

Image Edge Enhancement for Effective Image Classification

Bu Tianhao¹, Michalis Lazarou² and Tania Stathaki²

¹Glory Engineering & Tech Co., LTD, China

²Imperial College London, U.K.

Keywords: Data Augmentation, Image Classification, High Boost Filtering, Edge Enhancement.

Abstract: Image classification has been a popular task due to its feasibility in real-world applications. Training neural networks by feeding them RGB images has demonstrated success over it. Nevertheless, improving the classification accuracy and computational efficiency of this process continues to present challenges that researchers are actively addressing. A widely popular embraced method to improve the classification performance of neural networks is to incorporate data augmentations during the training process. Data augmentations are simple transformations that create slightly modified versions of the training data, and can be very effective in training neural networks to mitigate overfitting and improve their accuracy performance. In this study, we draw inspiration from high-boost image filtering and propose an edge enhancement-based method as means to enhance both accuracy and training speed of neural networks. Specifically, our approach involves extracting high frequency features, such as edges, from images within the available dataset and fusing them with the original images, to generate new, enriched images. Our comprehensive experiments, conducted on two distinct datasets—CIFAR10 and CALTECH101, and three different network architectures—ResNet-18, LeNet-5 and CNN-9—demonstrate the effectiveness of our proposed method.

1 INTRODUCTION

Deep learning has undeniably been at the forefront of technological advancement over the past decade, with a transformative influence, particularly in the field of computer vision (He et al., 2016). However, training neural networks effectively for computer vision tasks poses many limitations, because of the reliance on substantial amount of data to avoid overfitting (Zhang et al., 2019).

Data augmentation is a widely recognized technique for mitigating model overfitting and enhancing generalization, effectively addressing the issue of insufficient sample diversity (Shao et al., 2023). A common practice within the realm of data augmentation is to apply colour space and/or other geometric transformations to the original images in order to increase the size and diversity of the training dataset. Some notable works of data augmentation techniques include image mix-up approaches, such as the creation of new images by directly merging two images from distinct classes (Zhang et al., 2017), replacing specific regions of an image with patches from other images (Yun et al., 2019) and various other inventive variations (Liu et al., 2022), (Yin et al., 2021), (Kim et al., 2020). Other approaches utilize the frequency domain information (Chen and Wang, 2021), (Mukai et al.,

2022) with a focus on generating new sets of images from diverse sources.

In our work, we present a novel data augmentation method that transforms the original images by merging them with their corresponding extracted edges, that we will refer to as *Edge Enhancement* (E^2). It operates on the same principle as high boost image filtering, a technique that accentuates the edges of an image without completely eliminating the background. The proposed E^2 method extracts the edges (high frequency components) from each of the RGB channels of the original images and subsequently integrates these extracted features with the original image to yield the edge-enhanced images. E^2 adheres to the fundamental concept of introducing subtle modifications by utilizing these extracted features to generate new images, all without relying on complex algorithms, thus ensuring efficient preprocessing.

The main contributions of our work are summarized as follows:

1. We propose a novel data augmentation method that enhances the semantic information of every image in the training dataset which in turn improves the training of neural networks.
2. Our method has undergone rigorous evaluation on two distinct datasets, CIFAR10 and CALTECH101, employing neural network models in-

