**Abstract**

Neural networks often face challenges such as overfitting, limited data availability, noisy data, lack of data diversity, and co-adaptation between neurons in deep networks etc. To address these issues, techniques such as data augmentation, dropout, and cutout have been introduced in deep learning. These techniques offer several benefits and enhance the robustness of models, ultimately resolving the challenges. By incorporating these techniques, deep learning models can achieve improved accuracy and reliability. In this project, we conduct a comprehensive comparative analysis of various cutout and augmentation techniques on the CIFAR-10 dataset using the ResNet18 architecture. The primary objective is to evaluate the impact of different cutout configurations and data augmentation techniques through two steps. First, we assess the influence of various cutout shapes and intensities on the model's performance. Then, we examine enhancing the model's performance by implementing diverse data augmentation techniques, once we have identified the best Cutout configuration. To achieve this, we train seven models, including six different cutout options and one model without cutout, and 5 combinations of augmentation techniques , over 150 epochs. We investigate the effectiveness of different data augmentation methods, such as random rotation, color jitter, random horizontal flip, random crop, and random vertical flip. These augmentation techniques are examined through various combinations to determine the optimal configuration for enhancing the model's accuracy and robustness. The results reveal that the square cutout shape, without intensity, yields the highest test accuracy of 95.42% in the last epoch and maintains consistent performance throughout the training. Furthermore, our findings demonstrate the effectiveness of specific data augmentation combinations in enhancing the model's convergence speed and overall performance. Overall, this research sheds light on the importance of cutout and augmentation techniques in deep learning models and provides valuable insights for improving their generalization and performance on real-world datasets like CIFAR-10.

**Introduction**

Convolutional Neural Networks (CNNs) are a class of deep learning models specifically designed for processing and analyzing visual data, such as images or videos. They have achieved remarkable success in various computer vision tasks, including image classification, object detection, and image segmentation. Modern networks commonly contain on the order of tens to hundreds of millions of learned parameters which provide the necessary representational power for such tasks. However, training deep CNNs ,with multiple layers, presents several challenges.

One significant challenge is the vanishing problem. The vanishing problem refers to the difficulty of training deep neural networks, including CNNs, when the gradients become extremely small during backpropagation. As the gradients propagate through many layers, they tend to diminish exponentially, making it challenging for the earlier layers to update their weights effectively. Consequently, the network fails to learn meaningful representations in those layers, limiting its overall performance.

To address this issue, techniques such skip connections have been introduced, one of the popular techniques implemented by the ResNet architecture, that have emerged as a powerful solution to the vanishing problem. ResNet is a groundbreaking architecture that addresses the vanishing problem and enables the training of extremely deep CNNs. ResNet introduced the concept of residual connections, also known as skip connections, which allow the network to bypass certain layers and propagate gradients directly to earlier layers. By doing so, the gradients have a shorter path to propagate through the network, enabling the successful training of extremely deep networks **[Deep Residual Learning for Image Recognition].** This architecture has significantly contributed to the development of state-of-the-art CNNs and has been widely adopted in various computer vision tasks.

Deep networks are particularly susceptible to overfitting due to their high complexity and the potential for capturing spurious correlations within the training data. To combat the potential for overfitting, several different regularization techniques can be applied, such as data augmentation.

Data augmentation is a widely used technique in deep learning, especially for CNNs. It involves applying various transformations to the training data, such as rotation, translation, scaling, flipping, and more. The goal of data augmentation is to artificially increase the diversity and size of the training set, thereby reducing the risk of overfitting. By exposing the network to a wider range of variations and distortions, data augmentation helps CNN learn more robust and invariant features, leading to improved performance on unseen data.

Another widely used technique is dropout. Dropout is a regularization technique introduced **by Geoffrey Hinton et al.** It involves randomly deactivating a fraction of neurons during training, i.e. random subsets of units are temporarily ignored during training. During testing, the entire network is used without dropout, but the learned weights are scaled by the dropout rate to ensure proper inference. Dropout reducing the reliance on any single unit and encouraging the network to learn more robust and generalizable features. Dropout helps prevent the model from over-relying on specific activations or connections and acts as a form of ensemble learning within the network.

