Fourier-basis functions to bridge augmentation gap: Rethinking frequency augmentation in image classification

Puru Vaish* Shunxin Wang* Nicola Strisciuglio
University of Twente
{p.vaish, s.wang-2, n.strisciuglio}@utwente.nl

Abstract

Computer vision models normally witness degraded performance when deployed in real-world scenarios, due to unexpected changes in inputs that were not accounted for during training. Data augmentation is commonly used to address this issue, as it aims to increase data variety and reduce the distribution gap between training and test data. However, common visual augmentations might not guarantee extensive robustness of computer vision models. In this paper, we propose Auxiliary Fourier-basis Augmentation (AFA), a complementary technique targeting augmentation in the frequency domain and filling the robustness gap left by visual augmentations. We demonstrate the utility of augmentation via Fourier-basis additive noise in a straightforward and efficient adversarial setting. Our results show that AFA benefits the robustness of models against common corruptions, OOD generalization, and consistency of performance of models against increasing perturbations, with negligible deficit to the standard performance of models. It can be seamlessly integrated with other augmentation techniques to further boost performance. Codes and models are available at https://github.com/nis-research/afa-augment.

1. Introduction

Computer vision models usually encounter performance degradation when deployed in real-world scenarios due to unexpected image variations [10, 14, 17]. Improving the robustness of computer vision models to out-of-distribution (OOD) data is thus essential for their reliable practical use. Among the methods addressing the robustness and generalization of computer vision models [1, 8, 9, 11, 38, 39, 47, 50, 53], data augmentation is mostly used for its easy-to-apply characteristics and effectiveness at reducing the distribution gap between training and test data [45]. Popular augmentation techniques, such as AugMix [15], AugMax [42], AutoAugment [2], TrivialAugment [34], and PRIME [33] have shown great improvements in corruption and perturbation robustness benchmarks and OOD datasets for generalisation, e.g. ImageNet-C, ImageNet-$\bar{C}$, ImageNet-3DCC, ImageNet-P, ImageNet-R and ImageNet-v2 [13, 14, 18, 32, 35]. These approaches mainly focus on adding visual variations to images through random or policy-based combinations [2, 15, 16, 26, 27, 30, 31, 34] of visual transformations aiming at increasing the diversity of training images (expanding on their domain, see visual augmentations in Fig. 1), and adversarial-based augmentations, which address the hardness of training samples but are computationally heavy (see Tab. 1, AugMax). However, even if trained with visual augmentations, models are still sensitive to image variations not included in the training [25] and frequency perturbations [49]. This occurs due to the pre-defined frequency characteristics of visual transformations, which cannot ensure the complete robustness of models against noise with different frequency characteristics from those encountered during training. Attackers may exploit this weakness and degrade model performance in operational settings [23]. This raises a question: *Is there a complementary augmentation technique that can bridge the gap left by visual augmentations?*

*Equal contribution

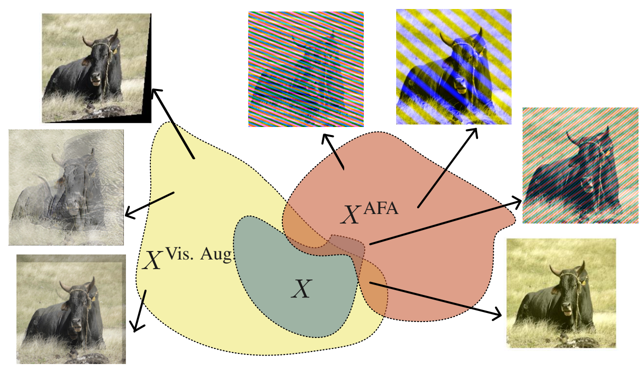

Figure 1. Frequency augmentation with Fourier-basis functions is complementary to common visual augmentations. They appear *unnatural* and can be used as adversarial examples.

Common visual augmentations impact different frequency components of images simultaneously, which are difficult to explicitly control, and might not encompass all possible frequency variations present in unseen corruptions or variantions happening in real-world scenarios [36]. We thus rethink image augmentation in the frequency domain, and complement visual augmentation strategies with explicit use of Fourier basis functions in an adversarial setting. There has been exploration into frequency-based augmentations to discover capabilities beyond what visual augmentations can achieve. [6, 41, 48] swap or mix partial amplitude spectrum between images, aiming to induce more phase-reliance for classification. [43] augments images with shortcut features to reduce their specificity for classification. AugSVF [37] introduces frequency noise within the AugMix framework and [24, 28] adversarially perturb the frequency components of images. These augmentations are computationally heavy, due to the complicated augmentation framework [37], computation of multiple Fourier transforms for training images and their augmented versions [6, 41, 48], identification of learned frequency shortcuts [43], or adversarial training [24, 28].

In this work, we propose Auxiliary Fourier-basis Augmentation (AFA). We use additive noise based on Fourier-basis functions to augment the frequency spectrum in a more efficient way than other methods that apply frequency manipulations [6, 37, 43]. The effect of additive Fourier-basis functions on image appearance is complementary to those of other augmentations (see Fig. 1). These images can be interpreted as samples representing an adversarial distribution, distinct from those augmented by common visual transformations. We thus expand upon the conventional idea of adversarial augmentation, moving beyond the generation of imperceptible noise through gradient backpropagation. We employ a training architecture and strategy with an auxiliary component to address the adversarial distribution, and a main component for the original distribution, similarly to AugMax [42]. However, the adversarial distribution that we construct using additive Fourier-basis is much less computationally expensive than that of AugMax (and other visual augmentation methods - see Tab. 1). It contributes to comparable or higher generalization results, while allowing for the training of larger models on larger datasets (e.g. ImageNet). Our contributions are:

- We propose a straightforward and computationally efficient augmentation technique called AFA. We show that it enhances robustness of models to common image corruptions, improves OOD generalization and consistency of prediction w.r.t. perturbations;
- We expand the augmentation space, complementary to that of visual augmentations, by exploiting amplitude- and phase-adjustable frequency noise, and use it in an adversarial setting. Our method reduces the augmentation gap of common visual augmentations.

