Fourier-Based Augmentations for Improved Robustness and Uncertainty Calibration

Ryan Soklaski[*]
MIT Lincoln Laboratory
Lexington, MA 02421-6426
ryan.soklaski@ll.mit.edu

Michael Yee
MIT Lincoln Laboratory
Lexington, MA 02421-6426
myee@ll.mit.edu

Theodoros Tsiligkaridis
MIT Lincoln Laboratory
Lexington, MA 02421-6426
ttsili@ll.mit.edu

Abstract

Diverse data augmentation strategies are a natural approach to improving robustness in computer vision models against unforeseen shifts in data distribution. However, the ability to tailor such strategies to inoculate a model against specific classes of corruptions or attacks—without incurring substantial losses in robustness against other classes of corruptions—remains elusive. In this work, we successfully harden a model against Fourier-based attacks, while producing superior-to-`AugMix` accuracy and calibration results on both the CIFAR-10-C and CIFAR-100-C datasets; classification error is reduced by over ten percentage points for some high-severity noise and digital-type corruptions. We achieve this by incorporating Fourier-basis perturbations in the `AugMix` image-augmentation framework. Thus we demonstrate that the `AugMix` framework can be tailored to effectively target particular distribution shifts, while boosting overall model robustness.

1 Introduction

Despite the chart-topping performances of CNN-based models across both standard and domain-specialized computer vision benchmarks, the tendency for these models to behave unreliably in the face of subtle changes to data distribution make them liable to fail in deployment [3, 4, 14]. This lack of robustness has been broadly attributed to the proclivity for CNNs to fit on unintuitive and superficial patterns that exist among the training data [3, 9, 10]. Thus the need for more comprehensive benchmarks—including those that test models against common corruptions and perturbations [5, 6, 8, 13, 15, 17, 18]—and the need for more varied data augmentation strategies have manifest.

An analysis of these common image corruptions from a Fourier perspective characterized the degree to which natural perturbations can be summarized by high, mid, and low-frequency variations in intensity over pixel-space [22]. It demonstrated that performance trade-offs in robustness between different categories of corruptions often manifest across different characteristic frequency regimes. E.g., a targeted improvement in robustness to Gaussian noise (high-frequency characteristics) will be concomitant with marked loss in robustness to a fog corruption (low-frequency characteristics) [2, 22]. Whereas targeted data-augmentation strategies suffer from this trade-off, those strategies that

[*] DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited. This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering. © 2021 Massachusetts Institute of Technology. Delivered to the U.S. Government with Unlimited Rights, as defined in DFARS Part 252.227-7013 or 7014 (Feb 2014). Notwithstanding any copyright notice, U.S. Government rights in this work are defined by DFARS 252.227-7013 or DFARS 252.227-7014 as detailed above. Use of this work other than as specifically authorized by the U.S. Government may violate any copyrights that exist in this work.

Preprint. Under review.

mix a diverse set of augmentation primitives – to include low, mid, and high-frequency characteristics – have proven to be effective at boosting overall model robustness[1, 7]; the `AugMix` augmentation method is particularly outstanding in this regard [7].

Still outstanding is the need to inculcate a model with robustness to particular distribution shifts – in a targeted manner – while maintaining robustness to common corruptions. There are well-known universal adversarial perturbations (UAP) that can be applied uniformly across images in order to induce high error rates across convolution-based computer vision models [14]. Indeed, it has been shown that even simple sinusoidal noise, designed to target particular frequency-and-orientation regimes, can serve as an effective UAP; these so-called Fourier-basis attacks can reduce image-classification models to near-guessing performance [7, 20, 22]. It is of manifest importance that models can be made robust both to common, naturally-occurring corruptions, as well as known black-box adversarial perturbations such as these.

Thus, in this work we investigate the capacity for `AugMix` to be modified to target particular data distribution shifts, while maintaining its outstanding ability to instill robustness against common corruptions. We specifically seek protection against Fourier-basis attacks, which are noteworthy for their simplicity, their efficacy at diminishing model performance, as well as their ability to be tailored to target any frequency regime in input space.

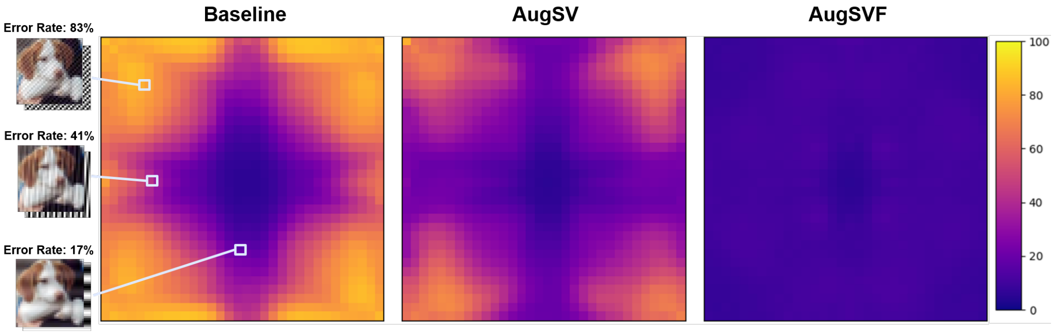


Figure 1: Classification error heatmaps for a model trained using three data augmentation techniques, in association with Fourier-basis perturbations applied to all CIFAR-10 test images. `AugSV` (i.e, `AugMix` [7]) improves robustness to some high and mid-frequency regions. `AugSVF`, which explicitly incorporates distinct Fourier-basis augmentations, greatly reduces the model's susceptibility across all frequency regimes.