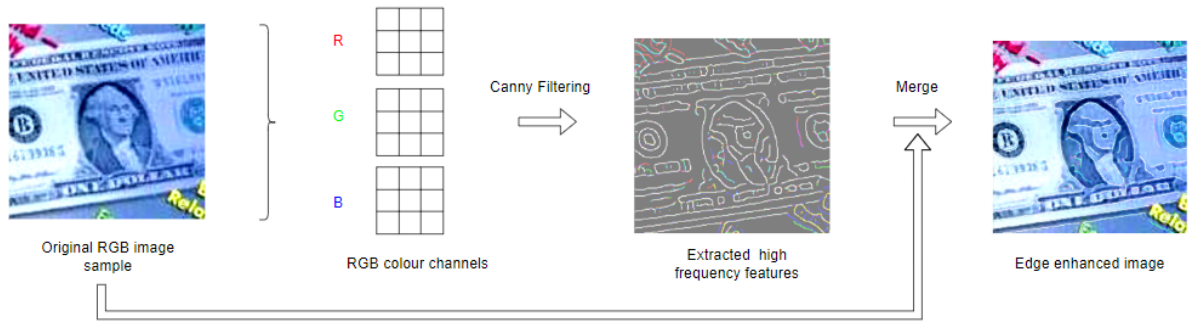


Figure 1: Edge enhancement process. Leftmost: original image. Mid left: corresponding R,G,B channels of the original labelled data. Mid right: corresponding extracted edges. Rightmost: edge enhanced image.

cluding CNN-9, LeNet-5 and RESNET-18. These experiments have consistently showcased substantial improvements in classification accuracy, accompanied by a reduced requirement for training epochs to achieve optimal classification accuracy.

2 BACKGROUND

2.1 Data Augmentation

Data augmentation is a technique that expands dataset sizes by introducing subtle alterations.

The simplest data augmentation strategies are through applying geometric transformations as rotation and flipping (Chlap et al., 2021). Our research aligns with this principle of generating new labeled data by modifying the original dataset. However, it deviates from the conventional methods of rotation and flipping. Instead, we propose a novel approach that prioritizes the enhancement of semantic information. This innovative method draws inspiration from the concept of high-boost filtering, which will be elaborated upon in the subsequent section.

2.2 High Boost Filtering

High-boost filtering is a well-established image processing technique, representing a variation of high-pass filtering. In contrast to high-pass filtering, high-boost filtering aims to accentuate an image's high-frequency information while preserving the background. Let's denote an image undergoing this process as $I(u,v)$ (Srivastava et al., 2009), where u,v correspond to spatial coordinates. High boost filtering can be formally defined as follows:

$$I_{boost}(u,v) = A \cdot I(u,v) - I_{low}(u,v) \quad (1)$$

where $I_{boost}(u,v)$ is the high boosted image pixel at location u, v , and A is a scaling factor greater than

1 (common choices are 1.15, 1.2) and $I_{low}(u,v)$ is the low pass filtered image pixel. Respectively, the high pass filtering can be presented in a similar way as,

$$I_{high}(u,v) = I(u,v) - I_{low}(u,v) \quad (2)$$

where $I_{high}(u,v)$ is the high pass filtered image pixel at location u, v . Our method employs the concept of high-boost filtering, where we combine the high-pass filtered image with the original version, with $A = 2$. This selection for A can be viewed as a particular instance of high-boost filtering.

2.3 Canny Edge Detection

Canny edge detection algorithm was invented by John F Canny in 1986 (Canny, 1986). The procedure of canny edge detection can be shown as follows: 1) Noise suppression within the image is accomplished by employing a Gaussian filter (a low-pass filter) (Deng and Cahill, 1993). The kernel for the 2-D Gaussian filter can be represented as:

$$G(u,v) = \frac{1}{2\pi\sigma^2} e^{-\frac{u^2+v^2}{2\sigma^2}} \quad (3)$$

The denoised result is achieved by convolving the kernel with the image, where σ^2 represents the filter variance. 2) Determine the edge strength ∇G and the four orientations θ (horizontal, vertical and the two diagonals), expressed as

$$\nabla G = \left(\frac{\partial G}{\partial u} + \frac{\partial G}{\partial v} \right)^{0.5} \quad (4)$$

$$\theta = \tan^{-1} \left(\frac{\partial G}{\partial v} / \frac{\partial G}{\partial u} \right) \quad (5)$$

where $\frac{\partial G}{\partial u}, \frac{\partial G}{\partial v}$ are the first order derivatives (gradients) in the horizontal and vertical directions. 3) Non-maximum suppression, which utilizes the edge strength to suppress non-edge pixels. 4) Double thresholding, a method which involves establishing both high and low edge strength thresholds, which

