SCRUTINY OF DIVERSE IMAGE AUGMENTATION STRATEGIES FOR DEEP LEARNING

# Abstract [400 words]

Deep learning models have performed remarkably well when there is abundance of data and high computing power is available to estimate parameters present in the deep learning architectures. With the recent advances in deep learning, the size of networks have grown exponentially. Those networks require a large amount of data to train and state of the art GPUs, among both of which 2nd is readily available but we often lack in data for example in cases of biomedical images where rarity of the diseases can be issue, on top of which we have to take care of patients privacy and requirements of medical experts for labelling the images, or fault detection as the likelihood of occurrence of such an event is very low. Even for transfer learning such image features, a certain amount of data is needed. And if variance in data is not sufficient enough then the generalization ability of the deep learning CNN based model is challenged, thus it becomes very important to address this problem

These obstacles have led researchers to perform wide experiments with image data augmentation techniques specially based on Generative Adversarial Networks, the positive advances of which can be applied to solve application issues related to medical images. A lot of existing studies on image data augmentation utilizes popular datasets like ImageNet, MNIST handwritten digit dataset, tiny-imagenet-200, MIT\_Adobe-5k-datset, Pascal VOC, Cifar-10, Cifar-100 datasets. Many of the above listed datasets can be classified as big data so researchers have drawn a subset of these datasets to simulate limited image datasets. Additionally while addressing the problem of data generation, we will also try to address data imbalance problems and test how data augmentation can be a potential solution.

To solve the problem of limited data I will try to find a data-space solution based on Image Augmentation. Under the umbrella of Image Augmentation various methods will be covered which will be responsible for enhancing the sizes and quality of training datasets such that better generalizable deep learning models can be built. I will try to cover Image data augmentation methods like geometric transformation, color space augmentations, kernel filters, mixing images, random erasing, feature space augmentation, adversarial training, generative adversarial networks, neural style transfer, and meta-learning. In my work I will talk about image augmentation methods, present developments and meta-level decisions required to implement those techniques. At the end I want to present how people can leverage data augmentation techniques to improve performance of deep learning models and how to expand limited datasets to utilize capabilities of big data learning

# Literature review [700 words]

Deep learning models have progressed a lot in discriminating between the tasks. This process has further accelerated due to the availability of complex deep network architectures, compute intensive and efficient Graphic Processing Units (GPUs) and access to open sourced and state of the art big data algorithms to handle huge amount of data. By using convolutional neural nets people have successfully solved problems pertinent to computer vision like Object segmentation, Object Detection, Pose Estimation, Activity recognition. There are many different applications and field of studies which are waiting for applying concepts of Deep Learning CNNS, RNNS, GANs to improve their current benchmarks. But generalization of model or making a generalized model is one of the biggest challenge faced by researchers in deep learning. By generalization we mean that when model encounters unseen data, how does it behaves. Ideally it should be able to make correct prediction or near to correct prediction for unseen data as well, in short it should be able to mimic its performance on training data. But to this contrary, models which are not generalized are called overfitted models and thus they perform very poorly when exposed to new data. To find out whether a model is prone to overfitting or not visually, training and validation accuracies can be plotted and their difference should be analyzed. If there is a high correlation between the training and the validation accuracy, it suggests that model is generalized, meanwhile if the correlation starts dipping, and in training set error is reducing with each epoch but in validation set error is increasing after certain parameters are tuned, that means model has started to overfit. Data augmentation is one of the most powerful method in reducing overfitting in any model. The idea of data augmentation is to create different scenarios and bring in variance in data for the model so that model can be generalized properly, which should in turn reduce the difference between the training and the validation errors and between other new future test sets as well.