	APR-SP / AFA (ours) w/o aux.	AFA (ours)	AugMix†	AFA w/ AugMix	PRIME	AFA w/ PRIME	AugMax†
FLOPs	$\times 1$	$\times 2$	$\times 3$	$\times 2$	$\times 1$	$\times 2$	$\times 8$
Memory	$\times 1.02$	$\times 1.62$	$\times 2.66$	$\times 1.83$	$\times 2.50$	$\times 3.06$	$\times 2.35$

Table 1. AFA adds minimal computational burden to existing methods and is more efficient compared to other adversarial methods. It requires only $\times 1.62$ memory and just $\times 2$ the FLOPs of standard augmentation [12] training whereas AugMax uses $\times 2.35$ the memory and $\times 8$ the FLOPs when using 5 PGD steps. Methods with † denote the use of loss with JSD.

2. Related works

Data augmentation includes a set of techniques to increase data variety, thus reducing the distribution gap between training and test data. Generalization and robustness performance of models normally benefits from the use of data augmentation for training [45] or at test-time [19].

Image-based augmentations. Common image augmentation techniques include transformations, e.g. cropping, flipping, rotation, among others [45]. Applying the transformations with fixed configuration lacks flexibility when the models encounter more variations in the inputs at testing time. Thus, algorithms were designed to combine transformations randomly, e.g. AugMix [15], RandAug [3], TrivialAugment [34], MixUp [52], and CutMix [51]. However, random combinations might not be optimal. In [2], AutoAugment was proposed, based on using reinforcement learning to find the best policy on how to combine basic transformations for augmentation. AugMax [42] instead combines transformations adversarially, aiming at complementing augmentations based on diversity with others that favour hardness of training data. PRIME [33] samples transformations with maximum-entropy distributions. [40] augments images based on knowledge distilled by a teacher model. However, these approaches address variations limited by visually-plausible transformations only.

Frequency-based augmentations. In [49], it was discovered that models trained with visual transformations might be vulnerable to noise impacting certain parts of the frequency spectrum (e.g. high-frequency components), demonstrating that visual augmentations do not completely guarantee robustness. Complementary augmentation techniques are thus required to fill the augmentation gap left by visual augmentations. The straightforward approach is augmentation in the frequency domain. For example, [6] mixes the amplitude spectrum of images to reduce reliance on the amplitude part of the spectrum and induce phase-reliance for classification. [41, 48] swap or mix the amplitude spectrum of images. [43] augments images with shortcut features to reduce their specificity for classification, mitigating frequency shortcut learning. [37] introduces frequency noise in the AugMix framework. [24, 29] adversarially perturb images in the frequency domain. While these techniques address what visual augmentations may overlook, they also have limitations. Most frequency augmentation methods are based on manipulation of the frequency components of images. They usually have high computational requirements to identify frequency shortcuts [43] (f.i. using [44, 46]), implement adversarial training setup [24] or calculate multiple Fourier transforms of original and augmented images [6, 41, 43, 48].

We instead propose to use Fourier-basis functions as additive noise in the frequency domain. Our augmentation technique requires only one extra step during training rather than multiple pre-processing and expensive computations during training time as in other methods [6, 41, 43, 48], and works to complement image-based augmentations. Furthermore, we simplify the adversarial training framework of AugMax [42], not requiring an optimization process to maximize the hardness of adversarial augmentation, and achieving comparable or higher robustness. This allows the use of adversarial augmentations at larger-scale. We account for the induced distribution shifts in the frequency domain via an auxiliary component. The benefit of AFA is complementary to visual augmentations, and we can incorporate them seamlessly to further boost model robustness.

3. Preliminary: Fourier-basis functions

We utilize Fourier-basis functions in our augmentation strategy as an additive perturbation to the images. They are sinusoidal wave functions used as basic components of the Fourier transform to represent signals and images. A real Fourier basis function has two parameters, namely a frequency f and direction ω, and is denoted as:

$$A_{f,\omega}(u,v) = R\sin(2\pi f(u\cos(\omega)+v\sin(\omega)-\pi/4)), \quad (1)$$

where $A_{f,\omega}(u,v)$ represents the amplitude of the wave at position (u,v). The function involves the sine of a 2D spatial frequency $2\pi f$ to produce a planar wave with a specific frequency f, and angle ω that indicates the direction of propagation. R is chosen such that the planar wave has unit l_2-norm. A particular Fourier basis function, characterized by specific frequency (f) and direction (ω), can be associated with a Dirac delta function in the spectral domain. Therefore, when employed in an additive manner, as in our augmentation strategy, this Fourier-basis function facilitates the targeted modification of particular frequency components of images. Examples of Fourier-basis waves superimposed on images are shown in Fig. 2.

4. Auxiliary Fourier-basis Augmentation

The Auxiliary Fourier-basis Augmentation (AFA) that we propose is based on two lines of augmentations, one considered in-distribution (using visual augmentations) and an-

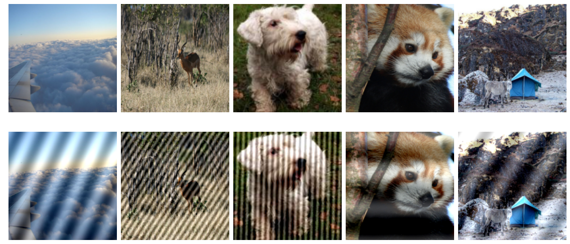

Figure 2. Example of Fourier-basis functions added to natural images. They appear as *gratings* that obscure spatial information.

other considered out-of-distribution or adversarial (using frequency-based noise) as shown in Fig. 3. We generate the adversarial augmented images by sampling a Fourier-basis and a strength parameter per colour channel, and adding them to the original images. Visually augmented and adversarially augmented training images are then processed using a main component and an auxiliary component, respectively. Joint optimisation of two cross-entropy functions encourages robust and consistent classification, as it promotes correctness under adversarially augmented images. Details of the different parts of the method are reported below.

Generation of adversarial augmented images. Randomly sampling augmentations and applying them to images with random strengths was shown to be sufficient to outperform more complex strategies [34].

We follow this design principle in our method to generate adversarial augmented images with Fourier basis functions, which allows us to avoid optimization steps to determine the worst-case combination of augmentations as in AugMax [42]. We produce adversarial augmented images by adding a different Fourier basis function $A_{f,\omega}$ per channel of the original RGB image. We generate the Fourier basis functions by sampling f and ω from uniform distributions as $f \sim \mathcal{U}_{[1,\mathrm{M}]}$ and $\omega \sim \mathcal{U}_{[0,\pi]}$, where M is the image size. The sampling space of all Fourier-basis is denoted as $\mathcal{V}$. We add the generated Fourier basis functions per channel c with a weight factor sampled from an exponential distribution $\sigma_c \sim \mathrm{Exp}(1/\lambda)$, with $c \in \{\mathrm{R},\mathrm{G},\mathrm{B}\}$. The selection of the exponential distribution for sampling augmentation magnitude is motivated by the concept of event rate, where perturbations with larger magnitudes become progressively less likely, albeit still possible. This is controlled by adjusting λ, ensuring a balance between maintaining diversity in sampled values while minimizing the occurrence of extremely large augmentation perturbations. In Sec. 5.3, we show how the parameter λ affects the augmentation results.