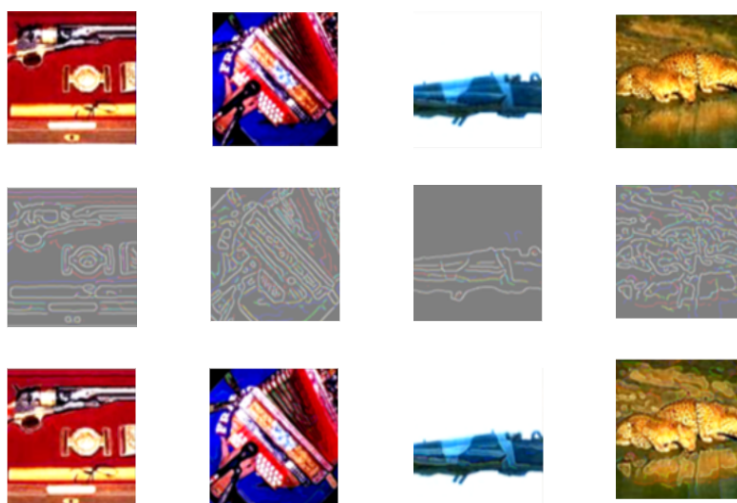


Figure 2: A batch of 4 images. Top row: original data. Mid row: corresponding extracted edges. Bottom row: corresponding edge enhanced data.

are employed to categorize the strength of edge pixels into three levels: strong, weak, and non-edges. Additionally, non-edge pixels are subsequently suppressed. 5) Hysteresis edge tracking represents the final step, employing blob analysis to further handle weak edges. Those weak edges which lack a strong edge neighbor connection are identified as potential noise or color variation-induced edges and subsequently removed. In our work we use Canny edge detection as a core technique to extract the edge information in each colour channel from the dataset of interest.

3 METHODOLOGY

3.1 Problem Definition

Let us define a labeled image dataset $D = (\mathbf{x}_i, \mathbf{y}_i)$, where \mathbf{x}_i represents the i^{th} image and \mathbf{y}_i represents the corresponding class label of image \mathbf{x}_i . This dataset comprises the images, each containing three RGB channels. The dataset is partitioned into three splits: the training set split, $D_{\text{train}} = (\mathbf{x}_i, \mathbf{y}_i)$, the validation set split, $D_{\text{val}} = (\mathbf{x}_i, \mathbf{y}_i)$ and testing set split, $D_{\text{test}} = (\mathbf{x}_i, \mathbf{y}_i)$. We use D_{train} to train a neural network, that consists of a backbone f_θ and a classifier g_ϕ (last layer of the network). Additionally, we denote the edge features extracted from the training data as \mathbf{v}_i , this is merged with \mathbf{x}_i to create the edge enhanced data \mathbf{e}_i .

The validation set D_{val} is used in order to save the model with the highest validation accuracy. Finally, we use D_{test} to calculate the test set classification accuracy.

3.2 Edge Enhancement

Our E^2 pre-processing can be illustrated in Figure 1, where the leftmost image is the original data. After applying Canny filtering the extracted edges are shown and the rightmost image is the edge enhanced result. We first apply normalization to the sample images, and then resize the images to ensure all images have exactly the same resolution, also maintaining consistent image size across the datasets. Canny filters are employed to the every image at each of its RGB channels to obtain the extracted edge information. The extracted edge information is then combined with their respective original versions, resulting in a transformed image set. Each image in this set is labeled according to the original class labels.

3.3 Training Phase

In each batch, we apply edge enhancement to a set of training images, as depicted in Figure 2. In this representation, the top four images represent the original data, while the bottom four belong to the edge-enhanced images, corresponding to the edge-enhanced dataset $\tilde{D} = (\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$, where $\tilde{\mathbf{x}}$ is the edge-enhanced image and $\tilde{\mathbf{y}}$ is the corresponding class label of image $\tilde{\mathbf{x}}$. These edge-enhanced images are combined with the original training set images to create the input data for the network model. Consequently, each batch sample comprises a total of eight images, contributing to the generation of classification results.

