Vulnerabilities of Neural Networks using Image Predictions of Morph Sequences

William Powers
Queens College, CUNY
USA
william.powers81@qmail.cuny.edu

Lin Shi
Queens College, CUNY
USA
lin.shi66@qmail.cuny.edu

Larry S. Liebovitch
Queens College, CUNY
USA
Columbia University, NY
USA
larry.liebovitch@qc.cuny.edu

ABSTRACT

We found that keras, a high-level machine learning TensorFlow API, can be shown to exhibit vulnerabilities when used in conjunction with images of clothing from the Fashion-MNIST dataset that were incrementally morphed in sequence using the autoimagemorph python command line application [1]. These image prediction and accuracy ratings between the three types of clothing in each group are plotted across a 2-dimensional line plot. Within these line plots, there occurs strong fluctuations as one article of clothing is morphed into the next, and the keras prediction tool displays an uncertainty over which article of clothing is presented.

CCS CONCEPTS

• **Computing methodologies** → Machine learning.

KEYWORDS

Machine Learning; Image Classification; Data Visualization; Deep Learning Networks; Loss Landscapes

ACM Reference Format:
William Powers, Lin Shi, and Larry S. Liebovitch. 2022. Vulnerabilities of Neural Networks using Image Predictions of Morph Sequences. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Machine learning models are increasingly becoming better capable of task automation, anomaly detection, and predicting outcomes based on data and training algorithms. However, it should be understood that these tools are prone to errors. For example, while certain articles such as a shoe and a coat may be easily distinguishable to a human being, a machine learning technology's capability depends on the training data it is supplied, and whether it has encountered factors within this training data that would allow it to categorize the article of clothing. There is a case of a neural network classifier incorrectly predicting an image of a revolver handgun, and an image of a vulture to be an orangutan after an image rotation and translation [2]. Finding new means by which the vulnerabilities of these artificial intelligence classifiers can be exposed can potentially lead to better understanding the limitations of these tools, and how they can be improved.

A novel approach to testing the capabilities of these machine learning technologies is to present it with images that have been modified using a morph sequence generator which takes two or more images as input and creates a specified number of new images between them which emulate one image progressively morphing into the next image. While an image classifier may have a high accuracy in predicting a shirt or a dress, a combination of the two of those images mid-morph can have an adversarial effect on ability of the classifier to make that prediction.

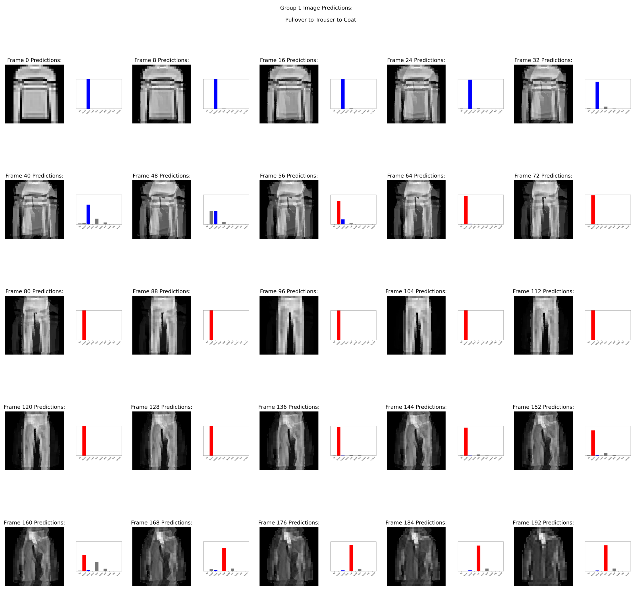

Figure 1: Predictions between 3 Articles of Clothing During Morph Sequence

2 IMAGE PREDICTIONS

We demonstrate these vulnerabilities by collecting the accuracy results of image predictions using 5 groups of 2 sets of 3 randomly

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

selected articles of clothing from 3 distinct classes from the 10 available within the dataset. This first set of classes is used to produce a second set using different articles of clothing belonging to the same classes as the first set. Each permutation of pairs of clothing from these sets of three are used as input for the autoimagemorph application, and produce a collection of 21 morph sequences each consisting of 21 images of the transition between the first and second image. Starting at 100 percent of the first image and 0 percent of the second image, the first image of each morph sequence changes at 5 percent increments until reaching 0 percent of the first image and 100 percent of the second image.