The proposed augmentation process results in a 3-channel image $x^a = [x^a_R, x^a_G, x^a_B]$, where:

$$x^a_c = \mathrm{Clamp}_{[0,1]}(x_c + \sigma_c A_{f_c,\omega_c}), \quad c \in \{\mathrm{R},\mathrm{G},\mathrm{B}\}. \quad (2)$$

An example of image x^a augmented with additive Fourier-basis functions is shown in our method schema in Fig. 3. We

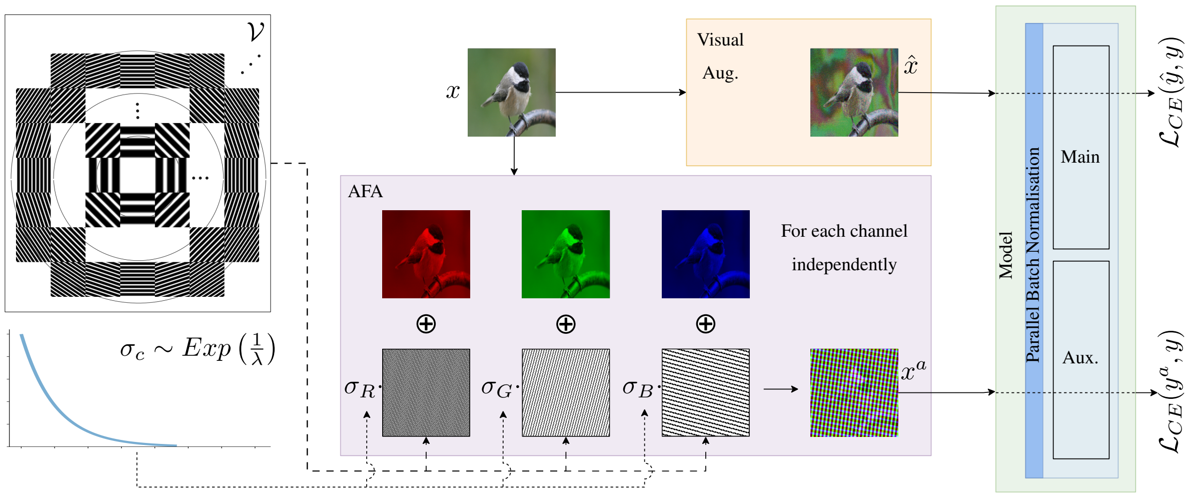

Figure 3. Schema of the AFA augmentation pipeline. The image x is augmented using AFA, which adds a planar wave per channel c of the image at a strength value σ_c sampled from an exponential distribution (eq.2). The AFA augmented image x^a is used for training, processed through the auxiliary component of the parallel batch normalisation layer (for models that use batch normalization to track batch statistics, e.g. ResNet). Other visual augmentations are applied in parallel, and used for training via the main component of the normalization layer. Finally, we train via optimizing two cross-entropy losses, one for the main and the other for the auxiliary component.

demonstrate the adversarial nature of augmented samples in the supplementary material.

Auxiliary component for distribution shifts. As shown in Figs. 2 and 3, the Fourier-basis augmentations result in images with an unnatural appearance due to substantial frequency perturbations. The presence of planar waves across the augmented images determines the *unnaturalness* of image appearance, which can be seen as adversarial attacks on the images. These augmentations disrupt the learned mean and variance in batch normalization layers, which are inconsistent with the distribution shifts induced by our augmentation and lead to inconsistent activations. This results in a negative impact on model convergence and generalization abilities.

We address these issues by deploying architectural components in the training, capable of handling distribution shifts explicitly by tracking statistics and adjusting the loss function accordingly. Namely, we incorporate auxiliary components into the model, such as Parallel Batch Normalization layers and an additional cross-entropy term in the loss function to specifically account for these adversarial augmented images. These modifications to the model architecture and training enhance performance, particularly in the presence of distribution shifts, contributing to better generalization, robustness to common corruptions and consistency to time-dependent increasing perturbations. The introduction of parallel batch normalization layers is motivated by the need to account for distribution shifts induced by adversarial (Fourier-basis) augmentations, as observed in [42]. With the parallel batch normalisation, the affine parameters and statistics of main and auxiliary distributions are recorded separately. This allows independent learning of distribution of the visually and adversarially augmented images. Without these additional normalization layers, the model training assumes a single-modal sample distribution, limiting its ability to differentiate between the main and the adversarial distribution, thus negatively affecting overall performance. In Sec. 5.3, we show the result of not employing the auxiliary components.

It is worth noting that for models that do not employ batch normalization layers (e.g. CCT that uses layer normalization and does not track statistics), the parallel normalization layers are not needed. However, the extra term in the loss function (see next paragraph) to generate consistent predictions across distribution shifts serves as a regularization mechanism, which is verified in the supplementary material.

Loss function. We work in the supervised learning setting with a training dataset $\mathcal{D}$ consisting of clean images x with labels y. We train the model in the main architecture stream (see Fig. 3) using a cross-entropy loss $\mathcal{L}_{\text{CE}}(\hat{y}, y)$, where y is the ground-truth label and $\hat{y}$ is the predicted label for images augmented with a given visual augmentation strategy (e.g. standard, PRIME, etc.). Under the non-auxiliary setting, models thus optimise the standard cross entropy loss.

In the auxiliary setting, we add an extra cross-entropy loss term $\mathcal{L}_{\text{CE}}(y^a, y)$, which optimise the model to predict the correct label on adversarial augmented images whose predicted label is denoted by y^a, contributing to robustness of the model w.r.t. aggressive distribution shifts. We refer to the combined loss function $\mathcal{L}_{\text{ACE}}$, taking the average of the two cross-entropy terms, as the Auxiliary Cross Entropy

(ACE) Loss:

$$\mathcal{L}_{\text{ACE}}(\hat{y}, y^a, y) = \frac{1}{2}\left[\mathcal{L}_{\text{CE}}(\hat{y}, y) + \mathcal{L}_{\text{CE}}(y^a, y)\right]. \quad (3)$$

It contributes to achieve comparable performance, with lower training time and complexity, than using the Jensen-Shannon Divergence (JSD) loss [15, 42]. Our motivation to not employ the JSD loss is the reduced training time due to less computational complexity. In our experiments, for comparison purposes, we also use the JSD loss in the auxiliary setting, where training batches are augmented using AFA and go through auxiliary components. We report results in Sec. 5.3 (Fig. 6).