During the forward propagation, we extract the features \mathbf{v}_i by inputting every image \mathbf{x}_i into the Canny edge detection function, denoted as $Canny()$. These

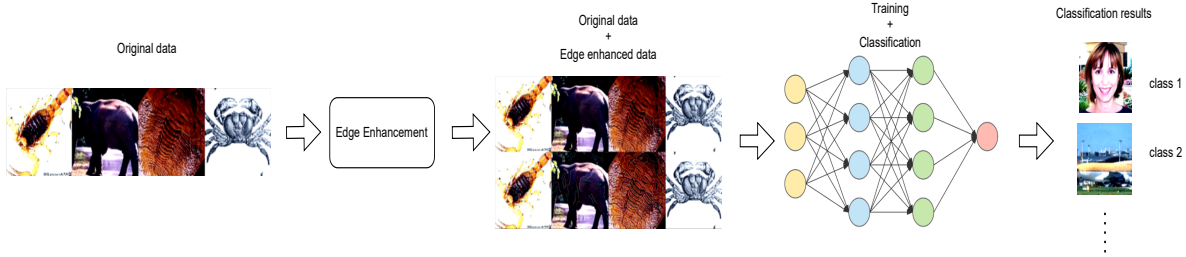


Figure 3: The flowchart of the edge enhancement method for model training. i) Original data is edge enhanced to the transformed data. ii) The original and transformed data is concatenated then feed into the network. iii) Model training. iv) Classification results obtained through test dataset.

features can be expressed as follows:

$$\tilde{\mathbf{x}}_i := \text{Canny}(\mathbf{x}_i) + \mathbf{x}_i \quad (6)$$

During the training phase, the edge-enhanced and original images are both used to train the network by concatenating them together before feeding them to the network. Therefore, every batch size now contains both the edge enhanced images, $\{\tilde{\mathbf{x}}_i\}_{i=1}^b$, and the original images, $\{\mathbf{x}_i\}_{i=1}^b$, where b represents the batch size. After the concatenation we have $\{\mathbf{u}_i\}_{i=1}^{2b}$ because the size of the new batch size is $2b$. We feed every image u_i from the new batch to network f_θ to obtain the feature vector \mathbf{z}_i that corresponds to image \mathbf{u}_i , expressed as

$$\mathbf{z}_i := f_\theta(\mathbf{u}_i) \quad (7)$$

We pass \mathbf{z}_i to g_ϕ so that

$$\mathbf{c}_i := g_\phi(\mathbf{z}_i) \quad (8)$$

where \mathbf{c}_i is the vector of logits of image i . We obtain the probability of each class \hat{p}_i by using the softmax activation function

$$\hat{p}_i := \frac{\exp(\mathbf{c}_i)}{\sum_{i=1}^N \exp(\mathbf{c}_i)}, \quad (9)$$

where N is the total number of classes. With the aid of equation (9), we calculate the cross-entropy loss L_{ce} as

$$L_{ce} := - \sum_{i \in N} \bar{y}_i \log(\hat{p}_i), \quad (10)$$

where \bar{y}_i represents the one hot encoded version of \mathbf{y}_i . Back propagation is then performed to calculate the gradient of L_{ce} with respect to the network weight.

Algorithm 1 summarizes the training procedure of E^2 , and the corresponding flowchart is illustrated in Figure 3

3.4 Inference Stage

During the inference stage, we calculate the probability vector, \hat{p}_i , for each image x_i in the test dataset,

D_{test} by embedding image \mathbf{x}_i using the functions f_θ , g_ϕ and equation (9). To predict the class label \hat{y}_i of image \mathbf{x}_i we employ the following formula:

$$\hat{y}_i := \arg \max_{j \in [N]} \hat{p}_{ij} \quad (11)$$

where N is the number of possible classes and the result corresponds to the position of the element with the highest class probability in the class probability vector p_i .

We calculate the test dataset accuracy using the following formula:

$$\mathbf{S}_c := \text{Acc}(\mathbf{y}_i, \hat{\mathbf{y}}_i) \quad (12)$$

where \mathbf{S}_c is the overall classification accuracy, and Acc is the accuracy function that sums up the numbers of correct comparison results between the prediction and true class labels, we denote the sum as N_{correct} . To scale them into percentage, the results are divided by total number of images and multiplied by a hundred, we express it as $\frac{N_{\text{correct}}}{N_{\text{total}}} \times 100$.