Keras is used to predict the class of clothing each of these frames within a morph sequence belong to. The accuracy rating of the predictions pertaining to the three classes of clothing used within the group are then collected within dataframes, and used to present the information on a 2-dimensional line plot using the plotly express libary of the matplotlib python tool [3]. As each line plot utilizes a different combination of two images into a third, there would be an expectation that the image predictions would less accurately predict the article of clothing toward the middle of the 21 frame morph sequence where fluctuations occur, and more accurately predict the article of clothing toward the end of the morph sequence as it morphs into a complete image.

The slices of each morph sequence are then utilized in creating 3-dimensional landscapes of each class of clothing for each set within the group (A and B). The x-axis of the landscape is determined by the sequential 21 combinations between the first and second article of clothing, and the y-axis is each of the 21 frames within each combination. The z-axis, or height of the 3-dimensional landscape is the prediction accuracy of the specified class used to label the landscape.

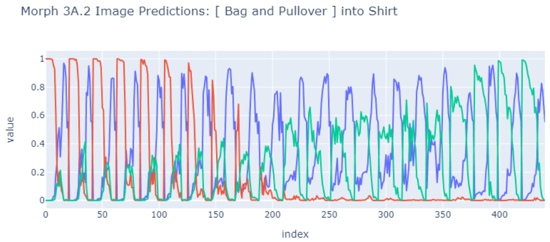

Figure 2: 2-dimensional line plot using image predictions the 2nd permutation of Group 3A (3A.2) from morph sequences of all combinations between a Bag and Pullover morphing into a Shirt

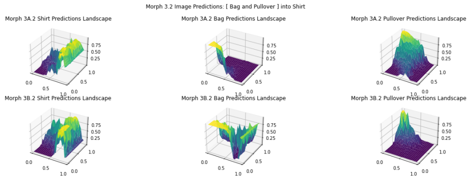

Figure 3: 3-dimensional landscapes using image predictions the 2nd permutation of Groups 3A and 3B from morph sequences of all combinations between a Bag and Pullover morphing into a Shirt

3 RESULTS

The visual 2-dimensional and 3-dimensional representations of image predictions between each of the 5 groups, while similar, present some discerning characteristics. For example, the line plot representations of Group 3A.2 to 3A.3 are of the same images of a Shirt, a Pullover, and a Bag but using a different permutation of two of the images morphing into the third. The graph of 3A.3 shows typical fluctuations between the Shirt and Pullover as they gradually morph into a complete image of the Bag, which is distinctly predicted at the end of each of the 21 cycles. Group 3A.2, instead using combinations of the Bag and Pullover morphing into a Shirt, is not so clearly defined. The Shirt is never so accurately predicted as was the Bag in the previous figure. The accuracy ratings between the first two images are also more pronounced, as the predictions for the Bag are almost entirely eliminated in favor of the Shirt and Pullover even while the composition of the frames between the Bag and Pullover are roughly 50 percent.

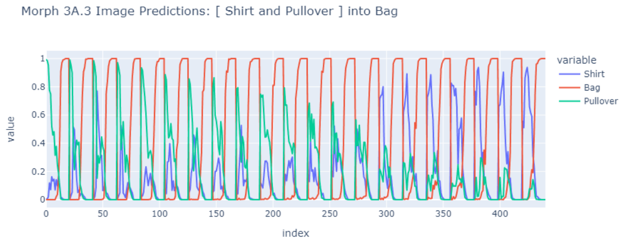

Figure 4: 2-dimensional line plot using image predictions the 2nd Permutation of Group 3 (3A.3) from morph sequences of all combinations between a Shirt and Pullover morphing into a Bag

REFERENCES

[1] András Jankovics. autoimagemorph. https://github.com/jankovicsandras/autoimagemorph, 2020.

[2] Logan Engstrom, Dimitris Tsipras, Ludwig Schmidt, and Aleksander Madry. A rotation and a translation suffice: Fooling cnns with simple transformations. *ArXiv*, abs/1712.02779, 2017.

[3] Plotly Technologies Inc. Collaborative data science, 2015.