Cutout is a data augmentation technique that enhances regularization by masking out regions of input images during training. This process forces the network to learn from the contextual information across different areas of the image .The main motivation for cutout comes from the problem of object occlusion, which is commonly encountered in many computer vision tasks. By generating new images with masked regions to simulate occluded examples during training, cutout helps the model become more robust and better prepared for real-world scenarios. This approach encourages the model to consider a broader context of the image, leading to improved generalization performance and more reliable decision-making in the presence of occlusions.

**Motivation**

The advancement of Deep Learning has revolutionized various fields, especially in computer vision, where Convolutional Neural Networks (CNNs) have achieved remarkable success in image recognition tasks. Our project is motivated by the desire to explore and optimize the performance of CNNs using insights from previous reaches (ResNet, improve, Dropout ).Our project proposal aims to explore and compare different combinations of cutout shapes and intensities to identify the best configuration that achieves the highest accuracy. We will investigate various shapes and different cutout transparency levels and their combinations. Then, we investigate the impact of additional data augmentation techniques on the best-performing Cutout configuration. Horizontal flipping, random rotation and other transformations will be examined to understand how they interact with the best cutout configuration and improve the model's ability to handle diverse and challenging real-world data.

Through this comprehensive analysis, we strive to contribute to the field of Deep Learning by uncovering strategies to enhance CNN performance. By building upon the concepts from the seminal articles and experimenting with novel combinations, our project aims to provide valuable insights for future advancements in image recognition tasks and inspire further research in regularization and data augmentation techniques. Ultimately, we seek to enable CNNs to achieve state-of-the-art accuracy while maintaining robustness and adaptability across various real-world scenarios.

**Experiments**

1. **CIFAR-10 and Analysis**

To evaluate the model's performance and determine the best cutout shape and intensity, we systematically apply each cutout type to the CIFAR-10 dataset. CIFAR-10 dataset contains 60,000 32x32 color images in 10 different classes: airplanes, cars, birds, cats, deer, dogs, frogs, horses, ships, and trucks. Each dataset is split into a training set with 50,000 images and a test set with 10,000 images.

1. **Implementation**

In the initial phase of our implementation, we focus on enhancing the Cutout technique by exploring various shapes and levels of transparency. This entails experimenting with different cutout shapes, to observe how each shape impacts the model's performance. Additionally, we will vary the transparency levels of the cutout regions to understand their influence on the model's ability to generalize and handle occlusions effectively. Having established the best Cutout configuration, we will proceed to the second step, where we aim to further improve the model's performance through data augmentation. By employing various data augmentation techniques (and examine different combinations), we try to enhance the robustness of the best Cutout configuration. The goal is to ensure that the model can learn from a diverse range of augmented training data, enabling it to adapt effectively to real-world scenarios with varying challenges and occlusions. Our aim is to find the most effective combination of cutout shape, transparency level, and data augmentation techniques, thereby maximizing the model's accuracy and generalization capabilities.

**2.1** **Cutout Experiment**

2.1.1 Cutout shape:

In our quest to identify the ideal shape for cutout, we conducted a comprehensive investigation encompassing three distinct shapes: rectangle, triangle, and circle. Through experimentation, we sought to discern the impact of these diverse shapes on the model's performance and generalization capabilities.