2 Fourier-Basis Perturbations

A real-valued 2D Fourier-basis matrix of shape $d_x \times d_y$, $U(\vec{k}) \in \mathbb{R}^{d_x \times d_y}$, has its elements defined by a 2D sinusoidal plane wave that is evaluated on the discrete grid $P = \mathbb{I}^{\{0,\ldots,d_x-1\}\times\{0,\ldots,d_y-1\}}$, i.e.,

$$U_{x,y}(\vec{k}) = A\cos(2\pi f \vec{r} \cdot \hat{k} - \phi) \tag{1}$$

where $\vec{r} = [x, y] \in P$. A is selected such that $||U||_2$ is fixed [2]. Each basis matrix is characterized by a so-called wave vector, $\vec{k} = [k_x, k_y] \in K \subseteq P$, whose magnitude determines the frequency of the sinusoid, $f = ||(k_x/d_x, k_y/d_y)||_2$, and whose direction determines the orientation of the wave. K ($\subseteq P$) corresponds to the set of non-degenerate plane waves with a common phase shift of $\phi = \frac{\pi}{4}$.[3] Three example basis matrices are depicted in Figure 1.

A Fourier-basis perturbation consists of adding $U(\vec{k})$, elementwise, to each color channel of a shape-$d_x \times d_y$ image, with strength controlled by $||U||_2$. The perturbation can also be applied to each channel with a randomly-drawn sign applied to it; this is referred to as a “random-flip” perturbation.

[2] The definition of a 2D Fourier-basis matrix provided in section 2 of [22] is written in terms of the behavior of $U_{x,y}$ under the action of a 2D discrete Fourier transform. While their definition may not immediately resemble that of Equation 1, the two are in fact identical.

[3] The cardinality of K has an upper bound of $d_1 \times (\lfloor \frac{d_2}{2} \rfloor + 1)$ [16].

Table 1: A comparison of augmentation strategies across datasets and performance metrics. CIFAR-C results are averaged across all corruptions and severities. AugSVF produces superior calibration metrics across the board, and yields outstanding robustness on corruption datasets.

	Augmentations	Base	AugF	AugS	AugSF	AugSV	AugSVC	AugSVF
Datasets	Error Metrics							
CIFAR-10	Classification	5.7	6.3	5.6	5.9	5.1	**5.1**	5.2
	RMS Calibration	7.4	7.2	5.4	5.4	3.6	4.2	**3.3**
CIFAR-10-C	Classification	27.0	18.1	15.2	11.6	10.9	10.9	**9.9**
	RMS Calibration	23.6	16.0	12.2	9.8	7.6	8.4	**6.7**
CIFAR-100	Classification	25.5	27.8	26.7	27.5	24.2	**23.5**	24.9
	RMS Calibration	15.9	14.7	12.6	12.4	7.4	7.9	**7.0**
CIFAR-100-C	Classification	53.6	46.4	42.0	38.6	36.3	35.5	**34.8**
	RMS Calibration	33.7	25.2	21.3	18.0	13.8	14.5	**12.1**

Susceptibility heatmaps for channel-aligned perturbations with $||U||_2 = 2$ is shown in Figure 1. It reports $d_x \times d_y$ metric values that measure a model's performance on $d_x \times d_y$ perturbed test sets. Entry $[k_x, k_y]$ $(\in K)$ corresponds to a perturbation using $U(\vec{k} = [k_x, k_y])$ applied identically to every image in the test set; the color of the pixel at $[k_x, k_y]$ thus conveys the model's classification error rate on that perturbed test set. The heatmap is arranged so the lowest-frequency resides at its center; it is symmetrized about its center, due to the noted degeneracies among $\vec{k} \in P$.

2.1 Incorporating Fourier-Basis Perturbations in AugMix

The standard AugMix implementation involves a set of five spatial primitives (S) p_{rotate}, $p_{\text{shear}_{\{\text{x,y}\}}}$, $p_{\text{trans}_{\{\text{x,y}\}}}$ and a set of four "vision" primitives (V) p_{contrast}, p_{EQ}, p_{poster}, p_{solar}. AugS, AugV, and AugSV will indicate the inclusion of the primitives in S, V, and $S \cup V$ in the AugMix framework, respectively. Thus AugSV corresponds to the original implementation in [7] (see Appendix for more details). We can naturally include a Fourier-basis perturbation primitive (denoted by F) among the primitives used by AugMix. The elements of stochasticity for this augmentation result from sampling: $\vec{k}$, ϕ, and $||U||_2$. We sample over frequencies ($||\vec{k}||_2$) uniformly.[4] E.g. AugSVF includes spatial, vision, and Fourier primitives. Lastly, there four primitives that overlap with the CIFAR-C corruptions, $p_{\text{brightness}}$, p_{color}, p_{contrast}, and p_{sharp}, which we will denote using C.

3 Experimental Results

We train a Wide ResNet architecture [23] on both CIFAR-10 and CIFAR-100, respectively, using a variety of data-augmentation strategies. We evaluate the following augmentation strategies: Base, AugF, AugS, AugSF, AugSV, AugSVC, and AugSVF. For each augmentation technique, we train the model using four independent seeds and report the average performance. The baseline augmentation entails randomly flipping and cropping an image, and then normalizing it. We evaluate the model's classification error, as well as the RMS calibration error [12]. To assess the robustness of these models to unforeseen, common image corruptions, we include evaluations against all categories and severities of corruptions in the CIFAR-10-C and CIFAR-100-C datasets [6]. Our architecture, hyperparameters, and training methods match exactly those used in [7]; all models are trained using Jensen-Shannon divergence as a consistency loss [21].