5. Experiments and results

We compare AFA with other popular augmentation techniques, evaluating robustness to common corruptions, generalization abilities and consistency to time-dependent increasing perturbations, on benchmark datasets.

5.1. Experiment setup

Datasets. We trained models on the CIFAR-10 (C10) [20], CIFAR-100 (C100) [21], TinyImageNet (TIN) [22] and ImageNet (IN) [4] datasets and evaluate them on the corresponding robustness benchmark datasets, namely C10-C, C100-C, TIN-C, IN-C [14], IN-$\bar{\text{C}}$ [32], and IN-3DCC [18]. For ImageNet-trained models, we further evaluate their generalisation performance on the IN-v2 [35] and IN-R datasets [13], and consistency of performance on time-dependent increasing perturbations on the IN-P dataset [14].
Architectures and training details. We train ResNet [12] and transformers (CCT [7], CVT [7] and ViT [5]). We train ResNet-18, CCT-7/3x1 (32 resolution), CVT and ViT-Lite on C-10, C-100, and only ResNet-18 on TIN. In the case of ImageNet, we train ResNet-18, ResNet-50 and CCT-14/7x2 (224 resolution). Under auxiliary setting, we use the DuBIN variant of ResNet [42]. We always use standard transforms [12] before other augmentations. Implementation details and hyperparameter configurations are in the supplementary material.
Evaluation. We evaluate the classification accuracy on the original test set, which we refer to as standard accuracy (SA), and the average classification accuracy over all corruptions in the robustness benchmarks as robustness accuracy (RA). This provides direct comparison between model performance on original and corruption benchmark datasets. We also compute the mean corruption error (mCE) [14] for TIN and IN (for CIFAR there are no baselines advised) to evaluate the normalized robustness of models against image corruptions, the mean flip rate (mFR) and the mean top-5 distance (mT5D) to evaluate the consistency performance of models against increasing perturbations. For the evaluation of generalization performance, we compute the accuracy on the ImageNet-R and ImageNet-v2 test sets (note that ImageNet-v2 has 3 test sets, and we report the average accuracy on them). We only use the main BN layers during testing, similar to AugMax. More details about the metrics are in the supplementary material.

5.2. Results

Comparison with AugMax. We first report a direct comparison with AugMax [42] in Tab. 2, as AFA addresses the computational shortcomings of generating adversarial augmentations via PGD iterations, and of using a JSD loss for alignment of the distribution of original and (adversarially) augmented images. We use AugMix as main augmentation, as in AugMax, and ablate on the use of JSD and ACE loss.

We show that AFA achieves comparable (or better) performance than AugMax, despite it being much less computational intensive. We indeed demonstrate that we can generate adversarial augmentations by only adding (weighted) Fourier-basis waves per color channel, not requiring PGD steps, and can train the models using an extra cross-entropy instead of the expensive JSD loss. The improvements granted by our approach are particularly evident in the case of ImageNet (using ACE), where we gain 1.6% of standard accuracy and 4.1% of robust accuracy (5.6% mCE) performance w.r.t. AugMax. Considering the increased computational efficiency and the simplicity of adversarial augmentation method, AFA is a more versatile and effective tool than AugMax. Hence, in the rest of the paper, we do not report further results of the AugMax framework, due to its high computational requirements, which complicate the training of larger models (e.g. ResNet-50 and CCT).
Robustness, generalization and consistency. In Tab. 3, we

-	Main	Auxiliary	SA↑	RA↑	mCE↓
C10	AugMix†	✗	95.47	86.48	-
	AugMix†	AugMax	95.76	**90.36**	-
	AugMix†	AFA	95.24	89.96	-
	AugMix	AFA	95.44	89.81	-
C100	AugMix†	✗	78.72	61.61	-
	AugMix†	AugMax	78.69	65.75	-
	AugMix†	AFA	78.99	65.96	-
	AugMix	AFA	77.80	**66.69**	-
TIN	AugMix†	✗	64.65	36.30	83.90
	AugMix†	AugMax	62.21	**38.67**	**80.72**
	AugMix†	AFA	64.34	38.53	**80.79**
	AugMix	AFA	62.51	**38.67**	80.83
IN	AugMix†	✗	65.2	31.5	87.1
	AugMix†	AugMax	66.5	36.5	80.6
	AugMix†	AFA	65.0	36.8	80.4
	AugMix	AFA	68.1	**41.1**	**75.0**

Table 2. Comparison of AFA and AugMax (with AugMix for visual augmentation [42]), with a ResNet18 backbone. The mark † indicates the use of the JSD loss, otherwise the ACE loss is used.

