

Data augmentation with Artistic Style

Anonymous Submission

Anonymous Affiliation

Abstract

This paper presents a novel approach to use the image style transfer by CNN as a data augmentation strategy for image-based deep learning algorithms. The success of training deep learning algorithms heavily depends on a large amount of annotated data. Recent neural style transfer approach can apply the style of an image to another image without changing high level semantic content, so we think it is reasonable to use this method as an data augmentation strategy in computer vision tasks. We explore state-of-art neural style transfer algorithms build a novel approach to apply it on the current image classification algorithms we compare to and combine with the the traditional approaches to show the effectiveness of this method. Pre-trained Vgg16 and Vgg19 are the baseline network architecture.

Keywords: Neural network, Style Transfer, Data Augmentation, Image Classification.

1 Introduction

We propose exploring the problem of data augmentation for image and video classification, and evaluating different techniques. It is common knowledge that the more data an ML algorithm has access to, the more effective it can be. Even when the data is of lower quality, algorithms can actually perform better, as long as useful data can be extracted by the model from the original data set. For example, text-to-speech and text-based models have improved significantly due to the release of a trillion-word corpus by Google [8]. This result is despite the fact that the data is collected from unaltered Web pages and contains many errors. With such large and unstructured data sets, however, the task becomes one of finding structure within a sea of unstructured data. However, alternative approaches exist. Rather than starting with an extremely large corpus of unstructured and unlabeled data, can we instead take a small, curated corpus of structured data and augment in a way that increases the performance of models trained on it? This approach has proven effective in multiple problems.

Data augmentation guided by expert knowledge [14], more generic image augmentation [18], and has shown effective in image classification [16].

The motivation for this problem is both broad and specific. Specialized image and video classification tasks often have insufficient data. This is particularly true in the medical industry, where access to data is heavily protected due to privacy concerns. Important tasks such as classifying cancer types [14] are hindered by this lack of data. Techniques have been developed which combine expert domain knowledge with pre-trained models. Similarly, small players in the AI industry often lack access to significant amounts of data. At the end of the day, we've realized a large limiting factor for most projects is access to reliable data, and as such, we explore the effectiveness of distinct data augmentation techniques in image classification tasks.

The datasets we examine are the tiny-imagenet-200 data and MNIST [2] [3]. Tiny-imagenet-200 consists of 100k training, 10k validation, and 10k test images of dimensions 64x64x3. There are a total of 500 images per class with 200 distinct classes. MNIST consists of 60k handwritten digits in the training set and 10k in the test set in grayscale with 10 classes with image dimensions of 28x28x1. To evaluate the effectiveness of augmentation techniques, we restrict our data to two classes and build constitutional neural net classifiers to correctly guess the class.

In particular, we will train our own small net to perform a rudimentary classification. We will then proceed to use typical data augmentation techniques, and retrain our models. Next, we will make use of CycleGAN [19] to augment our data by transferring styles from images in the dataset to a fixed predetermined image such as Night/Day theme or Winter/Summer. Finally, we explore and propose a different kind of augmentation where we combine neural nets that transfer style and classify so instead of standard augmentation tricks, the neural net learns augmentations that best re-

duce classification loss. For all the above, we will measure classification performance on the validation dataset as the metric to compare these augmentation strategies.

2 Related Work

This section provides a brief review of past work that has augmented data to improve image classifier performance and the state of the art techniques of neural style transfer.

2.1 Traditional Data Augmentation Techniques

Sed ut perspiciatis, unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam eaque ipsa, quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt, explicabo. Nemo enim ipsam voluptatem, quia voluptas sit, aspernatur aut odit aut fugit, sed quia consequuntur magni dolores eos, qui ratione voluptatem sequi nesciunt, neque porro quisquam est, qui dolorem ipsum, quia dolor sit amet consectetur adipisci[ng] velit, sed quia non numquam [do] eius modi tempora inci[di]dunt, ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim ad minima veniam, quis nostrum exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commodi consequatur? Quis autem vel eum iure reprehenderit, qui in ea voluptate velit esse, quam nihil molestiae consequatur, vel illum, qui dolorem eum fugiat, quo voluptas nulla pariatur?

2.1.1 Neural Style Transfer

Sed ut perspiciatis, unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam eaque ipsa, quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt, explicabo. Nemo enim ipsam voluptatem, quia voluptas sit, aspernatur aut odit aut fugit, sed quia consequuntur magni dolores eos, qui ratione voluptatem sequi nesciunt, neque porro quisquam est, qui dolorem ipsum, quia dolor sit amet consectetur adipisci[ng] velit, sed quia non numquam [do] eius modi tempora inci[di]dunt, ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim ad minima veniam, quis nostrum exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commodi consequatur? Quis autem vel eum iure reprehenderit, qui in ea voluptate velit esse, quam nihil molestiae consequatur, vel illum, qui dolorem eum fugiat, quo voluptas nulla pariatur?

3 Method

Method. How to keep high level content. how to apply styles. the formulars.