3.1 CIFAR-C Evaluations

The performance metrics reported in Table 1 reveal the efficacy of incorporating Fourier-basis permutations in the AugMix framework. AugSFV produces the best classification performance on the CIFAR-C datasets, and superior calibration results on all benchmarks. The inclusion of AugSVC helps to demonstrate that simply adding any additional primitives to AugMix does not necessarily net a gain in performance. Indeed, AugSVC harms calibration across the board and is inconsistent in its

[4] An un-weighted sampling of the Fourier-bases would favor high-frequency perturbations.

Table 2: Comparing vulnerabilities to Fourier-based attacks. AugSVF greatly reduces the impact that Fourier-basis corruptions have on the model's performance. The phase-shift used for all perturbations was held-out at train time for AugSVF, nor did AugSVF train on $||U||_2$ exceeding 2.

	Metric	Mean (Max) Class Error			Mean (Max) RMS Cal Error		
	Method	Base	AugSV	AugSVF	Base	AugSV	AugSVF
Dataset	$\|\|U\|\|_2$						
CIFAR-10	1	24 (83)	12 (38)	**6 (8)**	22 (70)	8 (28)	**5 (6)**
	2	46 (88)	28 (83)	**9 (15)**	38 (86)	19 (73)	**6 (11)**
	3	60 (90)	42 (89)	**13 (27)**	48 (89)	29 (79)	**9 (20)**
	4	68 (90)	53 (90)	**19 (44)**	54 (90)	37 (78)	**13 (32)**
CIFAR-100	1	55 (92)	39 (69)	**27 (29)**	34 (66)	16 (43)	**8 (10)**
	2	74 (98)	58 (95)	**31 (40)**	50 (91)	30 (82)	**11 (16)**
	3	83 (99)	70 (98)	**36 (55)**	58 (97)	43 (94)	**14 (26)**
	4	88 (99)	78 (99)	**43 (69)**	62 (98)	52 (96)	**17 (38)**

impact on classification error. This is despite the fact that AugSVC explicitly incorporates corruptions that overlap with CIFAR-C.

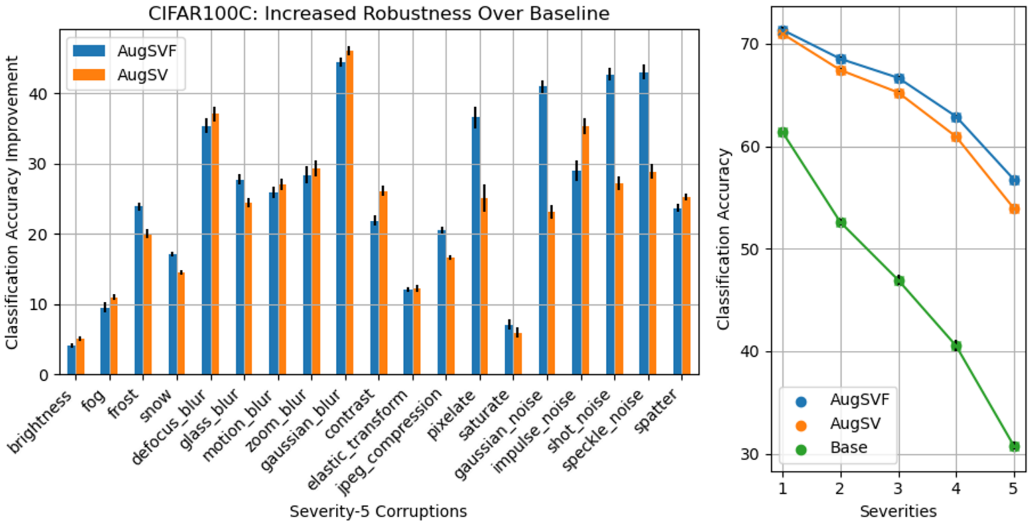


Figure 2: (Left) Enhanced robustness in classification accuracy over the baseline, across severity-5 corruption categories for CIFAR-100-C. AugSVF yields a significant boost in robustness to "noise" and "digital" corruptions, without any significant trade-off elsewhere. (Right) CIFAR-100-C classification accuracies averaged over corruptions, for each corruption severity.

Examining these results prior to aggregation—over severity and corruption-type—reveals that the gains in robustness achieved by AugSFV grows relative to AugMix with increasing corruption severity. Figure 2 shows, for each severity-5 corruption in CIFAR-100-C, the improvement that AugMix and AugSVF achieve over the baseline's classification accuracy for that category. AugSFV offers substantial boosts in robustness for several noise-type and digital-type corruptions, with improvements exceeding ten percentage points for multiple categories. Despite these marked improvements against distinctly "high-frequency" corruptions [22], AugSFV does not suffer any substantial trade-offs in other categories, other than against contrast and impulse noise; it otherwise sustains the gains in robustness achieved by AugMix. The error bars denote standard deviations across independent seeds.

3.2 Evaluating Against Fourier-Based Attacks

We assess the ability of models that leverage the Base, AugSV, and AugSV augmentation strategies, respectively, to generalize to data that has been corrupted with Fourier-basis noise. We corrupt the entire test using every Fourier-basis matrix; thus we evaluate all possible frequencies, orientations, and channel-alignments. The perturbations use a phase-shift held-out from training for AugSFV, nor did AugSVF include perturbations with $||U||_2 > 2$ during training. Table 2 reports results for both random and targeted Fourier attacks of varying strengths, as measured via mean and max error rates, respectively. The maximum error metrics correspond to a targeted attack where a single, worst-case Fourier-basis is selected to modify the entire test set.

