A Fourier Perspective on Model Robustness in Computer Vision

Dong Yin[*]
Department of EECS
UC Berkeley
Berkeley, CA 94720
dongyin@berkeley.edu

Raphael Gontijo Lopes[†]
Google Research, Brain team
Mountain View, CA 94043
iraphael@google.com

Jonathon Shlens
Google Research, Brain team
Mountain View, CA 94043
shlens@google.com

Ekin D. Cubuk
Google Research, Brain team
Mountain View, CA 94043
cubuk@google.com

Justin Gilmer
Google Research, Brain team
Mountain View, CA 94043
gilmer@google.com

Abstract

Achieving robustness to distributional shift is a longstanding and challenging goal of computer vision. Data augmentation is a commonly used approach for improving robustness, however robustness gains are typically not uniform across corruption types. Indeed increasing performance in the presence of random noise is often met with reduced performance on other corruptions such as contrast change. Understanding when and why these sorts of trade-offs occur is a crucial step towards mitigating them. Towards this end, we investigate recently observed trade-offs caused by Gaussian data augmentation and adversarial training. We find that both methods improve robustness to corruptions that are concentrated in the high frequency domain while reducing robustness to corruptions that are concentrated in the low frequency domain. This suggests that one way to mitigate these trade-offs via data augmentation is to use a more diverse set of augmentations. Towards this end we observe that AutoAugment [5], a recently proposed data augmentation policy optimized for clean accuracy, achieves state-of-the-art robustness on the CIFAR-10-C and ImageNet-C benchmarks.

1 Introduction

Although many deep learning computer vision models achieve remarkable performance on many standard i.i.d benchmarks, these models lack the robustness of the human vision system when the train and test distributions differ [23]. For example, it has been observed that commonly occurring image corruptions, such as random noise, contrast change, and blurring, can lead to significant performance degradation [7, 2]. Improving distributional robustness is an important step towards safely deploying models in complex, real-world settings.

Data augmentation is a natural and sometimes effective approach to learning robust models. Examples of data augmentation include adversarial training [13], applying image transformations to the training data, such as flipping, cropping, adding random noise, and even stylized image transformation [10].

However, data augmentation rarely improves robustness across all corruption types. Performance gains on some corruptions may be met with dramatic reduction on others. As an example, in [9] it

[*]Work done while internship at Google Research, Brain team.

[†]Work done as a member of the Google AI Residency program g.co/airesidency.

Preprint. Under review.