				Robustness						Generalisation		Consistency	
				IN-C		IN-C̄		IN-3DCC		IN-R	IN-v2	IN-P	
	Main	**Aux**	SA (↑)	RA (↑)	mCE (↓)	RA (↑)	mCE (↓)	RA (↑)	mCE (↓)	Acc. (↑)	Avg. Acc. (↑)	mFP (↓)	mT5D (↓)
ResNet18	-	✗	**68.9**	32.9	84.7	34.8	87.0	34.9	84.4	33.1	**64.3**	72.8	87.0
	-	AFA	68.2	35.9	81.0	41.7	78.3	37.1	81.7	32.8	63.7	64.2	76.8
	AugMix†	✗	65.2	31.5	87.1	34.6	87.3	32.1	88.3	28.2	59.5	80.2	86.2
	AugMix†	AFA	65.0	36.8	80.4	40.9	79.3	36.0	83.2	30.6	60.9	60.1	68.5
	AugMix	AFA	68.1	41.1	75.0	45.2	73.3	38.9	79.4	35.2	63.2	68.5	81.7
	PRIME	✗	66.0	43.6	72.0	42.0	78.1	42.4	75.2	36.9	61.4	54.7	65.3
	PRIME	AFA	67.2	**47.2**	**67.8**	**47.3**	**71.1**	**43.8**	**73.5**	**37.8**	63.0	**52.3**	**63.7**
	TA+	✗	**68.9**	36.9	80.1	35.9	85.6	38.6	79.7	32.6	63.7	68.1	81.4
	TA+	AFA	67.8	41.4	74.7	42.9	76.7	41.1	76.5	35.4	62.7	59.9	72.3
ResNet50	-	✗	75.6	39.2	76.7	39.9	79.4	41.2	76.1	36.2	70.8	58.0	78.4
	-	AFA	76.5	46.2	68.0	47.6	69.4	46.2	69.8	38.1	72.0	48.0	67.2
	APR-SP	✗	71.9	42.9	72.7	45.9	72.5	39.8	78.4	34.9	67.2	60.2	75.4
	APR-SP	AFA	74.4	47.6	66.7	51.4	64.9	42.6	74.6	38.7	69.3	54.9	72.6
	AugMix†	✗	74.7	43.4	72.0	44.6	73.3	41.9	75.5	33.0	70.0	60.9	72.5
	AugMix†	AFA	75.6	50.6	62.9	51.8	64.0	47.6	68.3	36.3	71.2	44.5	56.1
	AugMix	AFA	76.6	49.1	64.7	52.5	62.9	46.3	69.6	41.0	71.8	52.2	72.2
	PRIME	✗	72.1	49.2	64.9	46.4	71.5	47.2	68.8	38.5	67.8	45.4	58.1
	PRIME	AFA	74.5	**53.9**	**59.2**	**54.2**	**61.3**	**50.2**	**65.0**	**40.9**	69.8	**40.4**	**54.8**
	TA+	✗	75.9	43.4	71.7	41.8	77.1	44.7	71.6	37.1	70.3	51.9	70.4
	TA+	AFA	76.6	50.3	63.1	49.7	66.7	49.6	65.4	40.0	**72.2**	45.1	64.5
CCT	-	✗	76.4	43.9	70.7	50.3	65.6	43.4	73.2	35.6	71.2	48.3	72.9
	-	AFA	76.9	51.9	61.0	58.5	55.4	50.7	64.4	39.0	71.9	38.4	61.8
	AugMix	✗	76.1	47.3	66.8	52.2	63.1	45.3	71.0	37.9	70.7	49.3	72.8
	AugMix	AFA	**77.4**	56.5	55.6	60.8	52.2	51.8	62.8	41.0	**72.5**	37.9	59.9
	PRIME	✗	73.6	54.1	58.6	54.5	60.8	50.7	64.4	39.2	68.7	36.1	53.0
	PRIME	AFA	76.6	**58.7**	**52.8**	**61.2**	**52.0**	**54.5**	**59.4**	**43.2**	71.9	**31.9**	**51.2**
	TA+	✗	77.1	50.2	63.2	54.1	60.7	49.3	65.8	38.2	72.1	41.8	66.3
	TA+	AFA	76.9	56.0	56.0	59.1	54.6	53.1	61.1	41.1	72.1	36.4	58.5

Table 3. Robustness, generalization and consistency results on ImageNet-based benchmarks. Models with † use the JSD loss. TrivialAugment (TA) has overlapping augmentations with IN-C (+), and no other overlaps with other datasets. The green colour indicates an improvement when the main augmentation is combined with AFA, while red indicates no improvement. Results marked with **bold**/**bold** are the best for a particular architecture.

report results achieved by AFA combined with different visual augmentation methods, AugMix, PRIME, TrivialAugment (TA), to train different architectures (ResNet, CCT). We evaluate robustness to common corruptions on IN-C, IN-C̄ and IN-3DCC, OOD generalisation on IN-v2 and IN-R, and consistency w.r.t. increasing perturbations on IN-P.

AFA generally contributes to a boost of performance (green colored results in Tab. 3) when combined with different visual augmentation techniques, reducing the robustness and generalization gap for different model architectures. When compared to another Fourier-based augmentation technique, APR-SP [6], AFA outperforms it on all benchmarks when trained with only standard augmentation techniques. When models trained with AugMix and AFA, we record better overall performance than those trained with AugMix alone. For the transformer architecture CCT, training with AFA contributes to an even stronger improvement in all tests. These results stay consistent for smaller resolution datasets (CIFAR and TIN), as we report at the end of this section.

Robustness to high-severity corruptions. AFA contributes to a consistent improvement of robustness of models at increasing corruption severity. We compute the relative corruption error, namely the difference between the corruption error of models trained with a visual augmentation technique only and those trained with both visual augmentations and AFA, and report it in Fig. 4 for different corruption severity. A positive value indicates that models trained with the addition of AFA have better robustness. For higher corruption severity, AFA contributes to stronger robustness, measured by an increase in the relative corruption error in Fig. 4. The improvements obtained by AFA on IN-3DCC are slightly less pronounced than those on IN-C and IN-C̄. This is attributable to the specific corruptions in IN-3DCC that concern 3D geometric information, and are somewhat more complicated image transformations. However, AFA contributes to a substantial improvement w.r.t. to models trained without it. We thus highlight that AFA is very beneficial for increasing robustness to aggressive corruptions of the test images. Details of the results at different

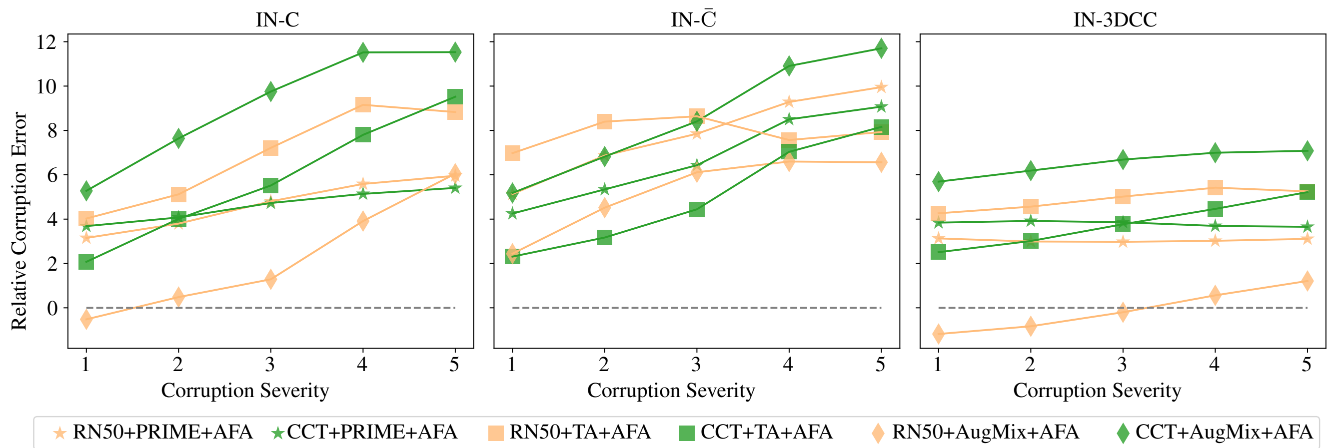

Figure 4. Relative error per corruption severity, computed by subtracting the classification error of models trained with PRIME, TrivialAugment, and AugMix with that of corresponding models trained with PRIME+AFA, TrivialAugment+AFA, and AugMix+AFA.