Figure 1 and Table 2 report that AugSVF is highly robust to corruptions of all frequencies, whereas Base and AugSV are reduced to near-guessing for targeted attacks with $||U||_2 \geq 2$, and a random $||U||_2 = 1$ corruption roughly doubles their clean error rates. Thus incorporating Fourier perturbations in AugMix is effective at targeting this distribution shift.

4 Conclusion

In this work we modify AugMix to protect against Fourier-basis attacks – a particular form of data distribution shift that can be used as an effective universal adversarial perturbation – while improving robustness against common corruptions and uncertainty calibration. We find that inoculating a model against such a fundamental class of perturbations does not degrade performance, despite the noted propensity for models to "attach" to spurious statistics. Furthermore, although these perturbations explicitly span all frequencies, there is only an enhancement – and no loss – in robustness to a wider variety of corruptions. Indeed, we obtain outstanding results on CIFAR-10-C and CIFAR-100-C, while greatly ameliorating the model's susceptibility to Fourier-based attacks.

Future work will investigate whether achieving robustness to Fourier-basis attacks provides robustness against other universal adversarial perturbations that have distinctive frequency characteristics. For instance, the UAP developed by Moosavi-Dezfooli et. al. [14] have been shown to exploit narrow regimes of high-frequency patterns to fool models [20], however we have shown here that our AugSFV augmentation method dramatically reduces a model's susceptibility in these regimes. This gets at a broader question to be investigated: how reliable is the Fourier perspective on robustness? In particular, does improved robustness to plane-wave perturbations in a particular sub-domain of frequencies and orientations reliably indicate that a model will be more robust to corruptions that are prominently comprised of these plane wave components? Lastly, the simplicity and composeability of Fourier-bases may make them an especially effective augmentation strategy in applied domains, such as medical imaging, where other common photography-inspired augmentations are not appropriate; thus Fourier-based augmentations should be tested on a broad range domain-specific datasets such as those provided by WILDS [11].

References

[1] Mane Cubuk Zoph and Le Vasudevan. "Autoaugment Learning augmentation policies from data". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 113–123.

[2] Nic Ford et al. "Adversarial examples are a natural consequence of test error in noise". In: *arXiv preprint arXiv:1901.10513* (2019).

[3] Robert Geirhos et al. "Generalisation in humans and deep neural networks". In: *Advances in Neural Information Processing Systems* 31 (2018), pp. 7538–7550.

[4] Amir Globerson and Sam Roweis. "Nightmare at test time: robust learning by feature deletion". In: *Proceedings of the 23rd international conference on Machine learning*. 2006, pp. 353–360.

[5] Yue He, Zheyan Shen, and Peng Cui. "Towards non-iid image classification: A dataset and baselines". In: *Pattern Recognition* 110 (2021), p. 107383.

[6] Dan Hendrycks and Thomas Dietterich. "Benchmarking Neural Network Robustness to Common Corruptions and Perturbations". In: *International Conference on Learning Representations*. 2018.

[7] Dan Hendrycks et al. "AugMix: A Simple Data Processing Method to Improve Robustness and Uncertainty". In: *International Conference on Learning Representations*. 2019.

[8] Dan Hendrycks et al. "Natural adversarial examples". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 15262–15271.

[9] Joern-Henrik Jacobsen et al. “Excessive Invariance Causes Adversarial Vulnerability”. In: *International Conference on Learning Representations*. 2018.

[10] Jason Jo and Yoshua Bengio. “Measuring the tendency of cnns to learn surface statistical regularities”. In: *arXiv preprint arXiv:1711.11561* (2017).

[11] Pang Wei Koh et al. “Wilds: A benchmark of in-the-wild distribution shifts”. In: *International Conference on Machine Learning*. PMLR. 2021, pp. 5637–5664.

[12] Ananya Kumar, Percy Liang, and Tengyu Ma. “Verified uncertainty calibration”. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. 2019, pp. 3792–3803.

[13] Shangyun Lu et al. “Harder or different? a closer look at distribution shift in dataset reproduction”. In: *ICML Workshop on Uncertainty and Robustness in Deep Learning*. 2020.

[14] Seyed-Mohsen Moosavi-Dezfooli et al. “Universal adversarial perturbations”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1765–1773.

[15] Norman Mu and Justin Gilmer. “Mnist-c: A robustness benchmark for computer vision”. In: *arXiv preprint arXiv:1906.02337* (2019).

[16] Mark Newman. “The Discrete Fourier Transform”. In: *Computational Physics*. Revised and Expanded. Mark Newman, 2013, pp. 299–300.

[17] Benjamin Recht et al. “Do cifar-10 classifiers generalize to cifar-10?” In: *arXiv preprint arXiv:1806.00451* (2018).

[18] Benjamin Recht et al. “Do imagenet classifiers generalize to imagenet?” In: *International Conference on Machine Learning*. PMLR. 2019, pp. 5389–5400.

[19] Ryan Soklaski and Justin Goodwin. *mit-ll-responsible-ai/hydra-zen: Release hydra-zen v0.3.0rc2*. Version v0.3.0rc2. Sept. 2021. DOI: `10.5281/zenodo.5517572`. URL: `https://doi.org/10.5281/zenodo.5517572`.

[20] Yusuke Tsuzuku and Issei Sato. “On the Structural Sensitivity of Deep Convolutional Networks to the Directions of Fourier Basis Functions”. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society. 2019, pp. 51–60.

