Transformations for Domain Generalization on Fruit Datasets

Brett O'Connor
Electrical and Computer Engineering
Cornell Tech
New York, NY 10044
bwo7@cornell.edu

Max Bonzulak
Computer Science
Cornell Tech
New York, NY 10044
mwb233@cornell.edu

Abstract

Domain Generalization is considered to be an "open problem" in the field of deep learning. In this paper, we will apply image transformation techniques to training data, aiming to improve generalizability of trained models. Specifically, we will train a CNN to classify fruit based on cropped images with no background. We expect that preprocessing these training images with transformations such as random noise and translation will improve generalizability to images of fruit in-context, with backgrounds. Additionally, we expect that this improvement will be more drastic when the technique is applied to smaller training datasets, as compared to larger training datasets.

1 Introduction

Convolutional neural networks (CNN) enable many crucial technologies in today's world. However, applying trained CNNs in practice can result in some problems. Specifically, models trained on datasets from a specific IID might not perform well when applied to out-of-distribution data. We are interested in exploring data augmentation to improve the generalizability of trained models. To address this issue, we plan to train a CNN with augmented data from a source domain, and evaluate the model on unseen out-of-domain data. Our hypothesis is that training a network with augmented data will lead to better accuracy results than using the original unaugmented data. Our results will help determine the efficacy of performing augmentation processing, and how this efficacy varies based on the size of the original training dataset.

2 Background

The problem of domain generalization was first introduced related to medical imaging from patient to patient [1]. Since the problem of Domain Generalization (DG) was introduced [2], major strides have been made. Today, there are several techniques that can address the DG problem, including domain alignment, ensemble learning, and data augmentation[3]. In this paper, we will focus on data augmentation techniques. Within the category of Data Augmentation techniques (DA), there are several approaches. However, many of these approaches, such as Style Transfer Models and Feature-Based Augmentation, require data from the target distribution to be available before model training. This can be impractical in real-world applications such as traffic scene segmentation, where data representative of all possible driving conditions may not be available [3]. Instead, we are focusing on transformation techniques that do not rely on the out-of-distribution target data. Image transformations such as flips, rotations, and color augmentation have been shown to improve domain generalizability [4],[5],[6],[7],[8].

Because our training set uses centered, cut-out images, we suspect that a weakness of the model will be adapting to examples that are off-center and in-context, with the background included. Therefore, we will aim to transform the training data with random translations and random noise.

3 Proposed Work

We have two datasets of different types of fruit, which we can address as "dataset A" and "dataset B". For each of these datasets we will create a subset of these datasets that will signify our "small" dataset. Therefore, the original datasets will be considered our "large" dataset. The following will then be done for both our "small" datasets and our "large" datasets: "Dataset A" will be split into train and test groups, in which we will perform different augmentations on the training data from "dataset A". This will include rotations, shearing, transposition, and adding noise to the image. Once the augmentation is done, we will process the data to prepare for training. We will train the data on a shallow CNN created for the purposes of this experiment. We intend to iterate to identify an effective architecture, with plans to have about 2 groups of 2 convolutional layers and Max pooling, followed by a 3-layer MLP and a softmax output layer.

After training, we will evaluate our model on the test data from "dataset A". After these results are obtained, we will deploy our model on "dataset B" and gather the results. Additionally, we will take the un-augmented versions of "dataset A" and process the data to train it on a fresh copy of our neural network. From here, we will evaluate the test data from "dataset A" and then evaluate the data from "dataset B" on this model.

In total, we will be looking to compare mean-per-class accuracy and loss for our augmentation-trained model on small and large datasets with the mean-per-class accuracy and loss for our normally trained model on small and large datasets. To succeed, we would expect the augmentation-trained models for both small and large datasets to report higher mean-per-class accuracies and lower losses than the un-augmentation-trained models. Additionally, we would like to see that there are minimal differences between the small datasets and the large datasets. The following diagrams depict the work we will be doing in the project. Figure 1 is known to be our controlled experiment, with no augmentation, while Figure 2 is our treatment experiment, with augmentation.

