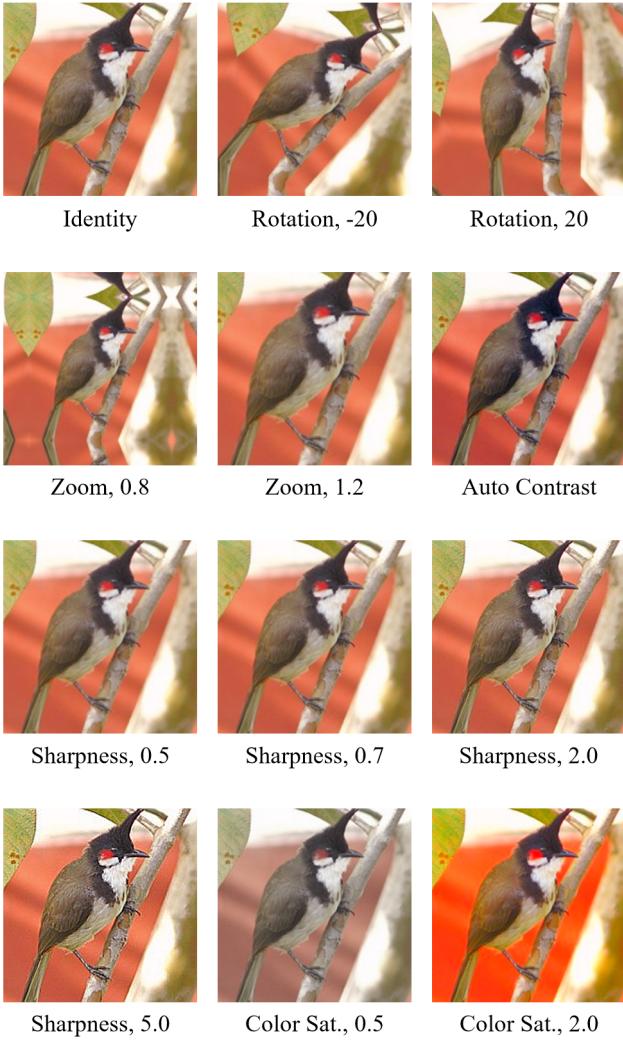


Figure 1: Illustration of the 12 predefined transformations on an image of a bulbul from ImageNet. 12 different transformations were used, identity transformation being the exit signal in cyclic TTA.

3 APPENDIX C: ORACLE-TTA PERFORMANCE

In this section, we demonstrate Oracle-TTA performance on ImageNet and ImageNet-C dataset. Oracle is hypothetically well trained loss predictor, which is able to perfectly predict which of transformation candidates is needed to result in the least loss value by the target network. The following experiment in Table 2 suggests the upper bound and potential rooms for improvement in the loss prediction pipeline. Cyclic behavior allows more opportunities for the the pipeline to generate well conditioned image, resulting

in lower error rate as the maximum number of iteration increases.

4 APPENDIX C: DEMONSTRATION OF ENTROPY WEIGHT METHOD

In this section, we will demonstrate how visually EWM corresponds to TTA. EWM calculates relative entropy of each augmentation and inversely reflects the entropy to the augmentation, as the uncertainty. The importance of each augmentation is represented by respective weight value, which corresponds to the “certainty”. The demonstration of EWM is presented in Figure 5, 6, and 7.