[21] Wikipedia contributors. *Jensen–Shannon divergence — Wikipedia, The Free Encyclopedia*. [Online; accessed 10-November-2021]. 2004. URL: `https://en.wikipedia.org/wiki/Jensen%E2%80%93Shannon_divergence`.

[22] Dong Yin et al. “A fourier perspective on model robustness in computer vision”. In: *arXiv preprint arXiv:1906.08988* (2019).

[23] Sergey Zagoruyko and Nikos Komodakis. “Wide Residual Networks”. In: *British Machine Vision Conference 2016*. British Machine Vision Association. 2016.

A Appendix

A.1 Additional Implementation Details of Fourier-Based Perturbations

Although our formulation of $U(\vec{k})$ presented in 2, which explicitly defines a real-valued plane wave sinusoid, does not immediately resemble the Fourier-based definition introduced in [22], the two are exactly equivalent, as are their implementations.

A channel-aligned Fourier-basis perturbation is applied to an image by adding U, which is scaled to have a specified ℓ_2-norm in order to control the strength of the perturbation, to each of the image's color channels. The image's pixel values are assumed to reside in $[0, 1]$ and are thus they are “clipped” to this domain following the channel-wise addition.

A random-flip perturbation follows this same process, but includes a randomly-drawn factor from $\{-1, 1\}$ that multiplies U in association with each color channel. We include both channel-aligned and random-flip perturbations in our analyses, whereas the random-flip type is exclusively considered in some prior works [7, 22].

A.2 Additional Implementations Details of AugMix

The AugMix data processing technique processes an image through parallel chains of randomly-selected and composed augmentation primitives; it randomly-weights and sums these parallel-processed images, and then randomly mixes the result with the original image [7].

An example involving a particular AugMix configuration, for AugSVF, for processing a single image is detailed in equation 2.

$$
\begin{aligned}
(w_1, w_2, w_3) &\sim \text{Dirichlet}(1, 1, 1) \\
m &\sim \text{Beta}(1, 1) \\
\text{AugMix}(x_{\text{img}}) &= (1 - m)\, x_{\text{img}} + m\, \big[w_1\, f_{\text{rotate}} \circ f_{\text{shear}_\text{x}}(x_{\text{img}}) \\
&\qquad + w_2\, f_{\text{equalize}}(x_{\text{img}}) \\
&\qquad + w_3\, f_{\text{fourier}} \circ f_{\text{rotate}}(x_{\text{img}})\big]
\end{aligned}
\tag{2}
$$

Here the "depth"—the number of augmentation primitives composed from each chain—of each of the three chains is drawn uniformly from $1, 2, 3$ with replacement. Once a chain's depth has been determined, that number of augmentations is drawn uniformly from $S \cup V \cup F$. The example in 2 thus shows depths of 2, 1, and 2, respectively.

Prior to this AugMix processing, a random flip-and-crop is applied to the image—as in the "baseline" augmentation strategy. Following the AugMix process, the per-channel data normalization—enforcing a per-channel mean and standard deviation of 0.5—is applied to the "mixed" image.

A.3 Methods

Our architecture, hyperparameters, and training methods match exactly those used in [7] except for one detail: for each seed, we select the best model checkpoint based on performance on a hold-out validation set, whereas the cited work evaluates directly on the test set. We leveraged hydra-zen [19] to produce all of our results in self-documenting and reproducible manner.

Each Fourier-basis perturbation used for the purpose of train-time augmentation has an ℓ_2-norm sampled uniformly from $[1, 2]$. The frequency and orientation are sampled in a weighted fashion such that the distinct frequencies are uniformly represented. Lastly, the phase-shift is sampled uniformly from $\{\frac{0}{3}2\pi, \frac{1}{3}2\pi, \frac{2}{3}2\pi\}$. All test-time evaluations involving Fourier-basis perturbations (e.g. generating a heatmap), utilize a distinct phase-shift of $\frac{3}{4}2\pi$.

A.4 Additional Results

Figure 3 is the complement to Figure 2: it reports the gains in robustness—due to AugSV and AugSVF, respectively—for the various corruption categories in CIFAR-10-C.

A.4.1 Further Presentation and Analysis of Susceptibility Heatmaps

Here we include heatmaps corresponding to Fourier-attacks of various channel-alignments and strengths (i.e. $||U||_2$) against the Base, AugSV, and AugSVF models, which were each trained separately on CIFAR-10 and CIFAR-100. Note the dramatic reduction in susceptibility that is achieved by AugSVF for all channel-alignments and strengths. It should be noted that perturbations with $||U||_2 = 4$ are roughly comparable to severity-5 corruptions from the CIFAR-C datasets, in terms of how visually disruptive they are.

It is interesting to see the different susceptibility patterns that manifest, between the channel-aligned and random-flip attacks; the former tend to have stronger effects in the high-frequency regimes, while the latter impact mid-to-low frequency regimes more severely. Prior papers [7, 22] only considered random-flip channels, which was to overlook an important class of Fourier perturbations and their impact on model performance. For instance AugSV appears to be somewhat robust to Fourier perturbations, in comparison to Base, when only considering random-flip attacks. However, e.g, for CIFAR-10, we see that high error-rates (80%) manifest when AugSV encounters high-frequency, channel-aligned perturbations that are only of strength $||U||_2 = 2$.