Square**:** The cutout square shape, characterized by a fixed side length which represents the length (in pixels) of each square patch and received as an argument by the user. In our experiment, we fixed the length to be 16 pixels. The application of cutout involves the random selection of a pixel coordinate within the image, serving as the center point for placing the cutout mask. By adopting this approach, the square region is dynamically positioned on the image during the training process. This approach allows for the possibility that parts of the cutout mask may extend beyond the image boundaries. (*Figure 1)*

Triangle: For the triangle shape, the initial steps are like those used for the square. The fixed side length of the triangle is set at 16 pixels also. We begin by randomly selecting a pixel coordinate within the image and applies a triangular mask centered around that location. The mask effectively "cuts out" a triangular region from the image. *(Figure 2)*

Circle: The circle shape implementation involves the following steps: First, random selection of a pixel coordinate within the image, serving as the center point. Then, a random size between 8 to 16 pixels is chosen, which influence about the circle size. The circle's radius is calculated as the minimum dimension between the image's height and width divided by the selected size. This calculation ensures that the circular cutout fits entirely within the smaller dimension (either height or width) of the image. pixel coordinates for the image's height and width are generated to create the image grid. Using a Boolean condition, a circular mask is formed by evaluating the distance of each pixel from the center point (x, y) and including pixels within the specified radius while excluding those outside it. Finally, the circular mask is applied to the image, effectively removing pixels within the circular region. (*Figure 3)*

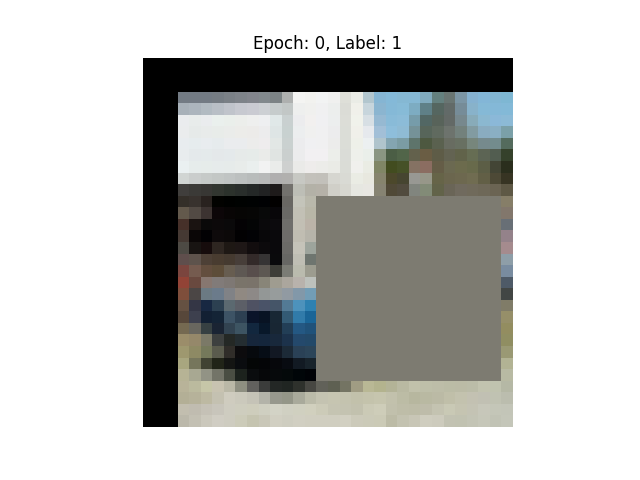


Figure 1: Square CutOut with padding



Figure 2: Triangle CutOut with padding

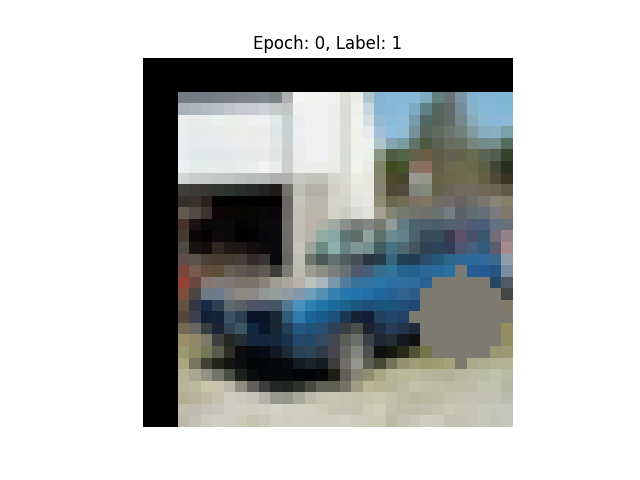


Figure 3: Circle CutOut with padding

**2.1.2 Cutout Intensity**



Figure 4 Transparency shapes

In our investigation, we explored the effect of different transparency levels (“intensity”) on the cutout technique. The transparency value, denoted as "intensity," is randomly generated within the range of 0 to the maximum transparency value, which, in this case, is 0.5. A transparency value of 0 represents complete opacity, where the cutout region is entirely covered, covering the selected portion of the image without revealing any of the original content. Conversely, a transparency value of 0.5 means that the cutout region is 50% transparent, allowing 50% of the original content underneath to be visible through the masked region. In practice, the intensity level is used as a scaling factor to modify the pixel values within the cutout region.

It acts as a multiplier that reduces the pixel values in the cutout area, making it darker or less intense compared to the original image. By varying the intensity of the cutout, we aimed to examine how different transparency levels influence the model's generalization capabilities and its ability to learn robust features. The investigation allowed us to explore the trade-off between transparency levels and their impact on the model's performance when evaluated on previously unseen data.

