# LMix:Regularization Strategy for Convolutional Neural Networks

**Linyu Yan**
Hubei University of Technology
**Kunpeng Zheng** ( zhengfreeking@163.com )
Hubei University of Technology
**Jinyao Xia**
Hubei University of Technology
**Ke Li**
Hubei University of Technology
**Hefei Ling**
Huazhong University of Science and Technology

**Research Article**

**Keywords:** Mixup, Data Augmentation

**Posted Date:** May 12th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1630095/v1

# LMix:Regularization Strategy for Convolutional Neural Networks

Linyu Yan[1], Kunpeng Zheng[1*], Jinyao Xia[1], Ke Li[1] and Hefei Ling[2]

[1*]Hubei University of Technology, wuhan, HuBei, China.
[2]Huazhong University of Science and Technology, wuhan, hubei, China.

*Corresponding author(s). E-mail(s): zhengfreeking@163.com;
Contributing authors: 361953203@qq.com; 2464274209@qq.com;
1028132487@qq.com;

**Abstract**

Convolutional neural network models, as well as the training samples necessary, have grown in size in recent years. Mixed Sample Data Augmentation is provided to further improve the model's performance, and it has yielded good results.Mixed Sample Data Augmentation allows the network to generalize more effectively and improves the baseline performance of the model. The mixed sample strategies proposed so far can be broadly classified into interpolation and masking. However, interpolation-based strategies distort the data distribution, while masking-based strategies can obscure too much information. Although Mixed Sample Data Augmentation has been proven to be a viable technique for boosting deep convolutional model baseline performance, generalization ability, and robustness, there is still room for improvement in terms of image local consistency and data distribution.In this research, we present a new Mixed Sample Data Augmentation that uses random masking to increase the number of image masks while retaining the data distribution and high-frequency filtering to sharpen the images in order to emphasize recognition regions. Our experiments on CIFAR-10, CIFAR-100, Fashion-MNIST, SVHN, and Tiny-ImageNet datasets show that the LMix improves the generalization ability of state-of-the-art neural network architectures. And our method enhances the robustness of adversarial samples.

**Keywords:** Mixup, Data Augmentation

# 1 Introduction

Deep convolutional neural networks have shone in various computer vision tasks, such as image classification [1, 2, 5], object detection [29, 30], semantic segmentation [14], and image super-resolution [15]. Deep convolutional neural networks follow the empirical risk minimization principle[16] to minimize the average error when performing training. Also, when the deep convolutional neural network is used to extract features from an input image, the larger the training sample, the greater the learning effect and generalization capacity of the model.For instance, the network of Pierre Foret et al[1]. was modeled using the JFT-300M dataset [31] with 4.8 billion parameters. Dhruv Mahajan et al. [2] used the ImageNet-22k dataset to model their network, which has 8.2 billion parameters. Tan et al[3]. used the ImageNet-22k dataset to model their network, which has 1.2 billion parameters.To further improve the training accuracy and speed, many scholars have proposed some training strategies, such as regularization techniques, data augmentation strategies [6, 8–10], etc. The regularization technique prevents overfitting in networks with more parameters than input data, as well as algorithmic generalization by avoiding training coefficients of perfect-fit data samples.Data augmentation can prevent model overfitting and increase the number of samples to improve model generalization, mainly including geometric space change, pixel color transformation, and multiple sample fusion.

(a) LMix. (b) Mixup. (c) CutMix.

**Fig. 1** Generated images of CutMix, LMix, and Mixup on the CIFAR dataset

Currently, mixed sample data augmentation [8–12] techniques based on Vicinal Risk Minimisation [13] have obtained good results in a variety of applications, particularly classification tasks. The Vicinal Risk Minimisation based data augmentation approach extracts additional dummy samples from the training samples to boost support for the training distribution.This also leads to the goal of expanding samples to increase data space without distorting the data distribution; nevertheless, larger samples unavoidably have distorted data distribution.To ensure that the data enhancement strategy can produce good results for the network, the following characteristics should be maintained: the virtual samples and the real samples should have a good acquaintance; the data augmentation strategy can improve the model's generalization ability; and the data augmentation strategy can improve the model's robustness against noise.