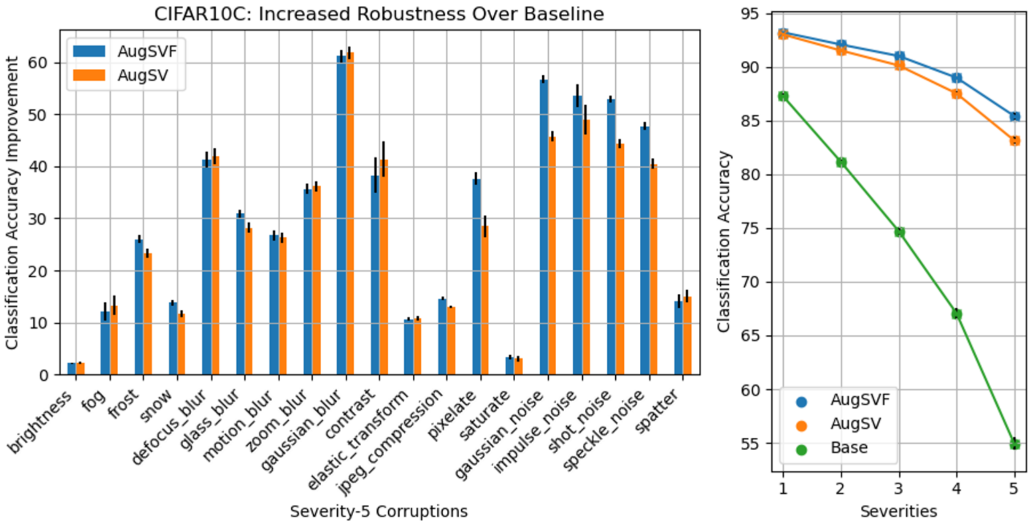
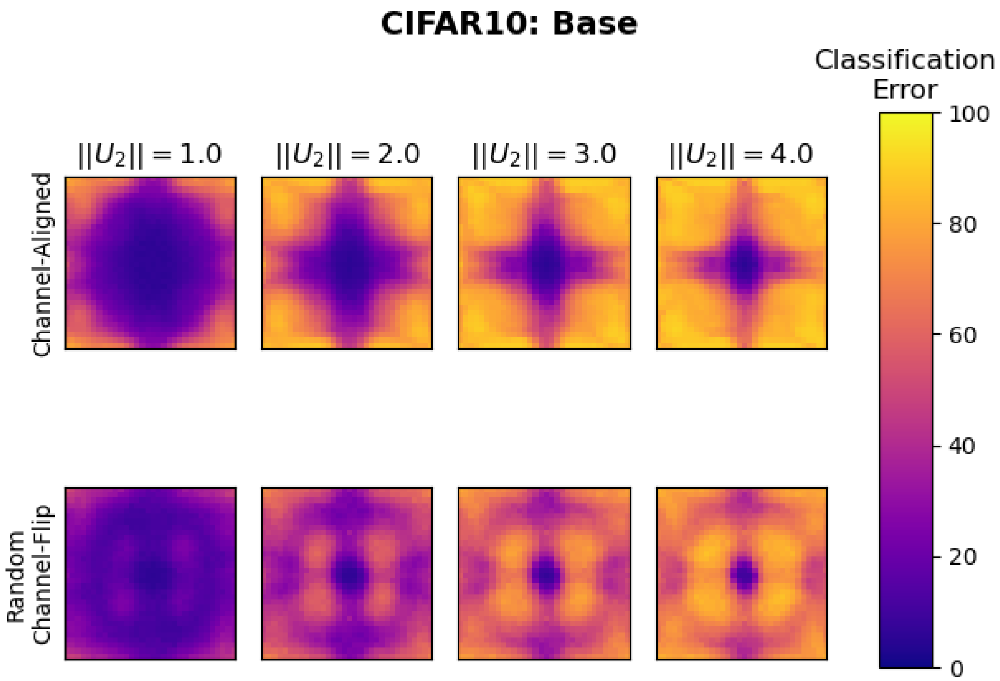
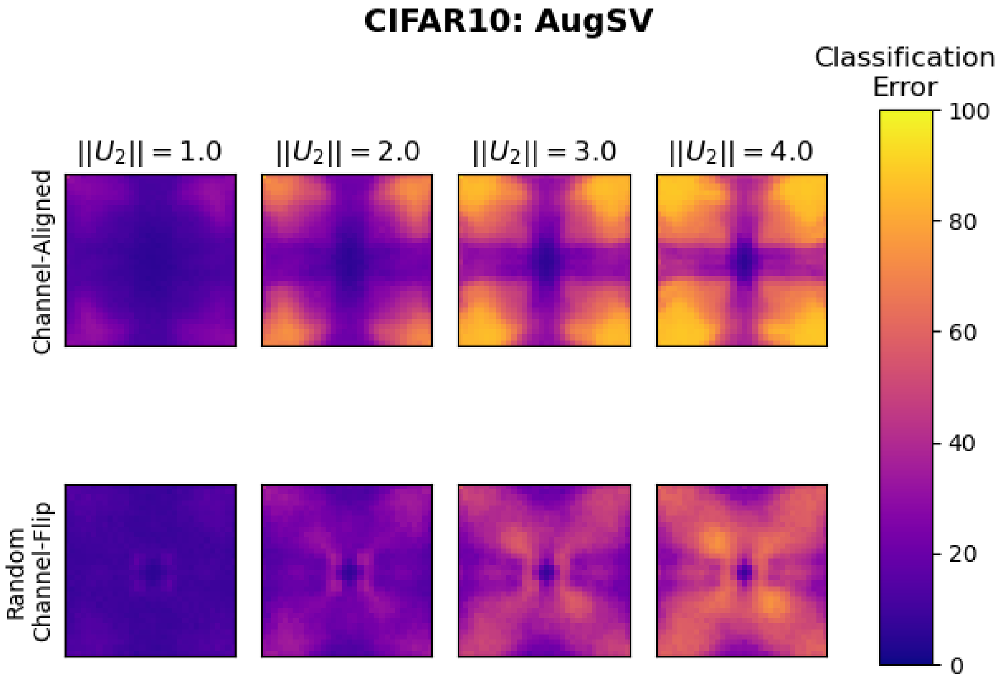
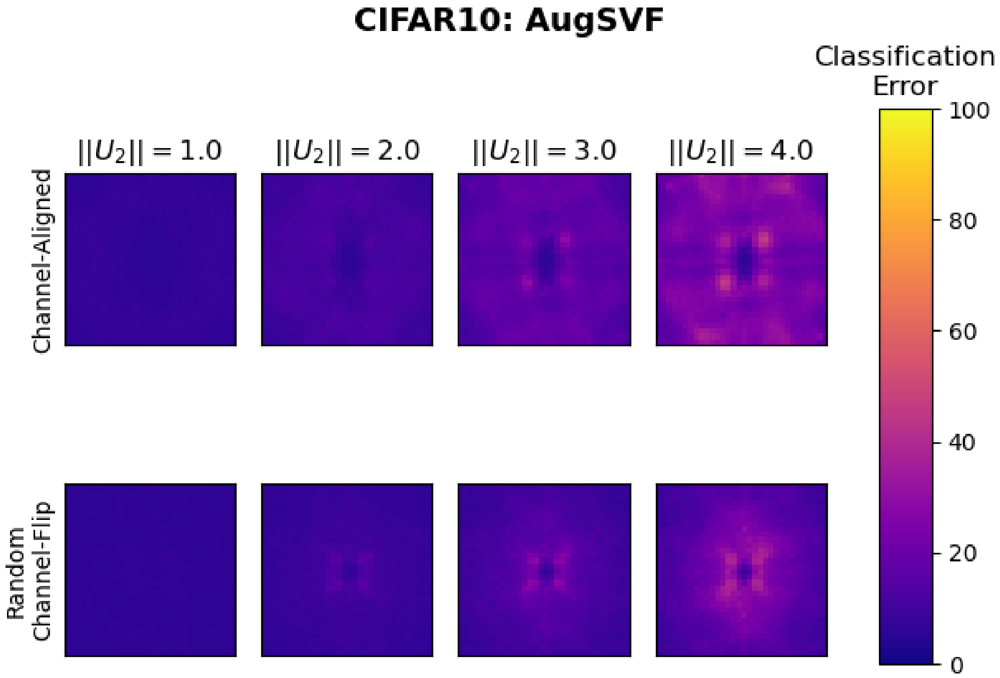
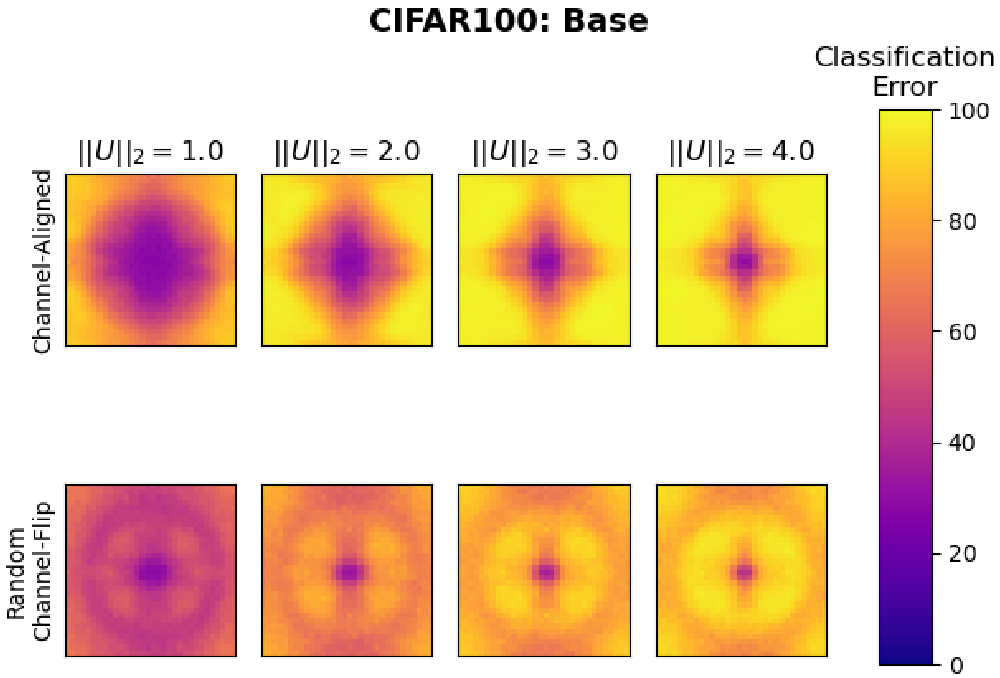


