CIS 4900 Independent Study Report: Comparing Various Image Augmentation Methods used in CNN based Deep Learning Models

Param Vipulkumar Chokshi
University of South Florida
Department of Computer Science and Engineering
pvc@usf.edu

1. Introduction

The scope of the study was to study the efficacy of various image augmentation techniques in improving the accuracy of CNN (Convolutional Neural Network) based deep learning models. Image augmentation generates similar but distinct training examples after a series of random changes to the training images, thereby expanding the size of the training set. Alternatively, image augmentation can be motivated by the fact that random tweaks of training examples allow models to less rely on certain attributes, thereby improving their generalization ability[2].

2. Experimental Setup

The setup used here to compare Image augmentation techniques is a model designed and trained for Classification of Global Microglia Proliferation. As proposed in [1], a custom CNN architecture is used, as shown in Figure 1. All convolution layers used ReLU activation, the two fully connected layers used ReLU with the addition of L2 regularization, the output consists of a softmax activation function, and finally the model is trained using Stochastic Gradient Descent (SGD) for optimization. To enhance performance, the model uses the snapshot ensemble approach with a cyclic learning rate in the form of cosine annealing to produce variability in the models as training proceeds. This technique allows for training a single time and saving intermediate models, which then can be used in an ensemble without the need for training the model multiple times. For each augmentation method, the model is trained 110 epochs with a cycle of 10 epochs and a maximum learning rate of 0.004. This approach produces 11 models used to predict on the images then those 11 predictions are averaged for a final classification.

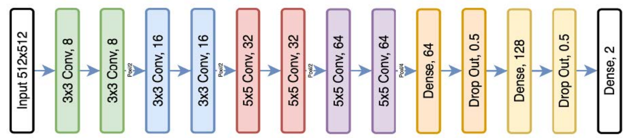
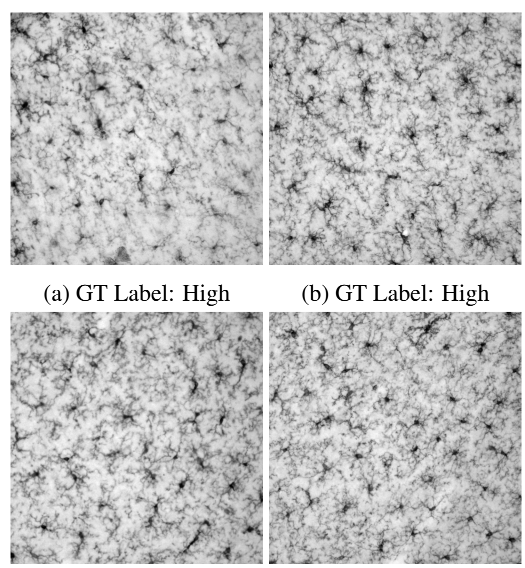


Figure 1: Network Architecture

The data on which the image augmentation are performed are low magnification images (20x) of Iba1-immunostained microglia cells. The model predicts high or low microglial proliferation at the global level using local images during training. The training dataset is annotated by an expert as high and low cases.



(a) GT Label: High (b) GT Label: High

(c) GT Label: Low (d) GT Label: Low

Figure 2: Example Images

3. Baseline Model

The model used as the basis of comparison is the original model used in the paper [1], trained on images which were augmented 11 times. Thus the data set consisted of 1 original image + 11 augmented images per each image in the original data-set. The augmentation method used here is a combination of elastic and rotation transformation. The original image goes through two elastic augmentations. These three new images (two elastic augmented and 1 original) are augmented/rotated three times at Angles [90, 180, 270] creating in total 12 images. Thus for consistency each singular experiment will be conducted by augmenting 1 image to produce 12 images. The following table shows the results from the baseline model.

Fold Number	Test Animal ID	Test Animal GT Class	Total # images to classify	# Images classified as High	# Images Classified as Low	Majority Class Predicted	Correct or Incorrect Prediction
1	943	High	869	677	192	High	Correct
	894	Low	815	654	161	High	Incorrect
2	979	High	863	783	80	High	Correct
	981	Low	893	666	227	High	Incorrect
3	1101	High	928	797	131	High	Correct
	1015	Low	910	416	494	Low	Correct
4	735	High	849	425	424	High	Correct
	888	Low	916	456	460	Low	Correct
5	924	High	883	513	370	High	Correct
	1016	Low	870	277	593	Low	Correct
6	978	High	906	459	447	High	Correct
	995	Low	564	115	449	Low	Correct
7	853	High	1112	800	312	High	Correct
	895	Low	1048	726	322	High	Incorrect

