A Fourier Perspective on Model Robustness in Computer Vision

Krishna Srikar Durbha

21st October 2020

Table of Contents

1. Introduction
2. Data Augmentation
3. Gaussian Data Augmentation and Adversarial Training
4. 2D-Gaussian Distribution
5. Hypothesis and Solution
6. Investigation

Problem with Distributional Shift I

Distributional Shift:

If Train and Test sets are not from the same Distribution such a shift is called Distributional Shift. Covariate Shift may be the most widely studied. We assume that while the distribution of inputs may change over time, the labeling function, i.e., the conditional distribution $P(y|\mathbf{x})$ does not change. Statisticians call this Covariate shift because the problem arises due to a shift in the distribution of the Covariates (features).

Example:

If Model is trained on Images from Fig.1 and when test on Images from Fig.2. There is a substantially difference in characteristics between the Train and Test Set as Training Set contains Real-World Images while Test set contains Cartoons.

Problem with Distributional Shift II

Figure: Train Images

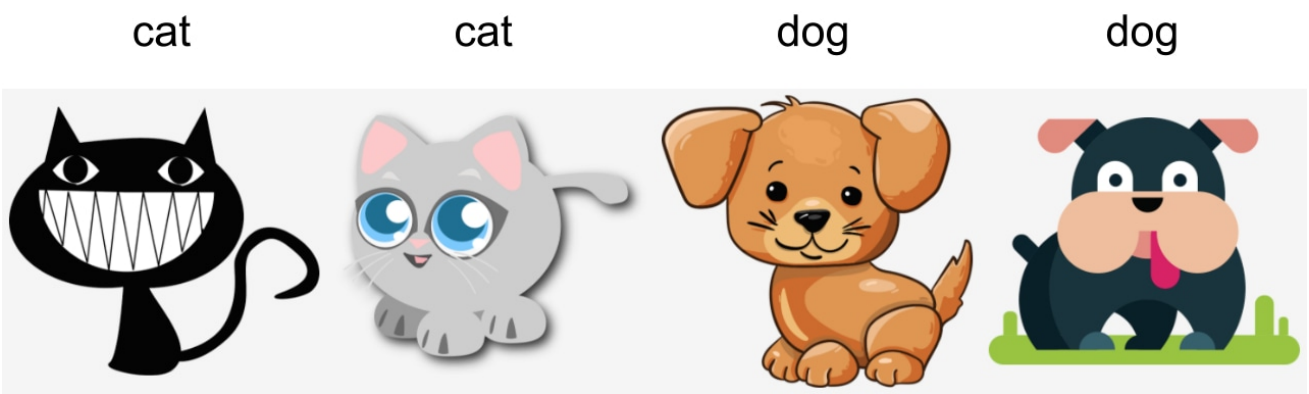


Figure: Test Images

Data Augmentation

Data Augmentation:

Data Augmentation is the process of increasing the amount and diversity of data. We do not collect new data, rather we transform the already present data. Data Augmentation is a natural and sometimes effective approach to learning robust models.

Examples of data augmentation include Adversarial Training, applying Image Transformations to the training data, such as flipping, cropping, adding Random Noise, and even stylized image transformation.

Gaussian Data Augmentation and Adversarial Training I

Gaussian Data Augmentation:

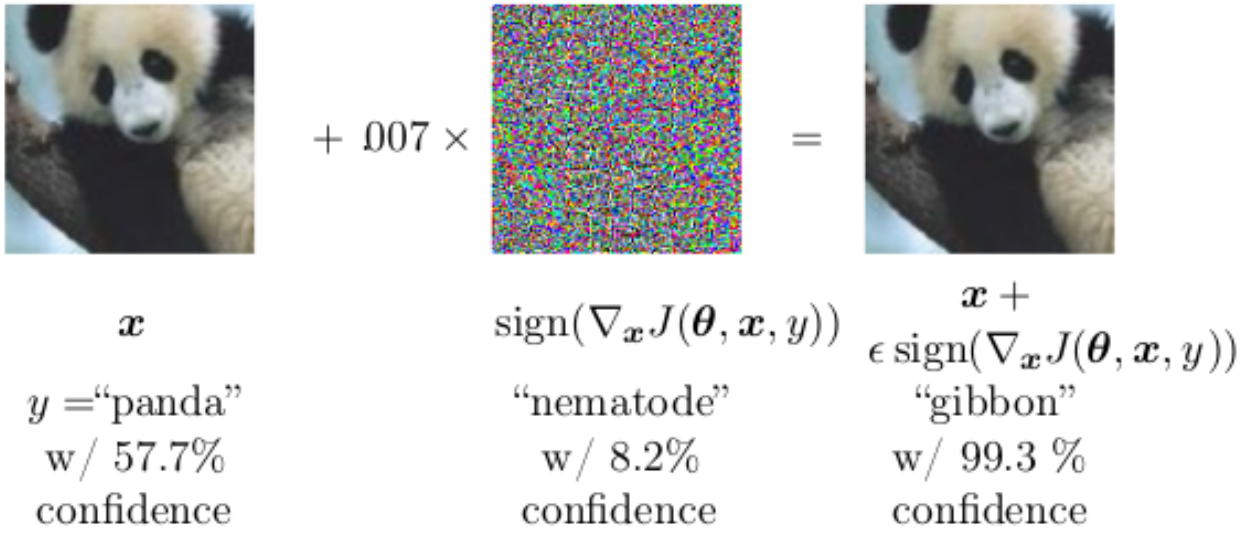
Gaussian Data Augmentation[allowframebreaks] Technique integrates Signal-to-Noise Ratio (SNR) with an Additive White Gaussian Noise (AWGN) to generate derived data samples suited for multi-class classification in various Deep Neural Networks Models.

Adversarial Training:

The process of training the model on adversarially perturbed examples from the training set in the context of Regularization i.e inorder to reduce error on Test set. Adversarial training discourages this highly sensitive locally linear behavior by encouraging the network to be locally constant in the neighborhood of the training data. This can be seen as a way of explicitly introducing a local constancy prior into supervised neural nets.

Gaussian Data Augmentation and Adversarial Training II

Examples of Adversarial Example Generation:



Training.png

Figure: A demonstration of adversarial example generation applied to GoogLeNet on ImageNet. By adding an imperceptibly small vector whose elements are equal to the sign of the elements of the gradient of the cost function with respect to the input, we can change GoogLeNet's classification of the Image. (Source: Deep Learning Book)