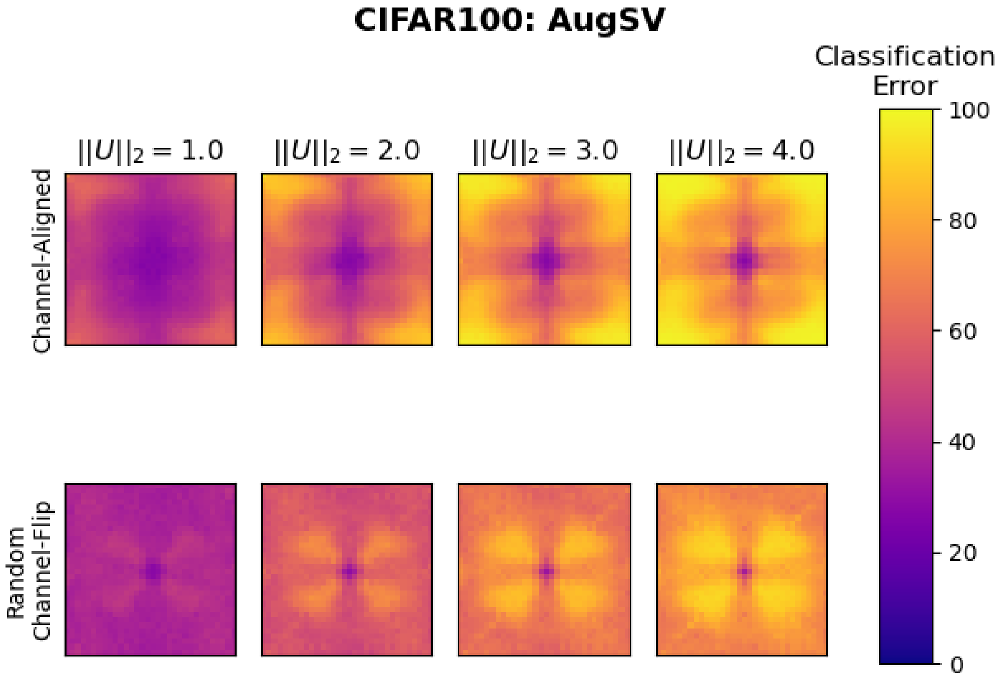
Figure 3: (Left) Enhanced robustness in classification accuracy over the baseline, across severity-5 corruption categories for CIFAR-10-C. `AugSVF` yields a significant boost in robustness to “noise” and “digital” corruptions, without any significant trade-off elsewhere. (Right) CIFAR-10-C classification accuracies averaged over corruptions, for each corruption severity.