4 EXPERIMENTS

4.1 Setup

Datasets. In our experiments two datasets are used: 1) CIFAR10, which is one of the most popular data sets in image classification research (Recht et al., 2018). It contains a total of 60000 color images of 10 classes, each with dimensions of 32 by 32 pixels. In our experiment we use 45000 images for training, 5000 images for validation and 10000 for testing.

2) CALTECH101, considered as one of the challenging datasets (Bansal et al., 2021), CALTECH101 features higher resolution images in comparison to CIFAR10, as depicted in Figure 1, Figure 2, and Figure 3. This dataset contains 101 classes and includes 8,677 images. For consistency in feeding images into the networks, we resized each image to 256 by 256 pixels and then center-cropped them to 224 by 224

Algorithm 1: Edge enhancement training procedure.

```

input : training dataset  $D_{\text{train}} = \mathbf{x}_i, y_i$ 
output : trained neural network model
1 batch size: 4; if  $D_{\text{train}} \in \text{CALTECH101}$  then
2    $D_{\text{train}} = \text{Resize}(D_{\text{train}}, 256 \times 256)$ ,
3    $D_{\text{train}} = \text{CenterCrop}(D_{\text{train}}, 224 \times 224)$ 
4  $D_{\text{train}} = \text{Normalization}(D_{\text{train}})$  while epoch not finished do
5   for  $i \leq D_{\text{train}}.\text{size}()$  do
6     feature = Canny( $\mathbf{x}_i, \text{sigma} = 3$ ),
7      $\tilde{\mathbf{x}}_i = \text{add}(\mathbf{x}_i, \text{feature})$ ,
8     Inputs = concat( $\mathbf{x}_i, \tilde{\mathbf{x}}_i$ )
9   Model = Model.train(Inputs,  $D_{\text{val}}$ )

```

pixels. Out of these images, 5,205 were used in the training set, 579 for validation, and 2,893 for testing.

Implementation Details. Our implementation is based in Python. We utilized the skimage package for the Canny edge detector and Pytorch for training our neural networks.

Networks. In our work, we used three different neural network architectures tailored to the two datasets to investigate the robustness of our method with different networks and datasets. 1) LeNet-5: This convolutional neural network (CNN) originally proposed by LeCun in 1998 (Lec, 1998), was initially designed for hand written character classification. Given its compatibility with 32 by 32 pixel images and suitability for datasets with small class sizes (10 in CIFAR10), we selected this network for the CIFAR10 dataset, due to its nature of accepting 32 by 32 images and suitable for small class sizes (10 for CIFAR10). 2) CNN-9: A standard neural network model with six convolutional layers and three linear layers, again suitable for 32 by 32 images. 3) ResNet-18: This network, proposed by (He et al., 2016) in 2016. It is designed for larger images of size 224 by 224 pixels. We used this network for the high-resolution CALTECH101 dataset after preprocessing the data to meet the network’s input requirements.

Hyperparameters. In the experiment, the batch size is fixed at 4, meaning 4 labeled images were enhanced with edge information simultaneously. We set the standard deviation σ of the Canny edge detector to 3 and trained the CIFAR10 dataset with LeNet-5 and CNN-9 for 50 epochs. For the CALTECH101 dataset, we trained ResNet-18 for 100 epochs. The loss func-

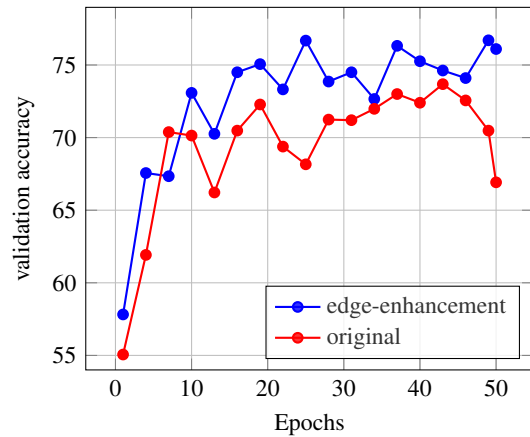


Figure 4: CIFAR10 through CNN-9.

tion used is the cross-entropy loss due to its strong performance in multi-class classification tasks. (Zhou et al., 2019). For weight optimization during network training, we used Stochastic Gradient Descent (SGD) with a momentum of 0.9, weight decay of 0.0005, and learning rate of 0.01. We use the validation set to calculate the validation accuracy and save the model with the highest validation accuracy through comparisons at each epoch.