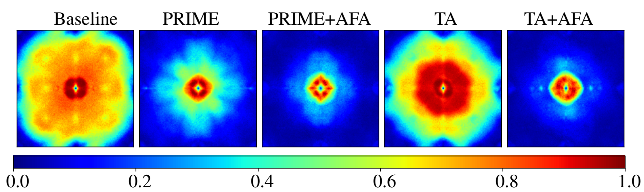

Figure 5. Fourier heatmaps of ResNet18 trained with standard setup, and PRIME and TrivialAugment, with and without AFA.

severity are in the supplementary material.

Fourier heatmap: robustness in the frequency spectrum. We further evaluate the robustness of models to perturbations at specific frequencies, using test images perturbed with frequency noises according to [49]. We present the results in the form of Fourier heatmaps, see Fig. 5 for heatmaps of ResNet18 models (trained on ImageNet), and the supplementary material for the heatmaps of CCT models. The intensity of a pixel at location (u, v) in the heatmap indicates the classification error of a model tested on images perturbed by Fourier noise at frequency (u, v) in the frequency spectrum (implementation details are in the supplementary material). ResNet18 trained with standard augmentations setting (baseline) is very sensitive to perturbations at low and middle-high frequency (see Fig. 5), while those trained with visual augmentations like PRIME and TrivialAugment (TA) still show vulnerability at low and middle-high frequency noise. When training models with AFA, i.e. PRIME+AFA and TA+AFA, the models become more robust to frequency pertubations, especially at middle-high frequency. AFA can provide extensive robustness to frequency perturbations and bridge the robustness gap that visual augmentation might not cover.

Results on CIFAR and TIN. In Tab. 4, we present the robustness results on smaller resolution datasets, C10 and C100. The results on TIN are in the supplementary material. These results are inline with those on IN in Tab. 3.

-	Main	Auxiliary	C10-C SA↑	C10-C RA↑	C100-C SA↑	C100-C RA↑
ResNet18	-	✗	94.15	73.67	78.27	48.30
	-	AFA	94.69	88.22	77.91	62.53
	AugMix†	✗	**95.47**	86.48	78.72	61.61
	AugMix†	AFA	95.24	89.96	**78.99**	65.96
	PRIME	✗	94.38	89.81	75.49	66.16
	PRIME	AFA	94.54	**90.64**	76.16	**68.48**
CCT	-	✗	95.67	80.45	**78.37**	54.20
	-	AFA	**95.94**	88.13	77.47	61.40
	AugMix	✗	95.10	85.42	75.79	60.83
	AugMix	AFA	**95.93**	90.57	77.22	66.18
	PRIME	✗	95.30	90.56	76.65	**67.92**
	PRIME	AFA	95.49	**91.40**	76.50	67.89
CVT	-	✗	94.31	77.02	75.53	48.25
	-	AFA	**94.53**	**87.03**	**76.96**	**60.12**
ViT	-	✗	94.46	75.97	74.26	50.88
	-	AFA	**94.58**	**86.71**	**75.13**	**58.25**

Table 4. Results for C10-C and C100-C with ResNet18, CCT. CVT and ViT-Lite. Models with † use loss with JSD.

5.3. Ablation

Auxiliary components. We investigate the contribution and importance of the auxiliary components in improving model robustness. We trained models with AFA-augmented images, passing through only the main components or the auxiliary components. The results in Tab. 5, i.e. lower RA and higher mCE of models trained with AFA applied only in the main components, highlight the importance of AFA auxiliary components. The auxiliary components play a crucial role in mitigating the impact of aggressive adversarial distribution shifts induced by AFA. By doing so, they contribute to model ability to learn from the original distribution, while AFA facilitates learning robustness to dis-

-	Main	Auxiliary	SA↑	RA↑	mCE↓
C10	-	✗	94.15	73.67	-
	AFA	✗	92.36	83.25	-
	-	AFA	**94.69**	**88.22**	-
C100	-	✗	78.27	48.30	-
	AFA	✗	72.34	58.70	-
	-	AFA	**77.91**	**62.53**	-
TIN	-	✗	61.64	23.91	100.00
	AFA	✗	59.04	28.87	93.45
	-	AFA	**62.52**	**33.35**	**87.58**
IN	-	✗	68.9	32.9	84.7
	AFA	✗	66.7	33.3	84.4
	-	AFA	**68.2**	**35.9**	**81.0**

Table 5. Ablation results ResNet18 trained with and without Auxiliary Components on C10, C100, TinyImageNet and ImageNet.

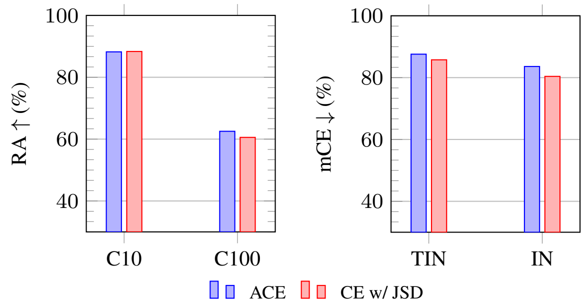

Figure 6. Comparison of using objective with and without the JSD term. All models are ResNet-18 trained with only AFA in the auxiliary component and no other augmentations. When used with JSD two batches passed through Auxiliary components and there was no main augmentation (in total 3 batches, 1 clean and 2 AFA).

tribution shifts. This is also highlighted in the substantial decrease in SA for models not employing auxiliary components. While model robustness improves under both settings, the performance gain for the auxiliary setting is three to five percentage points higher across all datasets.

ACE vs JSD. As part of our method, we replaced the use of JSD with ACE which is less computationally burdening. We thus performed an ablation analysis of the tradeoff of using JSD. We report results for robustness using mCE and Robust Accuracy (RA) in Fig. 6, and observe that JSD does not significantly improve the robustness of our model to image corruptions, despite it being more computationally heavy than using ACE. Using JSD also results in slightly worse robustness on C100. Given the minimal differences, we opt for the simpler ACE loss for training with the AFA augmentation pipeline and only using JSD if other techniques (e.g. AugMix) employ them.