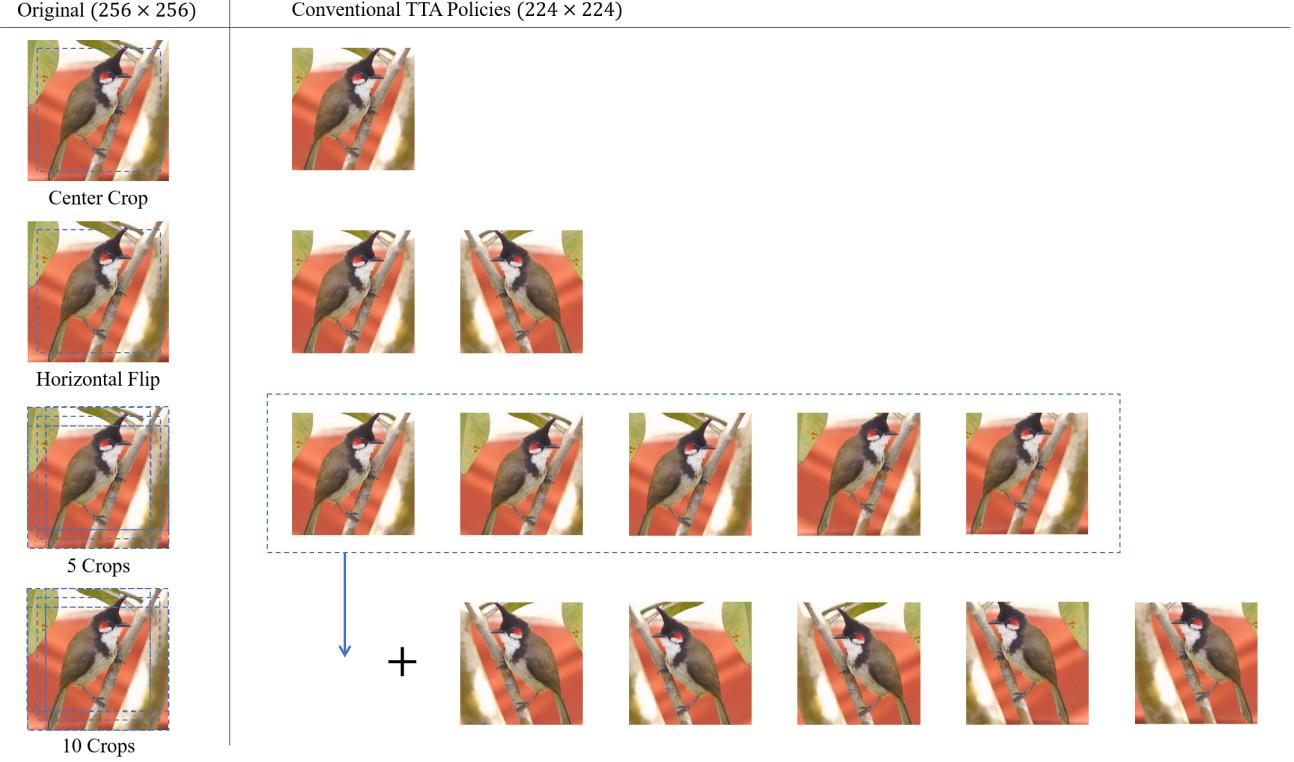


Figure 2: Demonstration of the conventional TTA policies. The cost (number of input images) increases from 1 to 2, 5, and 10 as goes from center crop to horizontal flip, “5 crops”, and “10 crops”.

References

- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- D. Hendrycks and T. Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *Proceedings of the International Conference on Learning Representations*, 2019.
- Ildoo Kim, Younghoon Kim, and Sungwoong Kim. Learning loss for test-time augmentation. In *Advances in Neural Information Processing Systems*, volume 33, 2020.
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, 2012.
- Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR*, 2015.
- Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with

convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

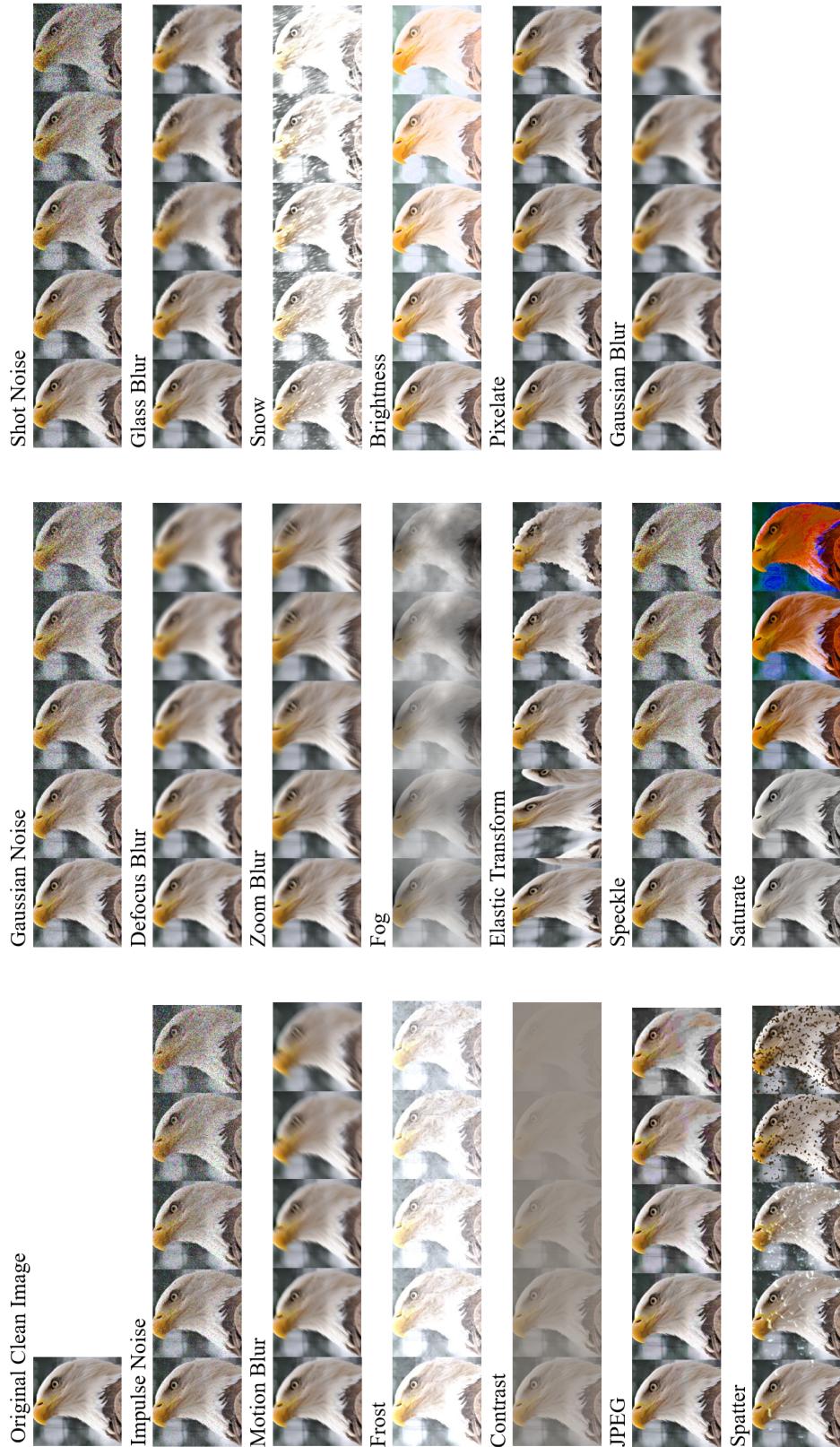


Figure 3: Illustration of the 19 different types of corruptions from ImageNet-C with a picture of an eagle.