The motive of this study is to research in depth and test about existing data augmentation techniques as well as find and develop most recent state of the art data augmentation technique which can help the researchers to solve a very big pain point of modelling generalizable deep learning models. One of the ways that many researchers have thought through to improve the generalizing capability of model is by focusing on the performance of the architecture of the model. This though process had led to a kind of evolution in the world of deep learning, starting from AlexNETs to VggNETs, then came the ResNETS, InceptionNETS, and then very latest DenseNETS. There has been many functional kind of solutions as well introduced like Batch Normalization, Dropout, Maxpooling. Transfer learning and Pre training were developed to extend the capabilities of deep learning to the problem statements with comparatively smaller datasets or less compute capabilities. But if we notice carefully, even after such advancements, more often than not if there is an issue with generalization then most probably first reason that comes into mind is model did not get enough exposure because data points were not sufficient. Whenever we are doing data augmentation it can be intuitively understood that we are trying to extract more information from a limited static dataset. We can call this transformations as data warping augmentation which might transform the training images so that their label is constant but the data would change. There can be many different transformation for the same purpose such as geometric Augmentation (Rotation and translation), Color enhancement, Color transformation, Random erasing, neural style transfer etc. It is usually believed that bigger the dataset, more appropriately the model will learn but sometime collecting bigger dataset becomes a bottleneck problem. Especially if we talk about the medical field, where the rarity of disease can reduce the chances manifold on collecting data of affected individual due to that disease. Even if we are somehow able to collect the data, labelling that data correctly becomes a time intensive and very crucial task and require expertise of a good physician which can be little expensive as well. These obstacles has actually promoted much research in the field of data augmentation especially generative adversarial networks based oversampling.

# Proposed work [1150 words]

In this project we will cover major image augmentation methodologies as displayed in the figure-1 below

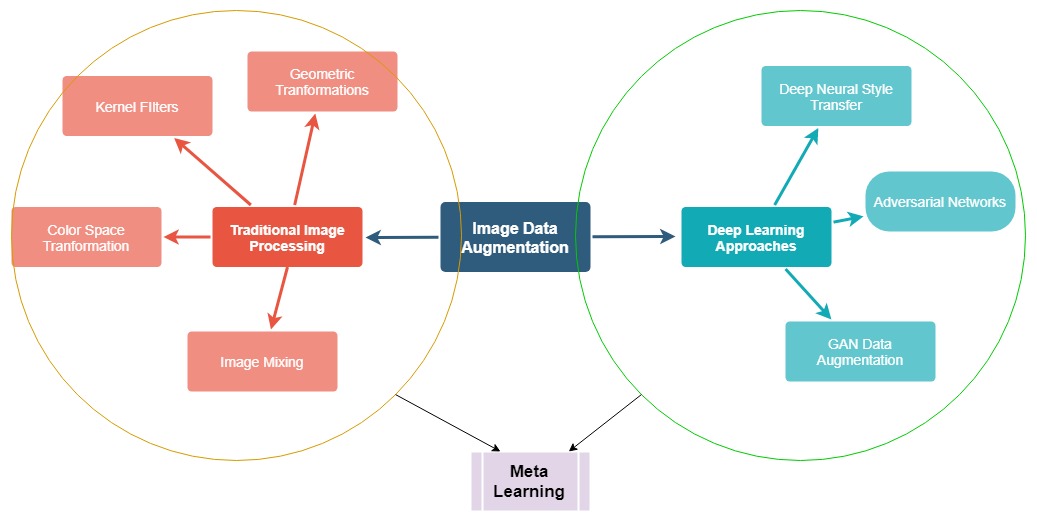


Figure 1 Displays proposed experiments we want to do to build a very robust data augmentation technique

## Traditional Image Processing based augmentation

### 1. Kernel Filters

Kernel filters are based on convolutions and they are very popular since the inception of image processing as they were being used for the sharpening and blurring of the images. One of the very interesting kernel filter is *PatchShuffle Regularization*, where pixel values are swapped in a n\*n window. This is a very popular augmentation technique which has shown promising results in cifar-10 dataset where it achieved 5.6 % error rate as compared to 6.33 % which was an earlier benchmark. Not a lot of exploration has been done when it comes to the usage of kernel filters in deep learning because many people think that these kernel filters can be embedded inside the deep neural networks by making few changes, we propose to explore those possible altercations in the project.

### 2. Geometric Transformations

Changing the geometry of the image is one of the common augmentation method that even exists in today’s modern world. SO it becomes very necessary to pay attention to the existing geometric augmentation methods and understand them so that we can apply them along with some other modified methods and get better data augmentation results. Some common geometric augmentation methods are listed below:

##### 2.1 Flipping

In flipping we see that if you take the center of the image as [0,0] coordinate then we try to swap left and right part across y axis. Although in the common world horizontal flipping is very common as compared to vertical flipping. A very imp point to be noted that flipping can’t be applied when we are trying to apply it any asymmetrical object as after applying flipping the label of the input image can be changed which is undesirable.