The incorporation of various cutout implementations, encompassing square, triangle, and circle shapes, alongside different transparency levels (intensity), results in a total of six distinct cutout types that need to be examine, also we examine model without cutout.

**2.1.3 Data augmentation**

The images are normalized using per-channel mean and standard deviation. Additionally, a standard data augmentation scheme for these datasets was applied (improved ResNet), which includes zero-padding the images with 4 pixels on each side to obtain a 40x40 pixel image. From the padded image, a crop of 32x32 pixels is extracted with a 50% probability. Furthermore, the images are randomly mirrored horizontally also with a 50% probability. As mentioned, the next step involves enhancing the model's performance by implementing diverse data augmentation techniques, once we have identified the best Cutout configuration.

* 1. **Comparative Analysis of Cutout Techniques on CIFAR-10 Using Improved ResNet**

To evaluate the cutout types on the CIFAR-10 datasets, we trained 7 models (6 different cutouts options and one model without cutout) using a deep residual network with depth of 18 (ResNet18 architecture) over 150 epochs.

Following previous reaches, we use an approach of learning rate decrease **(improved ResNet).** The learning rate is initially set to 0.1 but is scheduled to decrease by a factor of 5x after each of the 60th, and the 120th epochs.

* 1. **Cutout best model Results**

Upon executing seven distinct models, we arrived at the following results:

All models achieved excellent test and train accuracies, surpassing 94%. The investigation reveals that the most effective model is the one utilizing the cutout shape 'square' without intensity, achieving an impressive test accuracy of 95.42% in the last epoch. Remarkably, this model consistently demonstrates the highest test accuracy throughout all epochs, implying its superiority. The second-best model features the 'triangle' cutout shape, displaying a slight difference in accuracy (0.12% for the last epoch and 0.29% for the maximum test accuracy). Notably, the models incorporating intensity-based cutouts exhibit slightly lower performance, with the 'intensity\_triangle' model ranking third in terms of maximum test accuracy. These findings lead to the conclusion that employing the full cutout (without transparency) proves optimal, substantially enhancing model robustness and readiness for real-world scenarios.

|  |  |  |
| --- | --- | --- |
| **Model** | **Max Test Acc** | **Last Epoch Acc** |
| not cutout | 95.28 | 94.98 |
| Shape triangle | 95.33 | 95.30 |
| Shape square | **95.62** | **95.42** |
| Shape circle | 95.25 | 94.98 |
| Intensity triangle | 95.31 | 94.97 |
| Intensity square | 94.54 | 94.47 |
| Intensity circle | 95.08 | 94.85 |

Examining the learning curves (figure ---), we observe a proximity between the train and test curves of the model using the cutout shape 'square' without intensity. This proximity signifies its strong generalization ability during the training process, even though it may not have the fastest convergence rate. Notably, all models exhibit a comparable convergence speed, indicating that they converge to their optimal performance levels at an approximately similar pace.