Mixed Sample Data Augmentation is the modification of sample data to build an extended dataset for training models. Mixed Sample Data Augmentation proposed so far is broadly classified into two types: interpolation and masking.Mixed Sample Data Augmentation for Interpolation has Mixup [8], which is a Mixed Sample Data Augmentation based on the idea of Vicinal Risk Minimisation, and Mixup suggests a general vicinal distribution, the mixed distribution, as illustrated in Figure 1(b).Mixed Sample Data Augmentation for masking has CutMix[9], which proposes patches are cut and pasted among training images where the ground truth labels are also mixed proportionally to the area of the patches, as illustrated in Figure 1(c). Both strategies improve the baseline performance of the deep convolutional model. In terms of picture data distribution, CutMix trumps Mixup.

In this paper, we propose a new Mixed Sample Data Augmentation LMix, as illustrated in Figure 1(a). The main ideas are as follows: (1) use random masking to increase the number of image masks while effectively ensuring the local consistency of the image. (2) use high frequency filtering to sharpen the image to highlight the recognition area. The rest of this paper is organized as follows. In Section 2, we review the existing work on data enhancement strategies. Then, we present the implementation of the LMix algorithm in Section 3. In Section 4, we conduct a large number of experiments to demonstrate the effectiveness and efficiency of the proposed algorithm. Finally, we draw conclusions in Section 5.

## 2 Related Work

**Data Augmentation:** As the deep network deepens, the number of learning parameters required increases, which inevitably leads to overfitting. When the dataset is too small, too many parameters can fit all the characteristics of the dataset rather than the commonalities between the data [25, 26]. Data augmentation generates virtual samples from real samples to expand the dataset size, which can alleviate the model overfitting problem and make the training data as close as possible to the test data, thus improving the accuracy. At the same time, data augmentation can force the model to improve robustness and make the model more generalizable.Early data augmentation algorithms were transformations of images using geometric transformations including flip, rotate, crop, distort, scale, etc., and color transformations including noise, blur, color transformation, erase, fill, etc.Lopes et al. [4] added Gaussian blocks to Cutout to make the model more stable without losing model accuracy by adding noise to randomly selected blocks in the input image. Also, this method can be used in combination with other regularization methods and data enhancement strategies.He et al. [5] trained the deep residual network with random left-right flipping and cropping of the image data to improve the generalization ability of the model. This allowed the data samples to be expanded and greatly improved the generalization ability of the model.DeVries et al. [6] proposed masking regularization, a data augmentation approach comparable to random erasure.

They apply random masking on the image, masking it with a fixed-size rectangle. Within the rectangle, all values are set to 0 or other solid color values, and the erased rectangular section may or may not be totally in the picture.Taylor and Nitschke [7] analyzed the effectiveness of geometric and photometric (color space) transformations. They analyzed geometric changes such as flipping, as well as color space transformations such as color dithering (random color manipulation), edge improvement, and principal component analysis. Simply conducting simple image processing on individual photographs might lead to a slew of issues. For instance, operations such as flip, shear, and rotate are not safe for the dataset [27], while the color transformation enhancement approach is biased from a color space perspective with more diversity of color variations, resulting in insufficient enhancement and poor learning and underfitting of the color space, while the transformation is unsafe.

**Mixup:** Data augmentation not only has good generalization ability but also has excellent robustness, both for data containing noisy labels and against sample attacks. The fused images obtained by the multisample fusion approach are difficult to understand from the human perspective, yet the experimental results are excellent. Sample fusion proposes many data augmentation strategies to improve the model's accuracy and generalization capabilities in order to improve the model's baseline performance.Zhang et al. [8] proposed Mixup, a method for mixing two images, as a data-agnostic and simple data-expansion method. By performing a simple random weighted summation of two random samples from the real sample, while the labels of the samples are weighted summation correspondingly, the prediction results are lost with the labels after the weighted summation, and the parameters are updated in the reverse derivative. However, the mixup can distort the data distribution of the image, while the generated virtual samples are not very interpretable.Yun et al. [9] proposed Cutmix, Mixed Sample Data Augmentation for Masking. Two samples are randomly selected from real samples, a rectangular cut box is randomly generated to crop the corresponding position of one image, and then the corresponding position of another image is placed at the position where the image is cropped to generate a new sample, and the loss function is calculated using a weighted summation. However, using the regular cropping method can cause the image to lose a lot of information.Vera et al. [10] proposed an extension of input data blending to include intermediate hidden layer output mixing. The approach is designed to modify the input data in a smoother and more uniform manner, resulting in increased model performance and generalization. Kim et al. [11] proposed a method based on significance and local statistics for the given data. They added significance analysis to CutMix. The significance regions of individual samples are first calculated, only the significance regions are cropped, and then some complex optimization operations are added.Harris et al. [12] proposed an improved method based on CutMix, and they verified that the masking mixing method is more advantageous than the interpolated mixing method in terms of preserving image data distribution. They also designed an irregular mask to mask the image as the spatial size of the data samples increased.