Figure 3: Results from the Baseline model. 11/14 correct

3.1. General Experiment Process

The input dimension for the network used here are 512*512. Thus a data pre-processing step, The original color images are converted to gray scale image using correlation based gray-scale conversion. The images are then cropped into 726*726 sized images. These images are used to apply image augmentation. A size larger than 512*512 is used to remove the occurrences of black patches arising due to augmentation methods like rotating. Augmentation is applied to the larger images and at the end a centre crop of the size 512*512 is stored in the training data set.

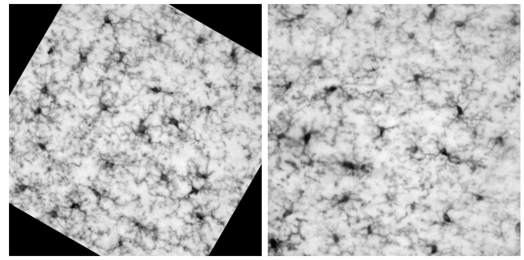
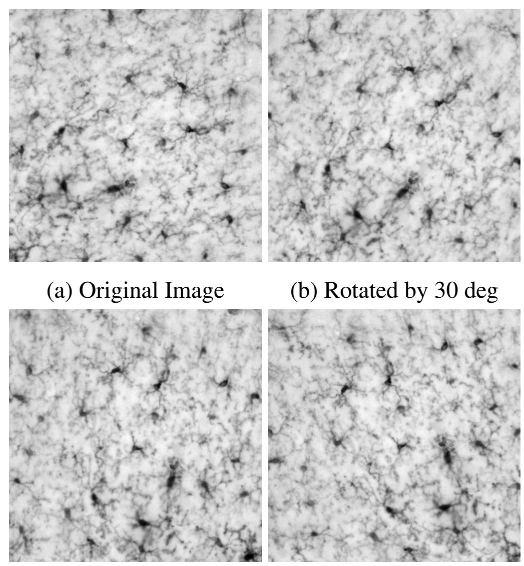


Figure 4: (left) A 512*512 image rotated by 60 degrees. (right) Centre crop from a 726*726 image rotated by 60 degrees.

3.2. Rotation Augmentation

A 2D Rotation transformation rotates the image about its origin using a rotational matrix. Rotation preserves the angle, and length of the image elements. Here to obtain 11 extra images per image we rotate each image by the following angles:
[30, 60 90, 120, 150, 180, 210, 240, 270, 300, 330]



(a) Original Image (b) Rotated by 30 deg

(c) Rotated by 60 deg (d) Rotated by 90 deg

Figure 5: Example of rotated images

The model was trained on these images and it produce the following results. Compared to our baseline, rotation augmentation performs worse. This decrease in performance, can be assumed to have risen from the fact the a combination of elastic and rotation transformation may represent variations in cells better than rotation alone. The shape of the cells are constantly changing and it does not follow any specific geometric pattern. Thus rotation alone is not adequate in mimicking/ creating images similar to the original images.

Fold Number	Test Animal ID	Test Animal GT Class	Total # images to classify	# Images classified as High	# Images Classified as Low	Majority Class Predicted	Correct or Incorrect Prediction
1	943	High	869	632	237	High	Correct
	894	Low	815	623	192	High	Incorrect
2	979	High	863	480	383	High	Correct
	981	Low	893	398	495	Low	Correct
3	1101	High	928	611	317	High	Correct
	1015	Low	910	310	600	Low	Correct
4	735	High	849	635	214	High	Correct
	888	Low	916	498	418	High	Incorrect
5	924	High	883	883	0	High	Correct
	1016	Low	870	883	0	High	Incorrect
6	978	High	906	451	455	Low	Incorrect
	995	Low	564	128	436	Low	Correct
7	853	High	1112	800	312	High	Correct
	895	Low	1048	672	376	High	Incorrect

Figure 6: Results from rotation augmentation. 9/14 correct

3.3. Elastic Augmentation

Elastic Augmentation applies elastic distortions deformation to the images. The implementation used here uses both local distortion and random affine transformation. To apply the deformation, a centre square is selected from the image. A 2D matrix of the same size is constructed comprising of rnadom numbers close to the centre square. An affine transform matrix is derived from the 2D matrix and the centre square. This affine matrix is used to transform the entire image.