Effect of hyperparameter $1/\lambda$. We studied also the contribution of the mean $1/\lambda$ of the exponential distribution that we use to sample the weight factor for the channel-wise application of the Fourier-basis augmentations. We provide the results in Fig. 7, and observe that our method has low sensitivity to the choice of the rate parameter. This is attributable to the choice of the exponential distribution that allows larger values to be sampled even if they are less likely. We indeed observe that larger values of $1/\lambda$, which result in larger perturbations (in the range of 10 to 15), result in stronger gains in robustness. At the same time, there is no clear trend in the standard accuracy on the clean dataset, with only minimal variations for the larger values, indicating that the choice of the $1/\lambda$ value does not have a specific influence on the correct functioning of AFA.

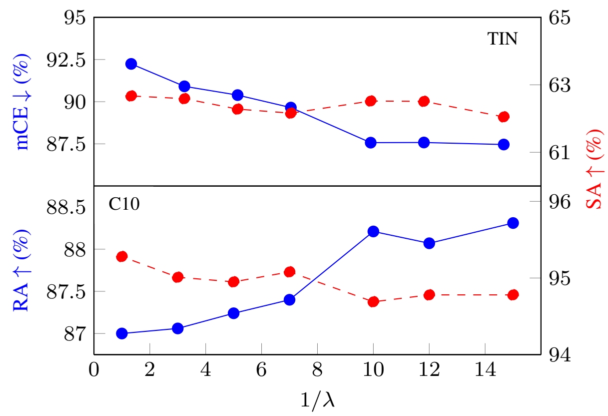

Figure 7. Trend of the mCE and SA with respect to the rate parameter. The models were trained using AFA in the auxiliary setting and no other augmentations for the main.

6. Conclusions

We proposed an efficient data augmentation technique called AFA, which complements existing visual augmentation techniques by filling the augmentation gap, that they do not cover in the Fourier domain. AFA perturbs the frequency components of images and generates adversarial samples. By leveraging Fourier-basis functions and the auxiliary augmentation setting we demonstrate that AFA allows the models to learn from aggressive/adversarial input changes. We performed extensive experiments on benchmark datasets, and demonstrated that AFA benefits the robustness of models against common image corruptions, the consistency of predictions when facing increasing perturbations, and the OOD generalization performance. Being complementary to other augmentation techniques, AFA can further boost the robustness of models, especially against strong corruptions and perturbation, and it also results in better robustness in the frequency spectrum. We foresee that investigating the use of Fourier-basis functions on the training process of neural networks would provide promising improvement to model performance, thus encouraging their reliability in real scenarios.

References

[1] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations, 2020. 1

[2] Ekin D. Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V. Le. Autoaugment: Learning augmentation policies from data, 2019. 1, 2

[3] Ekin Dogus Cubuk, Barret Zoph, Jon Shlens, and Quoc Le. Randaugment: Practical automated data augmentation with a reduced search space. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 18613–18624. Curran Associates, Inc., 2020. 2

[4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. 5

[5] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. 5

[6] Chen *et al*. Amplitude-phase recombination: Rethinking robustness of convolutional neural networks in frequency domain, 2021. 2, 3, 6

[7] Hassani *et al*. Escaping the big data paradigm with compact transformers, 2022. 5

[8] Fartash Faghri, Hadi Pouransari, Sachin Mehta, Mehrdad Farajtabar, Ali Farhadi, Mohammad Rastegari, and Oncel Tuzel. Reinforce data, multiply impact: Improved model accuracy and robustness with dataset reinforcement, 2023. 1

[9] Zhiqiang Gao, Kaizhu Huang, Rui Zhang, Dawei Liu, and Jieming Ma. Towards better robustness against common corruptions for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 18882–18893, October 2023. 1

[10] Antonio Greco, Nicola Strisciuglio, Mario Vento, and Vincenzo Vigilante. Benchmarking deep networks for facial emotion recognition in the wild. *Multimedia Tools and Applications*, 82(8):11189–11220, 2023. 1

[11] Xiaoshuai Hao, Yi Zhu, Srikar Appalaraju, Aston Zhang, Wanqian Zhang, Bo Li, and Mu Li. Mixgen: A new multi-modal data augmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*, pages 379–389, January 2023. 1

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 2, 5

[13] Dan Hendrycks, Steven Basart, Norman Mu, Saurav Kadavath, Frank Wang, Evan Dorundo, Rahul Desai, Tyler Zhu, Samyak Parajuli, Mike Guo, Dawn Song, Jacob Steinhardt, and Justin Gilmer. The many faces of robustness: A critical analysis of out-of-distribution generalization, 2021. 1, 5

[14] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations, 2019. 1, 5

[15] Dan Hendrycks, Norman Mu, Ekin D. Cubuk, Barret Zoph, Justin Gilmer, and Balaji Lakshminarayanan. AugMix: A Simple Data Processing Method to Improve Robustness and Uncertainty. *arXiv*, Dec. 2019. 1, 2, 5

[16] Ignacio Hounie, Luiz F. O. Chamon, and Alejandro Ribeiro. Automatic data augmentation via invariance-constrained learning. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 13410–13433. PMLR, 23–29 Jul 2023. 1

[17] Christoph Kamann and Carsten Rother. Benchmarking the robustness of semantic segmentation models with respect to common corruptions. *International Journal of Computer Vision*, 129(2):462–483, Feb 2021. 1

[18] Oğuzhan Fatih Kar, Teresa Yeo, Andrei Atanov, and Amir Zamir. 3d common corruptions and data augmentation, 2022. 1, 5

[19] Ildoo Kim, Younghoon Kim, and Sungwoong Kim. Learning loss for test-time augmentation. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 4163–4174. Curran Associates, Inc., 2020. 2

[20] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. Cifar-10 (canadian institute for advanced research). 5

[21] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. Cifar-100 (canadian institute for advanced research). 5

[22] Ya Le and Xuan S. Yang. Tiny imagenet visual recognition challenge. 2015. 5

[23] Xiu-Chuan Li, Xu-Yao Zhang, Fei Yin, and Cheng-Lin Liu. F-mixup: Attack cnns from fourier perspective. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 541–548, 2021. 1