# 3 Method

We discover that masked mixed sample data augmentation is more successful than interpolated mixed sample data augmentation in preserving data distribution, which is especially noticeable for convolutional neural networks. Convolutionalneural networks are locally consistent, which means that each neuron is only linked to one portion of the input neuron at a specified geographical position. Neurons are locally linked in the spatial dimension but completely connected in the depth dimension in picture convolution procedures.Local pixel connections are also stronger in the two-dimensional picture itself. This local connectedness guarantees that the learnt filter responds to the local input characteristics as strongly as possible. It is extremely critical for neural networks to successfully preserve the pictures' local consistency. Meanwhile, the interpolative mixed sample data augmentation has a weakness in that the number of masks is not fully guaranteed when just regular masks are used to act on the picture, thus we must increase the number of masks while keeping local consistency.

**Fig. 2** Virtual sample of sample fusion acquired from CIFAR-100

In this section, we propose LMix, mixed sample data augmentation that provides the greatest results in terms of local consistency and the number of masks in the picture, as shown in Figure 2. LMix employs an masking mixed sample data augmentation to preserve the image's local consistency.

Its algorithm is implemented as:

Let $x \in R^{W \times H \times C}$ denote a training set, $y$ denote the training set's label, and $(x_A, y_A)$ and $(x_B, y_B)$ represent two feature target vectors chosen at random from the training data. LMix's purpose is to create a new sample $(\hat{x}, \hat{y})$ by merging two training samples $(x_A, y_A)$ and$(x_B, y_B)$. The resulting training sample$(\hat{x}, \hat{y})$ is utilized to train the original loss function-trained model. It is defined as follows.

$$\hat{x} = mask \cdot x_A + (1 - mask) \cdot x_B \tag{1}$$

$$\hat{y} = \lambda y_A + (1 - \lambda) y_B \tag{2}$$

where $mask \in \{0,1\}^{W \times H}$ is the binary mask, which refers to the bits deleted and filled from the two pictures, and $'1'$ represents the binary mask filled with 1. As in Mixup [8], the combined ratio $\lambda$ between the two data points is obeying the $Beta(\alpha, \alpha)$ distribution.Compared to Cutmix [9],which directly intercepts a regular patch from a image to replace the image region of the target image, we use a mask made by combining a single region and its adjacent regions, which can reduce the number of binary mask conversions.To obtain the binary mask , we first apply the two-dimensional discrete Fourier transform to the image to convert it from the spatial domain to the frequency domain, and then obtain the low-frequency image of the image, the high-frequency component of the image through the low-frequency component, and the high-frequency filtered and enhanced image of the image through the high-frequency component.We define $Z$ as a complex random variable with a value domain of $Z = C^{H \times W}$ and densities of $P_{R(Z)} = N(0, I_{W \times H})$ and $P_{I(Z)} = N(0, I_{W \times H})$. The real and imaginary components of the input are denoted by $R(Z)$ and $I(Z)$, respectively. The amplitude of the sampling frequency corresponding to the $Ii, j$th bin of the two-dimensional discrete Fourier transform is denoted by $freq(w, h)[i, j]$. By attenuating the high-frequency component of $Z$, a low-pass filter is created. The two-dimensional discrete Fourier transform is employed in particular for a specified attenuation power. By attenuating the high-frequency component of $Z$, we may create a low-pass filter.

$$f_{LP}(z, \delta)[i, j] = \frac{z[i, j]}{freq(w, h)[i, j]} \tag{3}$$

The image high-pass filter is obtained by the obtained low-pass filter.

$$f_{HP}(z, \delta) = 1 - f_{LP}(z, \delta) \tag{4}$$

The sharpened image is then obtained by passing it through a high-pass filter.