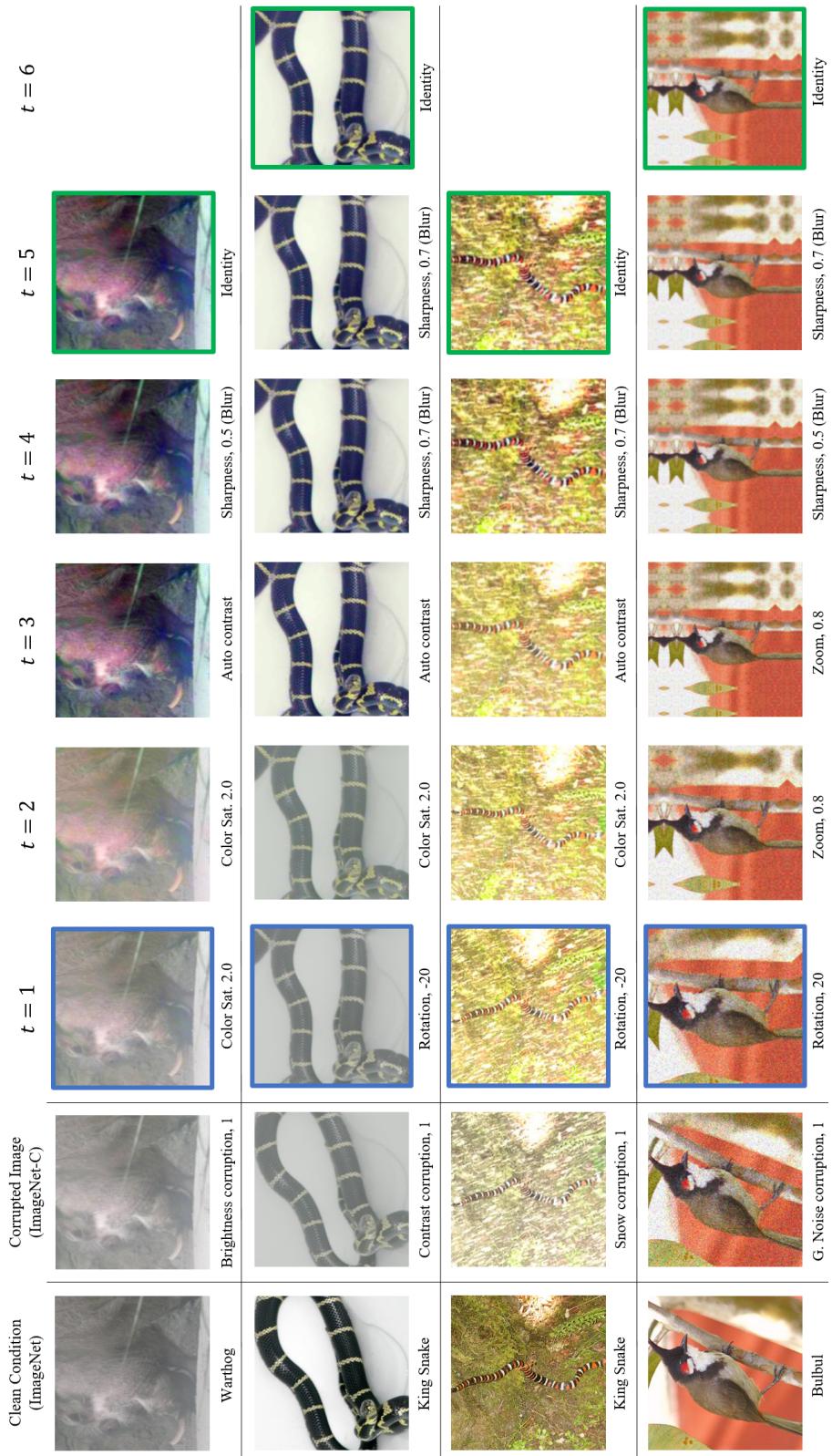


Figure 4: Illustration of exemplary cases of the loss prediction pipeline. The blue bordered images refers to the finalized input images by the previous method [Kim et al., 2020] with single iteration. That of green refers to the transformed input images by ours.