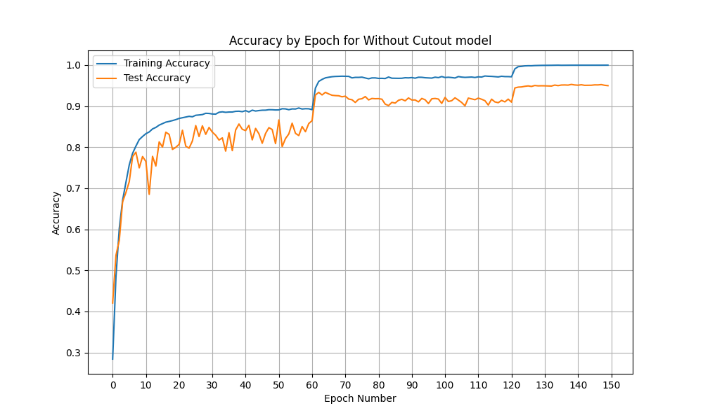
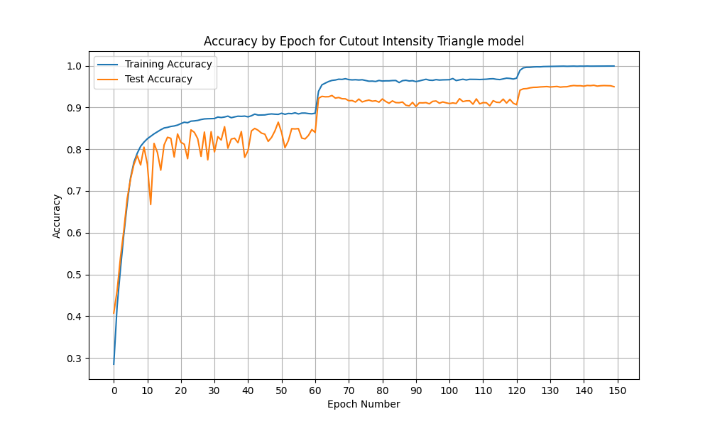
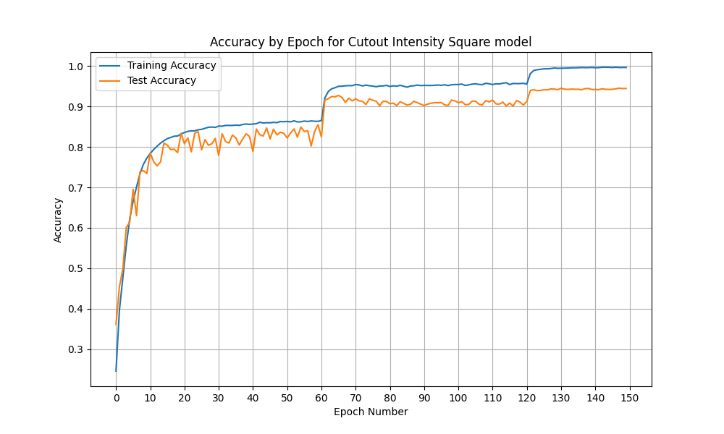
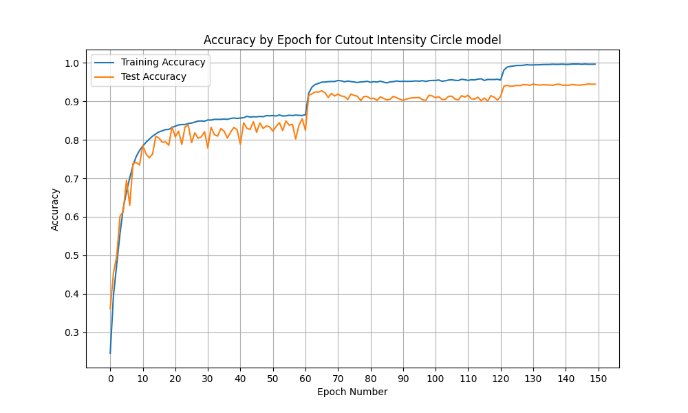
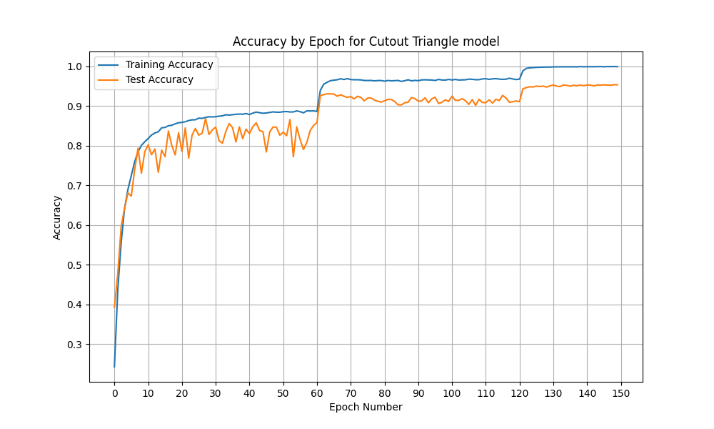
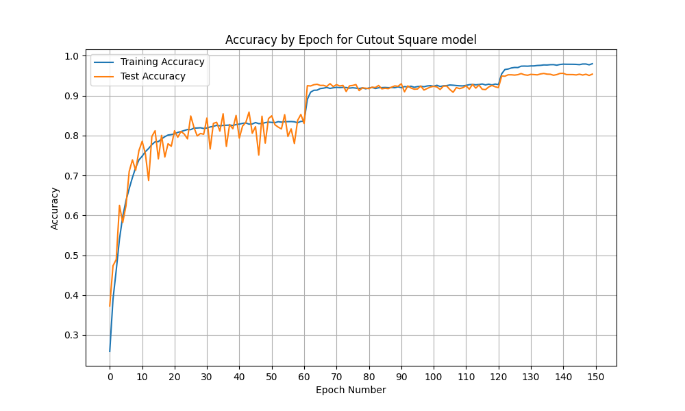
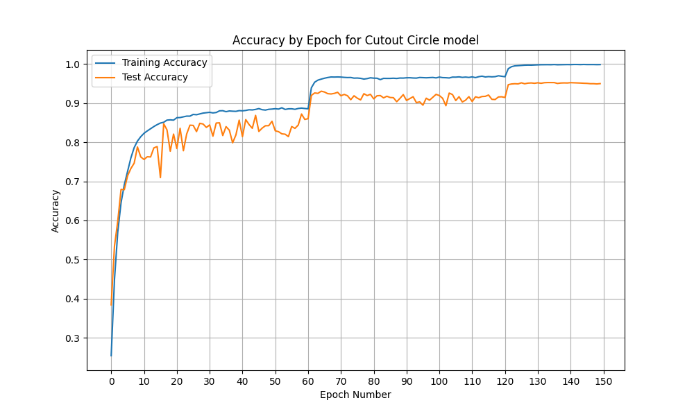