$$g_{mask}(x, y) = f^{-1}\{[1 + k \cdot f_{HP}(z, \delta)]F(u, v)\} \tag{5}$$

where $g_{mask}(x, y)$ denotes the high-frequency filtered enhanced image,$F(u, v)$ denotes the Fourier variation of the original image,$F_{HP}(z, \delta)$ denotes the high-pass filter, $\delta$ is the given attenuation frequency, and $f^{-1}$ denotes the discrete Fourier inverse transform.Finally obtaining the sampled binary mask mask now all that remains is to convert the grayscale image to a binary mask such that the average is some given $\lambda$. Let $top(n, x)$ return a set containing the top n elements of the input $x$. Setting the value of the top $\lambda, w, h$ elements of some grayscale image $g$ to $'1'$ and the value of all other elements to $'0'$, we obtain a binary mask with an average $\lambda$.

$$mask(\lambda, g)[i, j] = \begin{cases} 1, if g_{[i,j]} \in top(\lambda wh, g) \\ 0, otherwise \end{cases} \tag{6}$$

We first sample a random complex tensor whose real and imaginary components are both independent and Gaussian distributed. Then, each component is scaled according to its frequency by the parameter $\delta$, such that higher $\delta$ values correspond to increased attenuation of high-frequency information. Next, the Fourier inverse transform of the complex tensor is performed and its real part is taken to obtain a grayscale image. Finally, the top scale of the image is set to $'1'$ and the rest of the scale to $'0'$ to obtain the binary mask.

# 4 Experiment

In this section, we apply LMix to ResNet [19], DenseNet [21], and WideResNet [20] models on the CIFAR-10, CIFAR-100 [17], Fashion-MNIST, SVHN, and Tiny-ImageNet [18] datasets for image classification tasks to evaluate the enhancements and generalization improvements to the model baseline that LMix can provide. To compare and assess the performance of several mixed sample data augmentation for boosting the generalization effect and strengthening the baseline, the same hyperparameters were utilized for all models. In addition, the settings of the mixed-sample data augmentation techniques that provided the best results in the respective publications were picked for all of them. We replicate all studies when possible and publish the average performance and standard deviation following the last phase of training. In all tables, we highlight the best outcomes and those that are within their margin of error. The uncertainty estimate is the standard deviation of 5 replicates.

## 4.1 Image Classification

### 4.1.1 CIFAR Classification

This section first discusses the results of the image classification task on the CIFAR 10/100 dataset. On the CIFAR dataset we train: the PreAct-ResNet18 [19], WideResNet-28-10 [20], DenseNet-BC-190 [21], and PyramidNet-272-200 [22] models. We found that the regularization methods including cutout [6], mixup [8], CutMix [9], and FMix [12] need a longer training time to reach convergence. As a result, we set the epoch of all models to 300, the initial learning rate to 0.1, and decay at 75, 150, and 225 epochs in multiples of 0.1, with a batch size of 128. Table 1 compares the performance of the approach to that of other cutting-edge data augmentation and regularization methods. All trials were repeated five times, and the best performance during training is presented as the average.

**Hyperparameter setting:** We set the hyperparameter $\alpha$ of LMix to 1 and the decay rate $\delta$ to 3. Set the cropping area of Cutmix [9] and Cutout [6] to 16×16. For mixup, we set the hyperparameter $\alpha$ to 1, set the hyperparameter $\alpha$ and decay rate $\delta$ of FMix [12] to 1 and 3, and the hyperparameters $\alpha$ of Patchup [10],$Patchup_{prob}$,$x$, and block size are set to 2, 0.7, 0.5, and 7, respectively.

**LMix is applicable to a variety of models:**As shown in Table I, LMix applies to various convolutional neural networks, while LMix significantly

**Table 1** The accuracy of the approach for the image classification task in CIFAR-10 using the PreAct-ResNet-18, WideResNet-28-10, DenseNet-BC-190 (Dense), and PyramidNet-272-200 models.

| Data set | Model | Baseline | FMix | Mixup | CutMix | LMix |
|---|---|---|---|---|---|---|
| CIFAR-10 | ResNet-18 | 94.62 | 96.14 | 95.67 | 95.97 | 96.32 |
| | WRN | 95.32 | 96.41 | 96.69 | 96.63 | 96.58 |
| | Dense | 96.32 | 97.30 | 97.01 | 96.95 | 97.36 |
| | Pyrmid | 97.31 | 98.54 | 97.90 | 98.18 | 98.58 |