4.2 Ablation Study

In our experimental setup, for each training dataset and network architecture, we compared our edge enhancement method with the baseline method, which involved training on the original data without any modifications. We assessed this comparison through two aspects: 1) Validation Accuracy Variation: We monitored the validation accuracies for both methods throughout the training process, recording them at each epoch and presenting the results in graphical form. 2) Testing Accuracy: After identifying the models that achieved the best validation results, we tested them on the respective testing datasets to obtain optimal classification accuracies, which were then presented in a tabular format.

1) CIFAR10: Figure 4 and Figure 5 display the results obtained from both LeNet-5 and CNN-9 networks. In these figures, the edge enhancement method is represented by the blue lines, and it consistently outperforms the original method indicated by the red lines. The validation accuracy of the edge enhancement method surpasses that of the original at nearly every epoch. Under LeNet-5, the edge enhancement method achieves its highest accuracy at 55.74%, differing from its lowest accuracy (40.05%) by a substantial 15.69%. In contrast, the original

method exhibits a maximum-minimum difference of 15.24%. Similarly, with CNN-9, the edge enhancement method shows a maximum-minimum difference of 18.86%, compared to the original method's 18.62%. These results clearly indicate that the edge enhancement method consistently provides greater accuracy improvement during the training process. Furthermore, our results reveal a noteworthy trend: for both network architectures within the CIFAR10 dataset, the validation accuracy variations are notably more stable and compact when using the edge enhancement method. CALTECH101: In Figure 6, we present the results obtained through ResNet-18. Here, the edge enhancement method demonstrates a maximum-minimum difference of 44.73%, while the original method yields a value of 40.61%. Once again, these findings underscore the effectiveness of our method in achieving more substantial accuracy improvements.

2) In Table 1, we present the testing accuracy results, where E^2 denotes the edge enhancement method, and 'epoch' indicates the epoch at which the optimal model was achieved. CIFAR10: For the CIFAR10 dataset employing LeNet-5, our method stands out with an impressive 5.7% increase in accuracy over the original approach. Moreover, it achieves this remarkable performance while converging 12 epochs faster to reach an optimal model. In the case of CNN-9 applied to CIFAR10, our method still shines, delivering a substantial 4.03% accuracy improvement. Notably, it accomplishes this while requiring 18 fewer epochs to converge. These outcomes underscore the significant enhancements our method brings to both network architectures when working with the CIFAR10 dataset.

CALTECH101: Through ResNet-18, our method demonstrated a 1.56% improvement over the original method, required 30 less epochs to converge. To conclude, the experiment results are sufficient to show that the proposed edge enhancement method for neural network training provide noticeable classification accuracy improvement and it is more computationally efficient.

4.3 Comparisons with Transformation Methods

To further investigate the effectiveness of our edge enhancement method, we conducted a comprehensive evaluation by integrating it with various geometric transformation techniques. Specifically, we assessed its performance in combination with transformations like random horizontal flipping and random image

size cropping. Our objective was to compare the outcomes with a baseline approach that underwent the same corresponding transformations. We refer to the use of Random Cropping as C and Random Flipping as F.

In this experiment, we use the pytorch built-in transform functions and we split the comparisons within each dataset into three types: 1. exclusively applying random horizontal flipping. 2. solely implementing random image cropping. 3. simultaneously employing both random flipping and random cropping. The image classification accuracy results specified to each dataset section are compared, along with the results obtained without any geometric transformations, as shown in Table 1.

Random Horizontal Flip. The "random horizontal flip" operation involves flipping the input data horizontally with a specified probability. In our experiment, we maintained a default probability of 50 percent for applying this random flipping to the input data.