2D-Gaussian Distribution I

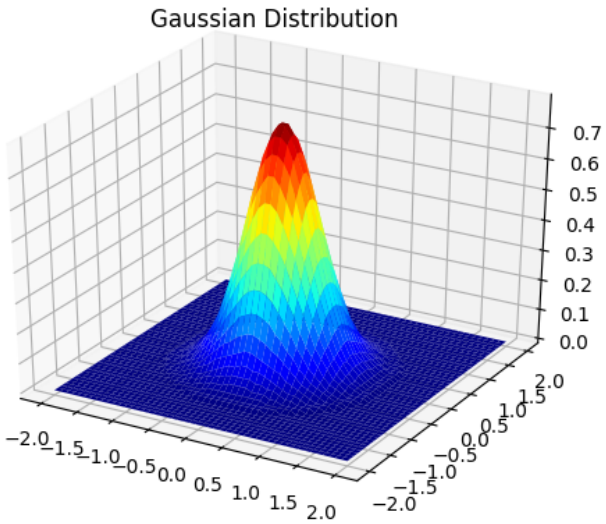


Figure: 2D-Gaussian Distribution

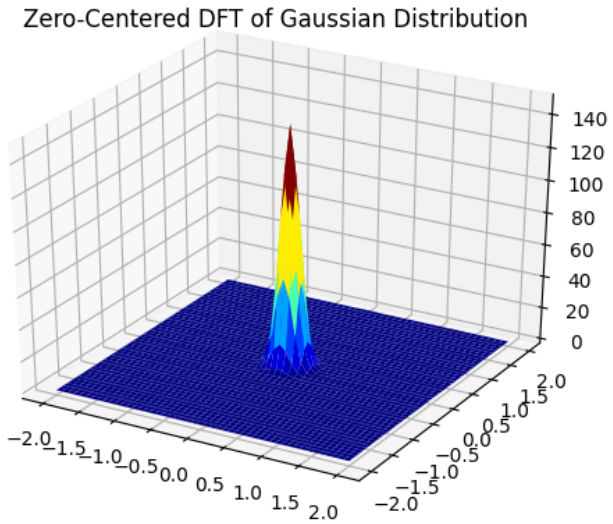


Figure: DFT of 2D-Gaussian Distribution

2D-Gaussian Distribution II

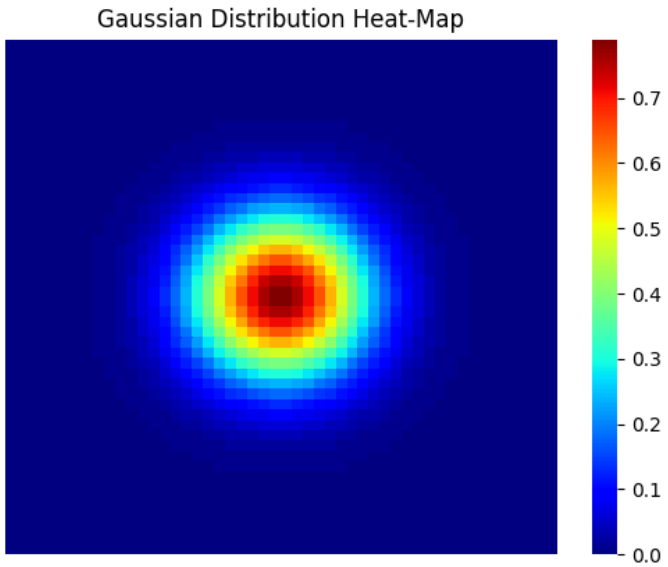


Figure: Heat-Map of 2D-Gaussian

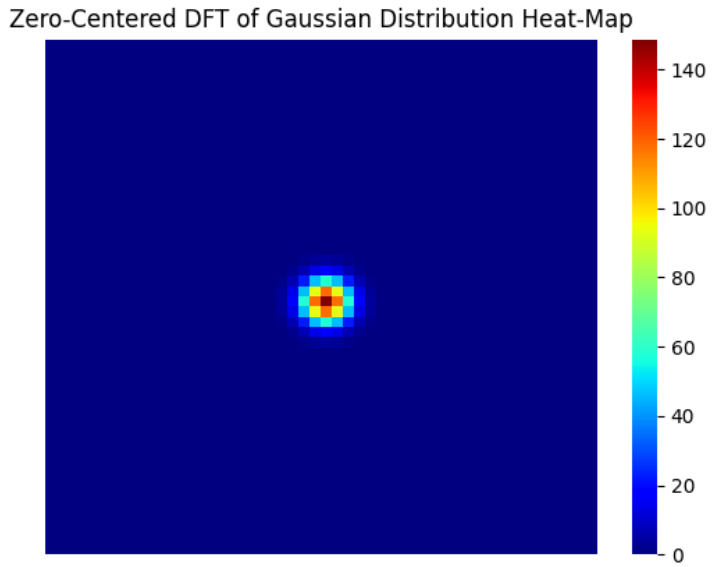


Figure: Heat-Map of DFT of 2D-Gaussian

Hypothesis and Solution

Hypothesis:

Operating Hypothesis of the paper is that the frequency information of these different corruptions offers an explanation of many of these observed trade-offs. Through extensive experiments involving perturbations in the Fourier domain, it is demonstrated that Augmentation Procedures like Gaussian Data Augmentation and Adversarial Training bias the model towards utilizing **Low Frequency Information** in the Input i.e if Input has low frequency content, corruptions with high frequencies increase the Robustness of Model while low frequency corruptions degrade the performance of model.

Solution:

More diverse Data Augmentation Procedures could be leveraged to mitigate the trade-offs called **AutoAugment**. **AutoAugment** Data Augmentation achieves state-of-the art results on the CIFAR-10-C benchmark and ImageNet-C

Investigation

Gaussian Data Augmentation:

A parameter σ for the following operation: In each iteration, we add i.i.d. Gaussian Noise $\mathcal{N}(0, \tilde{\sigma}^2)$ to every pixel in all the images in the training batch, where $\tilde{\sigma}$ is chosen uniformly at random from $[0, \sigma]$. When Gaussian Data Augmentation is used, $\sigma = 0.1$ for CIFAR-10 and $\sigma = 0.4$ for ImageNet. **Sensitivity of Models:**

In-order to investigate sensitivity of models towards corruptions, we define a $U_{i,j} \in R^{d_1 \times d_2}$ called 2D-Fourier Basis Matrices which have the following properties:

1. $||U_{i,j}|| = 1$
2. $\mathcal{F}(U_{i,j})$ only has up to two non-zero elements located at (i, j) and the its symmetric coordinate with respect to the image center.