3.1 Datasets and Features

3.2 Baseline Network

We propose two different approaches to data augmentation. The first approach is generate augmented data before training the classifier. For instance, we will apply GANs and basic transformations to create a larger dataset. All images are fed into the net at training time and at test time, only the original images are used to validate. The second approach attempts to learn augmentation through a prepended neural net. At training

time, this neural net takes in two random images from the training set and outputs a single image so that this image matches either in style or in context with a given image from the training set. This output, which represents an augmented image produced by the network, is fed into the second classifying network along with the original training data. The training loss is then backpropagated to train the augmenting layers of the network as well as the classification layers of the network. In test time, images from the validation or test set is ran through only the classification network. The motivation is to train a model to identify the best augmentations for a given dataset. The remainder of this section will go into detail of the data augmentation tricks we tried.

3.1. Traditional Transformations Traditional transformations consist of using a combination of affine transformations to manipulate the training data [9]. For each input image, we generate a duplicate image that is shifted, zoomed in/out, rotated, cropped, distorted, or shaded with a hue. Both image and duplicate are fed into the neural net. For a dataset of size N , we generate a dataset of $2N$ size.

3.2. Neural Style Transfer For each input image, we select a style image from a subset of 6 different styles: Cezanne, Enhance, Monet, Ukiyoe, Van Gogh and Winter. A styled transformation of the original image is generated. Both original and styled image are fed to train the net. More detail about the GANs and style transfer can be viewed on the cited paper [19].

Figure II: Neural Style Transfer via CNN

More details about the architecture of the layers will be described in the Experiments section. We implement a small 5-layer CNN to perform augmentation. The classifier is a small 3-layer net with batch normalization and pooling followed by 2 fully connected layers with dropout. This is much similar to VGG16 in structure but smaller in interest of faster training for evaluation. We aren't aiming for the best classifier. We are exploring how augmentation tricks improve classification accuracy, reduce overfitting, and help the networks converge faster.

4 Experiments and Results

There are three sets of images that we experimented with. Each dataset is a small dataset with two classes. A small portion of the data is held aside for testing. The remaining images are divided by a 80:20 split between training and validation.

Our first data set is taken from tiny-imagenet-200. We take 500 images from dogs and 500 images from cats. 400 images for each class is allocated to the training set. The remaining 100 in each class forms the validation set. The images are 64x64x3. RGB values are also normalized for each color in the preprocessing step.

The second data set is also taken from tiny-imagenet200 except we replace cats with goldfish. The reason for this change is that goldfish look very different from dogs whereas cats visually are very similar. Hence CNNs tend to have a harder time distinguishing cats. Finally, cats and dogs have similar styles whereas images from the goldfish tend to have very bright orange styles.

Lastly, the final dataset is 2k images from MNIST, 1000 from each class. We perform the task of distinguishing 0s from 8s. MNIST images are 28x28x1 and are in gray scale. Again, images are normalized in the preprocessing step. MNIST is much more structured than imagenet so that digits are always centered. The motivation is that MNIST provides a very simple dataset with simple images. Are patterns in the more complex images also observed in simpler images?

To test the effectiveness of various augmentation, we run 10 experiments on the imagenet data. The results of the experiments are tabulated in the following table. All experiments are run for 40 epochs at the learning rate of 0.0001 using Adam Optimization. The highest test accuracy at all the epochs is reported as the best score. Once we obtained the augmented images, we feed them into a neural net that does classification. We name this neural net the SmallNet since it only has 3 convolutional layers paired with a batch normalization and max pool layer followed by 2 fully connected layers. The output is a score matrix for the weights for each class. The layers of the network is detailed below although the specific net is not very important. Any net that can reliably predict the classes suffices. Hence, one can replace this net with VGG16 with fine-tuning

on the fully connected and last convolution layers to allow for sufficient training. Consider the variable $x \in \mathbb{R}$,

$$f(x) = x^2 + 2 \tag{1}$$

Equation (1) is a polynomial of order 2.

An example of table is shown Table 1.

| | | |
|---|---|---|
| x | y | z |
|---|---|---|

Table 1: Example of table.

5 Conclusion

Data augmentation has been shown to produce promising ways to increase the accuracy of classification tasks. While traditional augmentation is very effective alone, other techniques enabled by CycleGAN and other similar networks are promising. We experimented with our own way of combining training images allowing a neural net to learn augmentations that best improve the ability to correctly classify images. If given more time, we would like to explore more complex architecture and more varied datasets. To mimic industrial applications, using a VGG16 instead of SmallNet can help us determine if augmentation techniques are still helpful given complex

enough networks that already deal with many overfitting and regularization problems. Finally, although GANs and neural augmentations do not perform much better than traditional augmentations and consume almost 3x the compute time or more, we can always combine data augmentation techniques. Perhaps a combination of traditional augmentation followed by neural augmentation further improves classification strength.

Given the plethora of data, we would expect that such data augmentation techniques might be used to benefit not only classification tasks lacking sufficient data, but also help improve the current state of the art algorithms for classification. Furthermore, the work can be applicable in more generic ways, as style transfer can be used to augment data in situations where the available data set is unbalanced. For example, it would be interesting to see if reinforcement learning techniques could benefit from similar data augmentation approaches. We would also like to explore the applicability of this technique to videos. Specifically, it is a well known challenge to collect video data in different conditions (night, rain, fog) which can be used to train self-driving vehicles. However, these are the exact situations under which safety is the most critical. Can our style transfer method be applied to daytime videos so we can generate night time driving conditions? Can this improve safety? If such methods are successful, then we can greatly reduce the difficulty of collecting sufficient data and replace them with augmentation techniques, which by comparison are much more simpler.

References