Random Image Crop. The "random image crop" operation involves selecting a random location within the given input data and then resizing the cropped portion to a specified size. In our experiment, we adapted the resizing process based on the dataset being used: for CIFAR10, we resize the cropped data to 32 by 32 matching the original data size, suitable for LeNet-5 and CNN-9; for CALTECH101, since we use ResNet-18 for this data set, we again resize the cropped data to 224 by 224 fitting the requirement.

As the results shown in Table 2, when applied to the CIFAR10 dataset with network LeNet-5 our edge enhancement method provides significant accuracy improvements over the baseline under every transformation scenario. These improvements range from 3.16% to 3.97% when compared to the baseline. Notably, even when combined with random flipping, our method performs just 0.32% worse than our method without any transformations, as detailed in Table 1. Turning our attention to the Caltech101 dataset with ResNet-18, Table 3 showcases a similar trend. Our edge enhancement method consistently outperforms the original approach for all selected transformation combinations, with improvements ranging from 0.96% to 1.97%. In this case, both edge enhancement with random cropping and edge enhancement with both cropping and flipping surpass the unmodified method (with no transformations) from Table 1 in terms of performance. Furthermore, when considering the impact of these transfor-

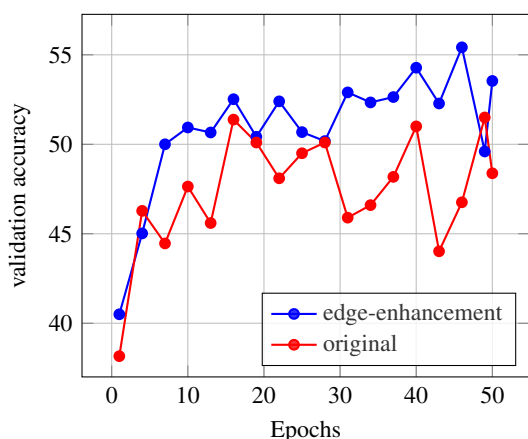


Figure 5: CIFAR10 through LeNet-5.

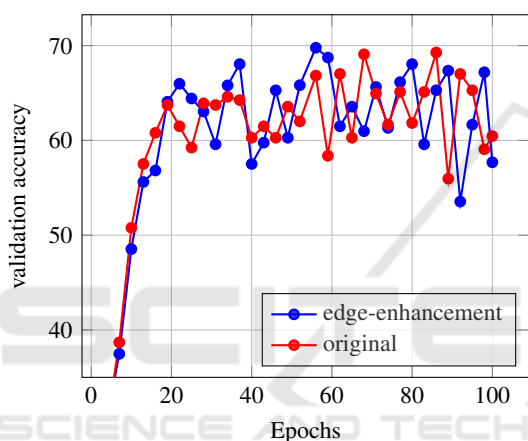


Figure 6: CALTECH101 through ResNet-18.

mation techniques, our results affirm the claim that edge enhancement substantially enhances image classification accuracy. Intriguingly, for both datasets using LeNet-5 and ResNet-18, the original method with random cropping and flipping demonstrates a faster convergence to optimal performance compared to all other results.

Table 1: Testing accuracy results.

METHOD	backbone	accuracy%	epoch
CIFAR10			
Original	LeNet-5	49.26	45
E ²	LeNet-5	54.96	33
Original	CNN-9	72.61	43
E ²	CNN-9	76.64	25
CALTECH101			
Original	ResNet-18	72.83	86
E ²	ResNet-18	74.39	56

Table 2: CIFAR10 transformation testing accuracy results.

METHOD	backbone	accuracy%	epoch
CIFAR10			
Original + F	LeNet-5	50.67	45
E ² + F	LeNet-5	54.64	32
Original + C	LeNet-5	46.74	29
E ² + C	LeNet-5	49.90	17
Original + F+C	LeNet-5	44.59	16
E ² + F+C	LeNet-5	48.54	49

Table 3: CALTECH101 transformation testing accuracy results.