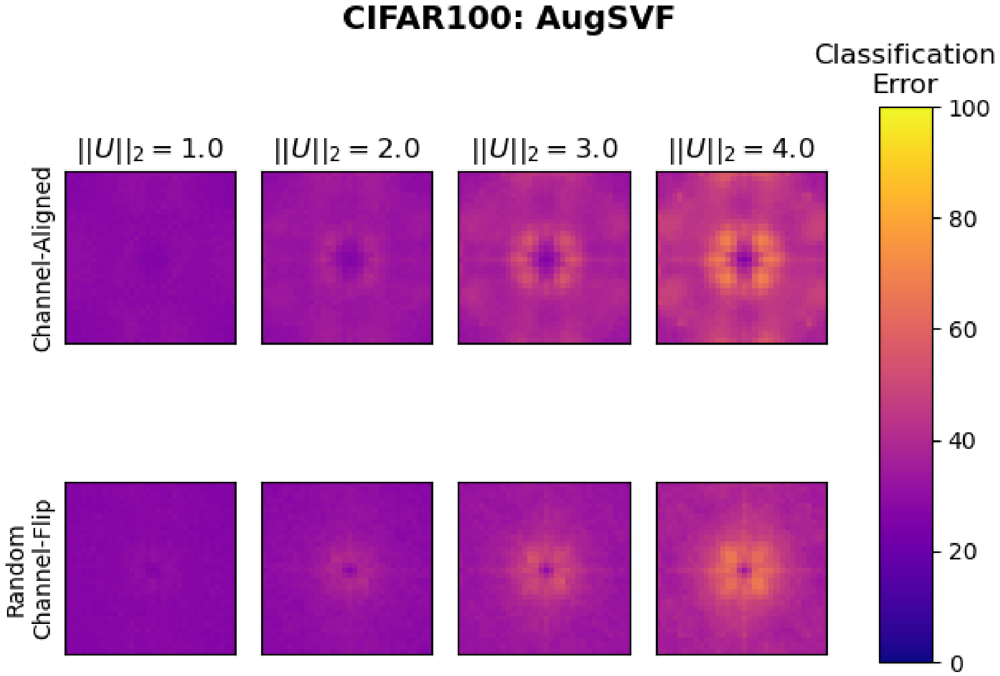



CIFAR10: AugSV
Classification Error
$||U_2|| = 1.0$
$||U_2|| = 2.0$
$||U_2|| = 3.0$
$||U_2|| = 4.0$
Channel-Aligned
Random Channel-Flip
100
80
60
40
20
0


CIFAR10: AugSVF
Classification Error
$||U_2|| = 1.0$
$||U_2|| = 2.0$
$||U_2|| = 3.0$
$||U_2|| = 4.0$
Channel-Aligned
Random Channel-Flip
100
80
60
40
20
0


CIFAR100: Base
Classification Error
100
80
60
40
20
0
$||U||_2 = 1.0$
$||U||_2 = 2.0$
$||U||_2 = 3.0$
$||U||_2 = 4.0$
Channel-Aligned
Random Channel-Flip


CIFAR100: AugSV
Classification Error
100
80
60
40
20
0
$||U||_2 = 1.0$
$||U||_2 = 2.0$
$||U||_2 = 3.0$
$||U||_2 = 4.0$
Channel-Aligned
Random Channel-Flip


CIFAR100: AugSVF
Classification
Error
100
80
60
40
20
0
$||U||_2 = 1.0$
$||U||_2 = 2.0$
$||U||_2 = 3.0$
$||U||_2 = 4.0$
Channel-Aligned
Random
Channel-Flip