Q1

(a) 4×9

(b)
$$\begin{bmatrix} W(2,2), 0, 0, 0, 0, 0, 0, 0, 0 \\ 0, 0, W(2,0), 0, 0, 0, 0, 0, 0 \\ 0, 0, 0, 0, 0, 0, W(0,2), 0, 0 \\ 0, 0, 0, 0, 0, 0, 0, 0, W(0,0) \end{bmatrix}$$

Q2.

① $x_0 = 2$

$z_1 = W_1 x_0 + b_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

$A_1 = \begin{bmatrix} \max(0,3) \\ \max(0,2) \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

$z_2 = W_2 A_1 + b_2 = \begin{bmatrix} 7 \\ 9 \end{bmatrix}$

$A_2 = \begin{bmatrix} \max(0,7) \\ \max(0,9) \end{bmatrix} = \begin{bmatrix} 7 \\ 9 \end{bmatrix}$

$z_3 = W_3 A_2 + b_3 = 7 + 9 - 1 = 15$

$W = \frac{dh(x)}{dx} = \frac{dz_3}{dA_2} \cdot \frac{dA_2}{dz_2} \cdot \frac{dz_2}{dA_1} \cdot \frac{dA_1}{dz_1} \cdot \frac{dz_1}{dx} = W_3 \odot I(z_2>0)^T \cdot W_2 \odot I(z_1>0)^T \cdot W_1$

$= [1,1] \odot [1,1] \cdot \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \odot [1,1] \cdot \begin{bmatrix} 1.5 \\ 0.5 \end{bmatrix} = 6$

$b = h(x_0) - W x_0 = 15 - 6 \times 2 = 3.$ ① $W = 6,\ b = 3$

Similarly. For ② $x_0 = -1$

$z_1 = \begin{bmatrix} -1.5 \\ 0.5 \end{bmatrix}$, $A_1 = \begin{bmatrix} 0 \\ 0.5 \end{bmatrix}$, $z_2 = \begin{bmatrix} 1 \\ 1.5 \end{bmatrix}$, $A_2 = \begin{bmatrix} 1 \\ 1.5 \end{bmatrix}$, $z_3 = 1.5$

$W = \frac{dh(x)}{x} = W_3 \odot I(z_2>0)^T \cdot W_2 \odot I(z_1>0)^T \cdot W_1$

$= [1,1] \odot [1,1] \cdot \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \odot [0,1] \cdot \begin{bmatrix} 1.5 \\ 0.5 \end{bmatrix} = 1.5$

$b = h(x_0) - W x_0 = 1.5 - 1.5 \times (-1) = 3$ ② $W = 1.5,\ b = 3$

For ③ $x_0 = 1$

$z_1 = \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix}$, $A_1 = \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix}$, $z_2 = \begin{bmatrix} 4.5 \\ 5.5 \end{bmatrix}$, $A_2 = \begin{bmatrix} 4.5 \\ 5.5 \end{bmatrix}$, $z_3 = 9$

$W = \frac{dh(x)}{x} = W_3 \odot I(z_2>0)^T \cdot W_2 \odot I(z_1>0)^T \cdot W_1$

$= [1,1] \odot [1,1] \cdot \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \odot [1,1] \cdot \begin{bmatrix} 1.5 \\ 0.5 \end{bmatrix} = 6$

$b = h(x_0) - W x_0 = 9 - 6 \times 1 = 3$ ③ $W = 6,\ b = 3$

Assignment 2 Theory Problem Set

DO NOT TAG

Name: Haoming Zhang
GT Email: hzhang961@gatech.edu

Theory PS Q1. Must show your work for full credit. Feel free to add extra slides if needed.

Theory PS Q2. Must show your work for full credit. Feel free to add extra slides if needed.

Assignment 2 Paper Review

DO NOT TAG

Provide a short preview of the paper of your choice.

The main contribution of this paper is that it shows that ImageNet trained CNNs are strongly biased towards recognizing textures rather than shapes, which is in stark contrast to human behavioral evidence and reveals fundamentally different classification strategies. In order to address this inconsistency, the paper presents Stylized-ImageNet(SIN), an adapted iteration of the ImageNet dataset where images undergo processing to reduce texture information and emphasize shape features.