Figure 5: Train and Test curves over epochs

* 1. **Data augmentation experiment:**

We are focusing on several types of data augmentation to improve the model's performance. After thorough examination and exploration, we have identified the best augmentations for this dataset, which include Random Rotation, Color Jitter, Random Horizontal Flip, Random Crop, and Random Vertical Flip.

While we have also explored other augmentations for small epoch sizes, they showed relatively lower performance compared to the chosen augmentations, therefore we have chosen to concentrate on the mentioned augmentations. Our investigation involves five combinations of data augmentation:

1. Random Crop + Random Rotation
2. Random Crop + Color Jitter
3. Random Crop + Color Jitter + Random Rotation
4. Random Crop + Color Jitter + Random Horizontal Flip
5. Random Crop + Random Horizontal Flip + Random Vertical Flip
6. Random Crop + Random Horizontal Flip + Color Jitter

By analyzing the results of these augmentations, we aim to determine the most effective combination that enhances the model's accuracy and robustness on this dataset, with the best cutout configuration model ( cutout shape 'square' without intensity).

* 1. **Data augmentation Results:**

|  |  |  |
| --- | --- | --- |
| **Data augmentation** | **Max Test Acc** | **Last Epoch Acc** |
| Random Crop + Random Rotation | 98.15 | 94.19 |
| Random Crop + Color Jitter | 98.1 | 94.3 |
| Random Crop + Color Jitter + Random Rotation | 97.15 | 94.19 |
| Random Crop + Color Jitter + Random Horizontal Flip | 96.9 | 95.2 |
| Random Crop + Random Horizontal Flip + Random Vertical Flip | 95.86 | 94.24 |
| Random Crop + Random Horizontal Flip + Color Jitter | 96.91 | 95.35 |
| Random Crop + Random Horizontal Flip | **95.62** | **95.42** |

**Conclusion**