METHOD	backbone	accuracy%	epoch
CALTECH101			
Original + F	ResNet-18	71.00	80
E ² + F	ResNet-18	71.96	53
Original + C	ResNet-18	79.23	88
E ² + C	ResNet-18	81.20	58
Original + C+F	ResNet-18	78.40	89
E ² + F+C	ResNet-18	79.81	92

5 CONCLUSION

In our research, we introduced a novel approach for enhancing image classification accuracy and computational efficiency. Our method, based on the high boost filtering principle, focuses on edge enhancement. This technique leverages the edge information from labeled data to transform the original dataset, thereby increasing both the size and diversity of the training samples. Through extensive experiments on popular datasets such as CIFAR-10 and CALTECH-101, using well-known neural network architectures including LeNet-5, CNN-9, and ResNet-18, we have demonstrated the effectiveness of our proposed method. Our results substantiate the validity of our hypothesis. Regarding our future research directions, we are interested in exploring the impact of data augmentation techniques in conjunction with our current state-of-the-art approach; additionally, scaling our method to larger datasets such as ImageNet and other domains as semi-supervised learning would be another potential research direction. Our goal is to further enhance classification accuracy while at the same time improving the training efficiency, pushing the boundaries of what is achievable in image classification.

ACKNOWLEDGEMENTS

In the loving memory of TianHao Bu's father BingXin Bu: 10/01/1965 - 13/09/2023.

REFERENCES

- (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- Bansal, M., Kumar, M., Sachdeva, M., and Mittal, A. (2021). Transfer learning for image classification using vgg19: Caltech-101 image data set. *Journal of ambient intelligence and humanized computing*, pages 1–12.
- Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698.
- Chen, X. and Wang, G. (2021). Few-shot learning by integrating spatial and frequency representation. In *2021 18th Conference on Robots and Vision (CRV)*, pages 49–56. IEEE.
- Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., and Haworth, A. (2021). A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, 65(5):545–563.
- Deng, G. and Cahill, L. (1993). An adaptive gaussian filter for noise reduction and edge detection. In *1993 IEEE conference record nuclear science symposium and medical imaging conference*, pages 1615–1619. IEEE.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- Kim, J.-H., Choo, W., and Song, H. O. (2020). Puzzle mix: Exploiting saliency and local statistics for optimal mixup. In *International Conference on Machine Learning*, pages 5275–5285. PMLR.
- Liu, Z., Li, S., Wu, D., Liu, Z., Chen, Z., Wu, L., and Li, S. Z. (2022). Automix: Unveiling the power of mixup for stronger classifiers. In *European Conference on Computer Vision*, pages 441–458. Springer.
- Mukai, K., Kumano, S., and Yamasaki, T. (2022). Improving robustness to out-of-distribution data by frequency-based augmentation. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 3116–3120. IEEE.
- Recht, B., Roelofs, R., Schmidt, L., and Shankar, V. (2018). Do cifar-10 classifiers generalize to cifar-10? *arXiv preprint arXiv:1806.00451*.
- Shao, S., Wang, Y., Liu, B., Liu, W., Wang, Y., and Liu, B. (2023). Fads: Fourier-augmentation based data-shunting for few-shot classification. *IEEE Transactions on Circuits and Systems for Video Technology*.
- Srivastava, R., Gupta, J., Parthasarthy, H., and Srivastava, S. (2009). Pde based unsharp masking, crispening and high boost filtering of digital images. In *Contemporary Computing: Second International Conference, IC3 2009, Noida, India, August 17-19, 2009. Proceedings 2*, pages 8–13. Springer.
- Yin, W., Wang, H., Qu, J., and Xiong, C. (2021). Batch-mixup: Improving training by interpolating hidden states of the entire mini-batch. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4908–4912.
- Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., and Yoo, Y. (2019). Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6023–6032.
- Zhang, H., Cisse, M., Dauphin, Y. N., and Lopez-Paz, D. (2017). mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*.
- Zhang, H., Zhang, L., and Jiang, Y. (2019). Overfitting and underfitting analysis for deep learning based end-to-end communication systems. In *2019 11th international conference on wireless communications and signal processing (WCSP)*, pages 1–6. IEEE.
- Zhou, Y., Wang, X., Zhang, M., Zhu, J., Zheng, R., and Wu, Q. (2019). Mpce: a maximum probability based cross entropy loss function for neural network classification. *IEEE Access*, 7:146331–146341.