The paper improved understanding of CNN representations and biases and advanced towards more plausible models of human visual object recognition. They proposed the SIN, which improved performance on classification and object recognition. The paper also demonstrates that a shape-base representation can increase robustness to diverse image distortions. In addition, some networks trained on SIN even outperform human performance in terms of robustness to distortions.

However, this paper doesn't generalize this method to other architectures besides ResNet-50.

My personal takeaway from this paper is that, this paper provides a novel method to align CNNs closer to human vision strategies. This provides insight into exploring advanced object recognition models that are both robust and accurate. In addition, it provides a useful starting point for future undertakings where domain knowledge suggests that a shape-based representation may be more beneficial than a texture-based one.

Paper specific Q1. Feel free to add extra slides if needed.

We care about the biases of the neural network, because these an impact the model's generalization. If the neural network is overly biased towards texture features, it may be hard for the network to generalize to new data which has similar shapes but irrelevant textures. On contrast, shape-based features trained on SIN generalize well to natural images. It's beneficial to align the network's bias with human because it can lead to models that are more interpretable and perform better on tasks that require human understanding of the visual world, even if it may not improve performance on the training set. However, different biases might be advantageous for different tasks or applications. It is not necessary and desirable that the neural networks have to have the same biases as humans. In other words, we may not always desire the network to have the same biases as humans.

Paper specific Q2. Feel free to add extra slides if needed.

Because in the stylized image, shape features are emphasized and texture features are reduced. Training on stylized image moves the bias of the network towards recognizing objects based on their shape. The change in bias can lead to enhanced generalization and robustness because shape-based features proves to be a more consistent and stable feature across image distortions and variations compared with texture. The Res-Net trained on SIN exhibits a much stronger shape bias compared to the same neural network trained on standard ImageNet. In some categories, the shape bias are almost as strong as that of humans. Therefore, Training on stylized image can help the network perform better on classification tasks and be more resilient to image distortions, which could help create more robust and generalizable CNNs.

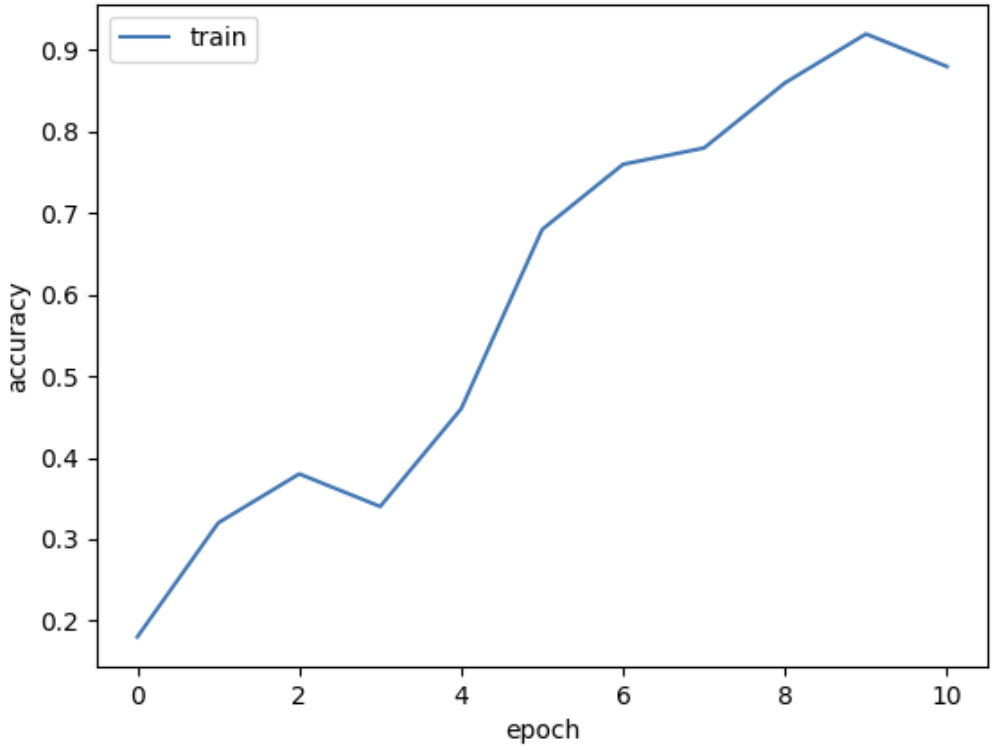
Assignment 2 Writeup

DO NOT TAG

Part-1 ConvNet

DO NOT TAG

Put your training curve here:



My CNN Model

DO NOT TAG

Describe and justify your model design in plain text here:

My model is a Convolutional Neural Network (CNN) with a sequence of 3 convolutional layers followed by ReLU activation functions and max-pooling layers. I used 3 sets of convolutional layers with increasing depth to capture more abstract and complex features the input images. Max-pooling layer is used to downsample the spatial dimensions. Batch Normalization is conducted to help stabilize the training process by reducing internal covariate shift, maintaining a more consistent distribution of inputs. And there will be 1 final fully connected layer performing classification.

Describe and justify your choice of hyper-parameters:

I chose a moderate batch size of 32 and regularization rate of 0.0004 to prevent overfitting. I used a learning rate of 0.001 which is a commonly used starting point and found it worked well. A momentum of 0.9 is passed in to help accelerate convergence by incorporating information from the previous step. I increased the epoch number to 12 in case the training curve wouldn’t reach a plateau with an epoch of 10.

What’s your final accuracy on validation set?

0.8052

Data Wrangling

DO NOT TAG

What’s your result of training with regular CE loss on imbalanced CIFAR-10?

Tune appropriate parameters and fill in your best per-class accuracy in the table

	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9
CE Loss	0.8500	0.7270	0.3850	0.0660	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000

What’s your result of training with CB-Focal loss on imbalanced CIFAR-10?

Additionally tune the hyper-parameter beta and fill in your per-class accuracy in the table; add more rows as needed

	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9
beta= 0.5000	0.9320	0.9540	0.5440	0.4660	0.3210	0.0480	0.1950	0.0740	0.0010	0.0050
beta= 0.9990	0.7380	0.5590	0.2670	0.2100	0.2160	0.3170	0.3990	0.4440	0.4420	0.3660

Put your results of regular CE loss and CB-Focal Loss(best) together:

	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9
CE Loss	0.8500	0.7270	0.3850	0.0660	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000
CB-Focal	0.7380	0.5590	0.2670	0.2100	0.2160	0.3170	0.3990	0.4440	0.4420	0.3660

Explain how you verified the correctness of your focal loss solution. Be specific; describe any testing you did and justify how the testing verifies correctness:

Firstly I conducted the unit test by setting the hyperparameter gamma to 0. In this case, the focal loss should be equal to cross-entropy loss. Secondly, I trained Res-Net on the imbalanced CIFAR-10 data with CE loss and with focal loss, and compared two sets of per-class accuracies by using Kolmogorov-Smirnov (KS) test. The KS test is a non-parametric test that compares the cumulative distribution functions (CDFs) of two samples. The null hypothesis of the KS test is that the distributions of acc_CE and acc_Focal are the same. The p-value of the KS test is 0.052, which means that there is a marginal statistical significance, suggesting there may be a significant difference between two distributions. Therefore, the correctness of the focal loss solution is verified. The focal loss solution has a better per-class accuracy overall.

Describe and explain your observation on the result:

*Explanation should go into **WHY** things work the way they do in the context of Machine Learning theory/intuition, along with justification for your experimentation methodology. **DO NOT** just describe the results, you should explain the reasoning behind your choices and what behavior you expected. Also, be cognizant of the best way to mindful and show the results that best emphasizes your key observations. If you need more than one slide to answer the question, you are free to create new slides.*

I observed that, when using CE loss, the model performs well on instances of majority classes 0 – 2 but bad on instances of minority classes 3 – 9. When using focal loss to add more weights to the minority classes, the model performs better on classes 3 – 9. As we know, focal loss is a modification of the standard cross-entropy loss by assigning larger weights to minority classes. It adjusts the contribution of each class based on its prevalence in the dataset. Because the data is highly imbalanced, we has to pass in a large \beta so that a large weight will be added to the minority classes. I tuned \beta, the hyperparameter of focal loss for reweighting of each class. By observing the per-class accuracy during the training process, I found that the best beta might fall into 0.9990 and 0.9999. With \beta < 0.9990, the prediction accuracies on instances of class 3 – 9 are low. While with \beta > 0.9999 the prediction accuracies on instances of class 0 – 2 will decrease, the accuracies on instances of class 3 – 9 will increase.