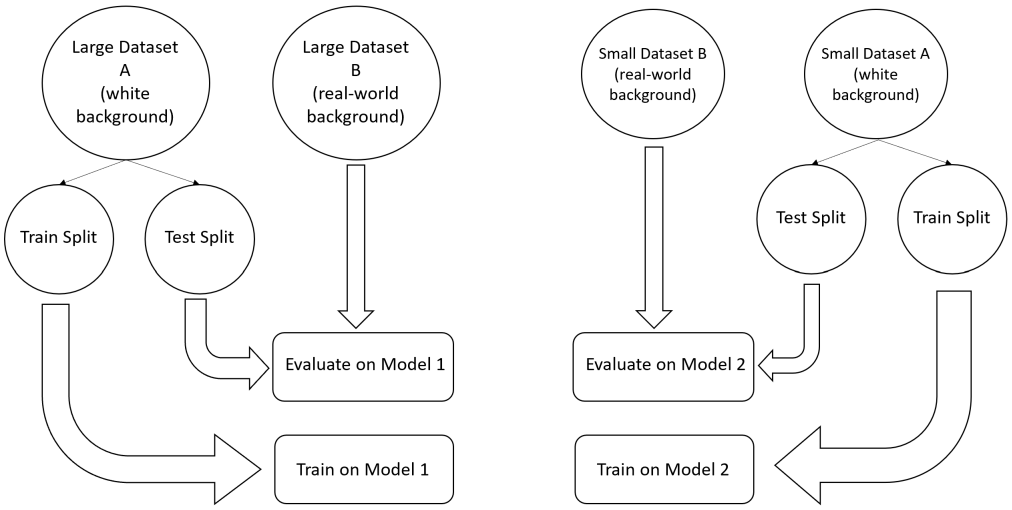

Figure 1: Control Experiment

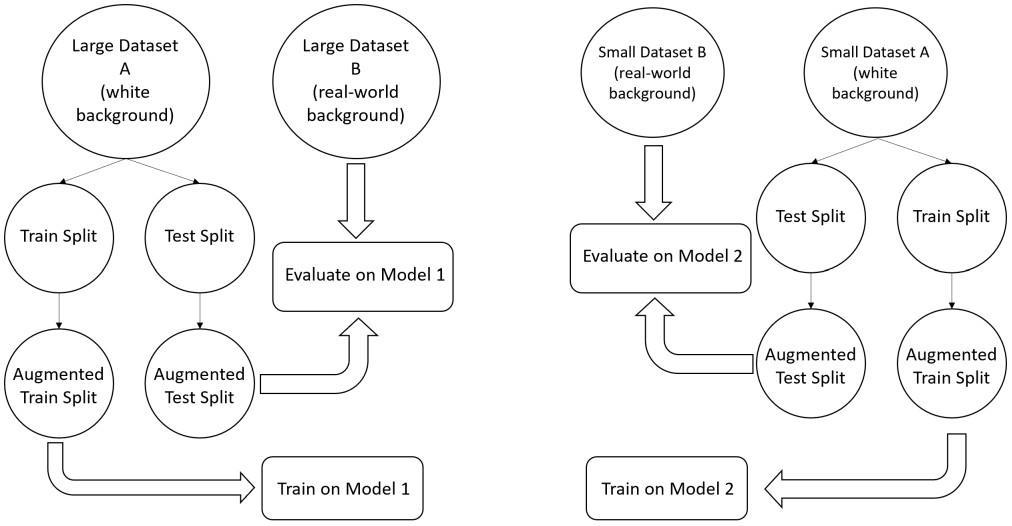

Figure 2: Treatment Experiment

	Original "A" Train Subset		*Augmented* "A" Train Subset	
	Original "A" Test Subset	*Original* "B" Test Set	*Augmented* "A" Test Subset	*Augmented* "B" Test Set
Small Train Sets				
Large Train Sets				