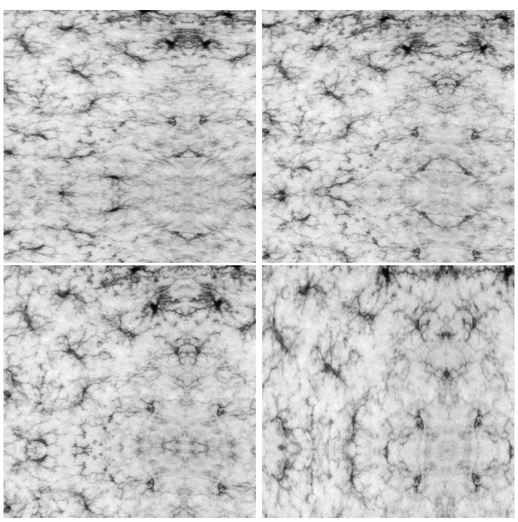


Figure 7: Example of elastic augmented images

The model was trained on these images and it produce the following results. Compared to the previous experiment, elastic augmentation perform much better. The results obtained are very close to the baseline experiment. Furthering our hypothesis from the previous section, the results from this experiments prove that elastic augmentation can better mimic cell images by performing random affine transformation. The augmented images produced due to stretching and shearing helps the model make better predictions. It is important to note here, that performing random affine transformations can produce out of proportions stretched images that vary from the original images by a lot. Thus here, the random seeds used to generate the random 2D matrix are produced in a way that the dice coefficient between the original images and the augmented images is between 0.9-1.0.

Fold Number	Test Animal ID	Test Animal GT Class	Total # images to classify	# Images classified as High	# Images Classified as Low	Majority Class Predicted	Correct or Incorrect Prediction
1	943	High	869	869	0	High	Correct
	894	Low	815	815	0	High	Incorrect
2	979	High	863	412	451	Low	Incorrect
	981	Low	893	294	599	Low	Correct
3	1101	High	928	526	402	High	Correct
	1015	Low	910	281	629	Low	Correct
4	735	High	849	557	292	High	Correct
	888	Low	916	467	449	High	Incorrect
5	924	High	883	529	354	High	Correct
	1016	Low	870	392	478	Low	Correct
6	978	High	906	472	434	High	Correct
	995	Low	564	157	407	Low	Correct
7	853	High	1112	695	417	High	Correct
	895	Low	1048	641	407	High	Incorrect

Figure 8: Results from elastic augmentation. 10/14 correct

3.4. Other Augmentations - Future work

There were a lot of other of augmentation methods which were implemented but not experimented with the model. These methods can be used to train the model in the future and study its effects:

1. Smoothing or Blurring augmentation: Applying smoothing filters on the images. Example: Gaussian Blur with various kernel sizes
2. Horizontal/Vertical Flipping: flipping or inverting the image over the centre horizontal axis or the center vertical axis.
3. Adding Noise: Adding some noise value to each pixel in the image. The values to be added can be drawn from a uniform random normal distribution.
4. Horizontal/Vertical Shift: shifting either some rows or columns to right/left or up/down.
5. Adding Occlusions: Adding occlusion (black patches) to random location in the image
6. Zooming: Zooming into various part of the image, and expanding the image using interpolation.

4. Conclusions

Comparing two augmentations in the experiment revealed that an augmentation which can efficiently produce images closer to the actual variants of the original image might prove more useful. In our case elastic augmentation perform better than stand alone rotation augmentations. Our baseline which is a combination of both performs better than both stand alone methods. Potentially based on the fact that cells can rotate and change shapes, the combination of the two augmentation methods produces a more realistic data set. Our hypothesis can be further explored by experimenting with other augmentation methods listed in the future work. Based on our current reasoning augmentations like shifting and flipping should be efficacious since they represent shifted cells, while methods like zooming might not be useful in predicting high/low density labels.

5. References

[1] Morera, H.,Dave, P., Kolinko, Y., Allen, K.,Alahmari, S., Goldgof, D., Hall, L.O., Mouton P.R., "Classification of global microglia proliferation based on deep learning with local images," Proc. SPIE 12032, Medical Imaging 2022: Image Processing, 120322K (4 April 2022); https://doi.org/10.1117/12.2611581

[2] Zhang, A., Lipton, Z., Li, M., & Smola, A. (2021). Dive into Deep Learning. arXiv preprint arXiv:2106.11342.