[24] Chang Liu, Wenzhao Xiang, Yuan He, Hui Xue, Shibao Zheng, and Hang Su. Improving model generalization by on-manifold adversarial augmentation in the frequency domain, 2023. 2, 3

[25] Jiashuo Liu, Zheyan Shen, Yue He, Xingxuan Zhang, Renzhe Xu, Han Yu, and Peng Cui. Towards out-of-distribution generalization: A survey, 2023. 1

[26] Siao Liu, Zhaoyu Chen, Yang Liu, Yuzheng Wang, Dingkang Yang, Zhile Zhao, Ziqing Zhou, Xie Yi, Wei Li, Wenqiang Zhang, and Zhongxue Gan. Improving generalization in visual reinforcement learning via conflict-aware gradient agreement augmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 23436–23446, October 2023. 1

[27] Yang Liu, Shen Yan, Laura Leal-Taixé, James Hays, and Deva Ramanan. Soft augmentation for image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16241–16250, June 2023. 1

[28] Yuyang Long, Qilong Zhang, Boheng Zeng, Lianli Gao, Xianglong Liu, Jian Zhang, and Jingkuan Song. Frequency domain model augmentation for adversarial attack. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, editors, *Computer Vision – ECCV 2022*, pages 549–566, Cham, 2022. Springer Nature Switzerland. 2

[29] Yuyang Long, Qilong Zhang, Boheng Zeng, Lianli Gao, Xi-

anglong Liu, Jian Zhang, and Jingkuan Song. Frequency domain model augmentation for adversarial attack. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, editors, *Computer Vision – ECCV 2022*, pages 549–566, Cham, 2022. Springer Nature Switzerland. 2

[30] Guozheng Ma, Linrui Zhang, Haoyu Wang, Lu Li, Zilin Wang, Zhen Wang, Li Shen, Xueqian Wang, and Dacheng Tao. Learning better with less: Effective augmentation for sample-efficient visual reinforcement learning, 2023. 1

[31] Juliette Marrie, Michael Arbel, Diane Larlus, and Julien Mairal. Slack: Stable learning of augmentations with cold-start and kl regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 24306–24314, June 2023. 1

[32] Eric Mintun, Alexander Kirillov, and Saining Xie. On interaction between augmentations and corruptions in natural corruption robustness, 2021. 1, 5

[33] Apostolos Modas, Rahul Rade, Guillermo Ortiz-Jiménez, Seyed-Mohsen Moosavi-Dezfooli, and Pascal Frossard. PRIME: A few primitives can boost robustness to common corruptions. *arXiv*, Dec. 2021. 1, 2

[34] Samuel G. Müller and Frank Hutter. TrivialAugment: Tuning-free Yet State-of-the-Art Data Augmentation. *arXiv*, Mar. 2021. 1, 2, 3

[35] Benjamin Recht, Rebecca Roelofs, Ludwig Schmidt, and Vaishaal Shankar. Do imagenet classifiers generalize to imagenet?, 2019. 1, 5

[36] Tonmoy Saikia, Cordelia Schmid, and Thomas Brox. Improving robustness against common corruptions with frequency biased models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10211–10220, October 2021. 2

[37] Ryan Soklaski, Michael Yee, and Theodoros Tsiligkaridis. Fourier-Based Augmentations for Improved Robustness and Uncertainty Calibration. *arXiv*, Feb. 2022. 2

[38] Nicola Strisciuglio and George Azzopardi. Visual response inhibition for increased robustness of convolutional networks to distribution shifts. In *NeurIPS 2022 Workshop on Distribution Shifts: Connecting Methods and Applications*, 2022. 1

[39] Nicola Strisciuglio, Manuel Lopez-Antequera, and Nicolai Petkov. Enhanced robustness of convolutional networks with a push–pull inhibition layer. *Neural Computing and Applications*, 32(24):17957–17971, 2020. 1

[40] Teppei Suzuki. Teachaugment: Data augmentation optimization using teacher knowledge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10904–10914, June 2022. 2

[41] An Wang, Mobarakol Islam, Mengya Xu, and Hongliang Ren. Curriculum-based augmented fourier domain adaptation for robust medical image segmentation. *IEEE Transactions on Automation Science and Engineering*, pages 1–13, 2023. 2, 3

[42] Haotao Wang, Chaowei Xiao, Jean Kossaifi, Zhiding Yu, Anima Anandkumar, and Zhangyang Wang. AugMax: Adversarial Composition of Random Augmentations for Robust Training. *arXiv*, Oct. 2021. 1, 2, 3, 4, 5

[43] Shunxin Wang, Christoph Brune, Raymond Veldhuis, and Nicola Strisciuglio. DFM-x: Augmentation by leveraging prior knowledge of shortcut learning. In *4th Visual Inductive Priors for Data-Efficient Deep Learning Workshop*, 2023. 2, 3

[44] Shunxin Wang, Raymond Veldhuis, Christoph Brune, and Nicola Strisciuglio. Frequency shortcut learning in neural networks. In *NeurIPS 2022 Workshop on Distribution Shifts: Connecting Methods and Applications*, 2022. 3

[45] Shunxin Wang, Raymond Veldhuis, Christoph Brune, and Nicola Strisciuglio. A Survey on the Robustness of Computer Vision Models against Common Corruptions. *arXiv*, May 2023. 1, 2

[46] Shunxin Wang, Raymond Veldhuis, Christoph Brune, and Nicola Strisciuglio. What do neural networks learn in image classification? a frequency shortcut perspective. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1433–1442, October 2023. 3

[47] Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V. Le. Self-training with noisy student improves imagenet classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1

[48] Qinwei Xu, Ruipeng Zhang, Ziqing Fan, Yanfeng Wang, Yi-Yan Wu, and Ya Zhang. Fourier-based augmentation with applications to domain generalization. *Pattern Recognition*, 139:109474, 2023. 2, 3

[49] Dong Yin, Raphael Gontijo Lopes, Jonathon Shlens, Ekin D. Cubuk, and Justin Gilmer. A Fourier Perspective on Model Robustness in Computer Vision. *arXiv*, June 2019. 1, 2, 7

[50] Mehmet Kerim Yucel, Ramazan Gokberk Cinbis, and Pinar Duygulu. Hybridaugment++: Unified frequency spectra perturbations for model robustness, 2023. 1

[51] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features, 2019. 2

[52] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations*, 2018. 2

[53] Stephan Zheng, Yang Song, Thomas Leung, and Ian Goodfellow. Improving the Robustness of Deep Neural Networks via Stability Training. *arXiv*, Apr. 2016. 1