Table 1: Intended Result Format

4 Datasets

We will be working with two different datasets for our experiments. One dataset - “dataset A”footnote (as shown in Figure 1 and Figure 2) - contains centered images of fruit with white backgrounds and the other dataset, “dataset B” footnote, contains images of fruit with real-world backgrounds. This change in backgrounds is considered to be the domain shift that makes these datasets useful for us. “Dataset A” contains about 90,000 images of fruit and vegetables with 131 different classes. “Dataset B” contains about 160,000 images of fruit and vegetables with 298 different classes. As depicted in Figure 1 and Figure 2, we plan to use the total dataset of “dataset A” and “dataset B” as our “large datasets” and we will use a subset of “dataset A” and a subset of “dataset B” for our “small datasets”. To determine the contents of our small datasets, we will randomly select 25% of the respective large datasets’ images. For our train and test splits, we will be using 70% of the data for training and 30% of the data for testing. Shown below are examples from our datasets, with Figure 3 being from "dataset A" and Figure 4 being from "dataset B".

Figure 3: "Dataset A" Example

Figure 4: "Dataset B" Example

5 Timeline

Week 1 (4/11 - 4/15) - Preprocess and augment all necessary data for both control and treatment experiments.

Week 2 (4/18 - 4/22) - Train model's 1 and 2 for both control and treatment experiments. Evaluate data on trained models for both experiments.

Week 3 (4/25 - 4/29) - Visualize data and compare results to our hypotheses.

Week 4 (5/2 - 5/6) - Begin conducting our final analysis. Begin writing our final report.

Week 5 (5/9 - 5/13) - Finish writing report and submit.

6 About the Team Members

Brett O'Connor is a M.ENG candidate at Cornell Tech, studying electrical and computer engineering. He graduated from Cornell University with a Bachelor's in Science, also studying electrical and computer engineering, with a minor in business. His interests currently lie in the overlap between digital signal processing and machine learning and has plans to begin his professional career as a signal processing engineer. Additionally, his hobbies involve playing and enjoying baseball, DJing, and being heavily involved in the Esports industry.

Max Bonzulak is a M.ENG candidate at Cornell Tech, studying computer science. He graduated from Lehigh University with an integrated BS in Philosophy and Computer Science, and a minor in entrepreneurship. Max is interested in deep learning and its applications in deep learning. Also, Max is intrigued by GANs and generative art, though he has yet to venture into subject.

References

[1] G. Blanchard, G. Lee, and C. Scott, "Generalizing from several related classification tasks to a new unlabeled sample," in NeurIPS, 2011.

[2] K. Muandet, D. Balduzzi, and B. Scholkopf, "Domain generalization via invariant feature representation," in ICML, 2013.

[3] Zhou, K., Liu, Z., Qiao, Y., Xiang, T. and Loy, C.C., 2021. Domain generalization in vision: A survey. arXiv preprint arXiv:2103.02503, 2021.

[4] R. Volpi and V. Murino, "Addressing model vulnerability to distributional shifts over image transformation sets," in ICCV, 2019.

[5] Y. Shi, X. Yu, K. Sohn, M. Chandraker, and A. K. Jain, "Towards universal representation learning for deep face recognition," in CVPR, 2020.

[6] S. Otalora, M. Atzori, V. Andrearczyk, A. Khan, and H. M ´ uller, ¨ "Staining invariant features for improving generalization of deep convolutional neural networks in computational pathology," Frontiers in bioengineering and biotechnology, 2019.

[7] C. Chen, W. Bai, R. H. Davies, A. N. Bhuva, C. H. Manisty, J. B. Augusto, J. C. Moon, N. Aung, A. M. Lee, M. M. Sanghvi et al., "Improving the generalizability of convolutional neural network based segmentation on cmr images," Frontiers in cardiovascular medicine, 2020.

[8] L. Zhang, X. Wang, D. Yang, T. Sanford, S. Harmon, B. Turkbey, B. J. Wood, H. Roth, A. Myronenko, D. Xu et al., "Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation," TMI, 2020.