Train Time Augmentation	TTA Method	Cost	Average		EWM (Ours)	
			Clean	<i>mCE</i>	Clean	<i>mCE</i>
Augmix	Center Crop	1	23.61	37.19		
	Horizontal Flip	2	22.59	35.96	22.51	35.96
	5 Crops	5	23.30	35.79	23.17	35.61
	10 Crops	10	22.35	35.64	22.29	35.50
	Single	1	23.48	35.69		
		2	22.13	34.78	22.06	34.42
		3	21.87	34.87	21.85	34.36
	Cyclic (Ours)	1	23.49	34.86		
		2	22.10	34.21	22.12	34.08
		3	21.93	34.45	21.87	34.02

Table 1: Performance comparison of the previous methods with the proposed method on CIFAR100 and CIFAR100-C. Note that *mCE* metric equation for CIFAR100-C is different from that of ImageNet-C. *mCE* for CIFAR100-C is not measured relative to the performance of AlexNet, but it measures the average of accuracies of all the corrupted versions of the dataset.

TTA Method	Iteration	Clean	<i>mCE</i>
Center Crop	0	22.39	65.07
Oracle-TTA	1	15.99	53.28
Cyclic Oracle	2	13.06	47.20
	3	11.68	44.02

Table 2: Performance using hypothetically perfect loss predictor named Oracle-TTA. Resnet-50 trained with the augmix data augmentation method was used as the target network. “Iteration” refers to the number of loss prediction done to each image. Clean and *mCE* refers to the error rate from ImageNet dataset and the mean corruption rate from ImageNet-C dataset. The lower values represent better performance. The correlation between the “Iteration” and the performance indicates that, upon selecting suitable transformation, more opportunities for transformation allows the input image to be better ready for the target network, which increases the upper bound of the general performance.



Figure 5: Demonstration of EWM on two images from ImageNet-C with 10 crops TTA policy. Numerical values below the each image refers to relative entropy, calculated by the target network. Red border refers to the patches with high entropy (uncertainty). **Top:** An image of dogs, divided into 10 patches. The red bordered patches contain image of a dog with exclusion of head part in the central dog’s face. **Bottom:** An image of a man, divided into 10 patches. The red bordered patches have missing part of the facial feature, leaving relatively high uncertainty.

Original Image	Modified Input Image: Single Iteration		
Gaussian Noise (severity = 1)  ruddy turnstone	$i = 1$ Sharpness, 0.5  0.2392	$i = 2$ Sharpness, 0.2  0.6142	$i = 3$ Zoom, 0.8  0.1465
 yawl	Zoom, 0.8  0.2274	Sharpness, 0.2  0.2474	Sharpness, 0.5  0.5252

Figure 6: Demonstration of EWM on two images from ImageNet-C with baseline loss prediction TTA policy [Kim et al., 2020] with single iteration of transformation prediction. 3 augmentations were used as hyper parameter, resulting in 3 input images in total. Numerical value below the each image refers to relative entropy, calculated by the target network. **Top:** With a symmetrical padding, the zoom out transformation has presented more coherent level of feature in bird’s facial structure ($i = 3$), as it had minor cut by the border of the image. **Bottom:** An image of a yawl is also aided by zoom out transformation ($i = 1$), the padding has created the fake reflection of the yawl to the water, resulting in less entropy (uncertainty). The set of predefined transformation from previous method was used.

Original Image	Modified Input Image: Cyclic Iteration		
	$i = 1$	$i = 2$	$i = 3$
Gaussian Noise (severity = 1)			
			
Maltese dog	0.2638	0.4260	0.3102
	Sub Policy 1: Sharpness, 0.5 Sharpness, 0.5 Sharpness, 0.5 Identity	Sub Policy 2: Zoom, 0.8 Sharpness, 0.5 Sharpness, 0.5 Sharpness, 0.5 Identity	Sub Policy 2: Sharpness, 0.7 Sharpness, 0.5 Sharpness, 0.5 Sharpness, 0.5 Identity

Figure 7: Demonstration of EWM on an images from ImageNet-C with our cyclic TTA method. 3 augmentations were used as hyper parameter, resulting in 3 input images in total. Each image has tried to remove the Gaussian noise by performing blurring and zoom out. The entropy among the images show somewhat relatively uniform values.