###### 2.2 Cropping

To create different viewpoint or view of the scene cropping can be proved as very useful data augmentation techniques where we can do zoom in or zoom out operations to generate little different perspective when we try to think in terms of camera. Usually not more than 5% or 10% of cropping is done in terms of distance from the center

###### 2.3 Rotation

Rotation is one of the very popular data augmentation technique. We usually apply rotation in those kind of images where some kind of symmetry is present in the input images, such that on rotating the label of the input image should not change. It has been observed in MNIST digit recognition that slight rotation applied to the input images can improve the results by a considerable margin.

### 3. Color Space Transformations

Modern digital images usually consists of three dimension i.e. height, width and color channels. Usually RGB images are found where R, G, B belongs to Red, Green and Blue respectively. So It is a very task in terms of mathematics to augment the color channels, it means you have to just move the matrix forward and backward and also augmentation in terms of values of colors can also be done such that to do those augmentation we can plot the histogram of color values distribution and change those histograms.

### 4. Mixing Images

Mixing images is such a image augmentation technique which is very un-intuitive and from a human point of view it becomes very difficult to understand. It becomes more difficult when you get a performance gain but you cannot visualize that why in the first place you got that improvement. In this data augmentation technique two images can be randomly cropped and added, or may be randomly flipped and added. They might be added by averaging the pixel values of each pixels. A tradition followed in the mixing images that after adding up the images, the label of first selected image is assigned. The resultant image of this operation is also a image which is used for training of classification model.

### 5. Random Erasing

It is also one of the very interesting technique used in data augmentation. It basically helps us to regularize the model where you will just randomly take an input image, crop some areas out of it and erase that particular cropped area. Post erasing the same images has to be fed to the model with label remain unchanged and you will look for the output.

## Deep Learning based image augmentation

### 1. Generative Adversarial Networks based image augmentation

A very exciting strategy has popped up in recent years, which goes by the name Generative Modelling. It refers to the process of creating new artificial data points from the exiting data such that they can retain the characteristics of existing data and be a new additional data point in itself. GAN like other data augmentation technique can e taken as a very intelligent process of deriving more information not from just single data point but from a collection of many other data points. While combining GAN framework with variational autoencoders we can successfully describe data point using low level representation, for example images with the three dimension of (height, width, color channel) can be represented with a single vector of size (n\*1). Low dimensional representation of data input points makes it difficult to visualize for humans, so method called t-SNE representation can be used to visualize those vectors. After getting a vector representation pair of images can be merged to created new images and then using decoder to create full fledge images. GAN was first proposed by Goodfellow et el. Consisting of a discriminator and a generator. GAN have successfully reproduced new dataset on MNIST dataset. From recent research it has been found that cyclicGAN, DCGAN have good potential to do data augmentation in modern deep learning science.

### 2. Neural Style Transfer based Augmentation

Neural style transfer is one of the finest capabilities of deep learning science Using neural style transfer a lot of awesome apps has been developed for creative and sometimes funny image manipulations. Those manipulations were the work of modern CNNs. Thus neural style transfer has got a lot of appreciation from artistic minds. The idea with which neural style transfer was developed to manipulate or add information in image in CNN so that new information from other images can be added to other image without actually changing the original representations. Fast Style Transfer developed by Gatys. Et el. Is one of the great advancement in neural style transfer. The algorithm converts the loss function from a per-pixel loss to the perceptual loss and uses feed forward network to stylize images. This new concept of perpetual loss has demonstrated great potential in solving problem statement of super resolution as well. Neural style transfer has proved its mettle in data augmentation science which can improve the generalization ability of convolutional neural networks.

# References [120 words]

[1] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012;25:1106–14.

[2] Karen S, Andrew Z. Very deep convolutional networks for large-scale image recognition. arXiv e-prints. 2014.

[3] Kaiming H, Xiangyu Z, Shaoqing R, Jian S. Deep residual learning for image recognition. In: CVPR, 2016.

[4] Christian S, Vincent V, Sergey I, Jon S, Zbigniew W. Rethinking the inception architecture for computer vision. arXiv e-prints, 2015.

[5] Gao H, Zhuang L, Laurens M, Kilian QW. Densely connected convolutional networks. arXiv preprint. 2016.

[6] Nitish S, Geoffrey H, Alex K, Ilya S, Ruslan S. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014;15(1):1929–58.