**Table 2** Accuracy of the algorithm using PreAct-ResNet-18,PreAct-ResNet-34, WideResNet-28-10, and DenseNet-BC-190 models to test the algorithm for the image classification task in CIFAR-100.

| Data set | Model | Baseline | FMix | Mixup | CutMix | LMix |
|---|---|---|---|---|---|---|
| CIFAR-100 | ResNet-18 | 75.12 | 79.65 | 77.34 | 78.48 | 79.85 |
| | ResNet-34 | 76.58 | 81.55 | 78.54 | 79.56 | 81.86 |
| | Dense | 78.24 | 82.03 | 81.95 | 81.84 | 81.91 |
| | Pyrmid | 81.64 | 83.75 | 83.23 | 82.69 | 83.95 |

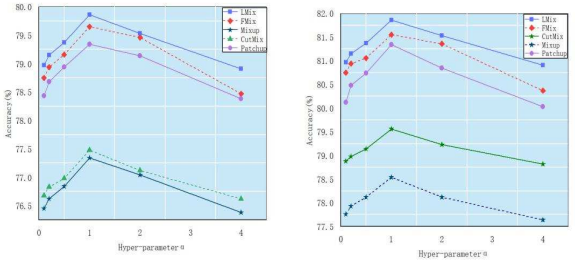


**Fig. 3** Effect of varying the value of hyperparameter $\alpha$ on the baseline accuracy of various algorithms under the CIRAF-100 data set.

improves the baseline performance of various lightweight models, and for the ResNet-18 [19] model, LMix improves the most accuracy over the baseline performance by 1.51% and on the average accuracy over the baseline performance by 1.68%. For the WideResNet-28 [20] model, LMix improves 1.23% over the maximum accuracy of the baseline performance and 1.29% over the average accuracy of the baseline performance. For the DenseNet [21] model, LMix improved the maximum accuracy over baseline performance by 1.05% and the average accuracy over baseline performance by 1.11%. For the Pyramid [22] model LMix improved the maximum accuracy over the baseline performance by 1.32% and the average accuracy over the baseline performance by 1.33%.

**LMix performance on CIFAR-10/100:**The results in Table II show that the same models were trained on the CIFAR-10 dataset, and LMix provided significant improvements over the other hybrid sample enhancement algorithms. For ResNet-18, LMix outperforms cutout by 1.16%, Mixup by 0.72%, Cutmix by 0.42%, FMix by 0.29%, and patchup by 0.62% in terms of accuracy for the image classification task. LMix also performs very well on the CIFAR-100 dataset, as shown in Table II. For ResNet-18, LMix outperformed the baseline by 4.73%, outperformed FMix by 0.2%, outperformed CutMix by 0.37%, and outperformed Mixup by 2.51% on the image classification task.

The results obtained in Figure 3 indicate that LMix has the highest accuracy for the image classification task trained with ResNet-18 on CIFAR-100 with hyperparameter $\alpha = 1$, while outperforming Mixup, CutMix, and FMix. we have explored the performance of LMix for ResNet-18 and ResNet-34 on CIFAR-100. As shown in Figure 4, we found an improvement in accuracy for both classification tasks.

### 4.1.2 Tiny-ImageNet

We trained the PreAct-ResNet18 network on the Tiny-ImageNet [18] dataset, which contains 200 classes with 500 training images and 50 test images per class with a resolution of $64 \times 64$ . We trained the model with an initial learning rate of 0.1 for 200 epochs, and we used a decay learning rate of 0.1 at 150 and 180 epochs. we set the momentum to 0.9. In the case of mixup weights $\lambda$, for the mixup, we set $\alpha = 1$ as described in the mixup. For CutMix, we chose $\alpha = 1$, which is the best performance in [0.2,0.5,1.0], while for FMix, we chose $\alpha = 1.0$, for Cutout and CutMix with a cropping region of $16 \times 16$. In the experiments

**Table 3** Accuracy of the image classification task using PreAct-ResNet-18 test algorithm in Tiny-ImageNet.

| Model | MaxAcc(% ) | Acc(% ) |
|---|---|---|
| Baseline | 55.94 | 55.86 |
| +CutMix | 64.08 | 63.84 |
| +FMix | 63.33 | 62.23 |
| +Mixup | 61.96 | 61.89 |
| +LMix | 64.16 | 63.92 |

using the Tiny-ImageNet dataset, compared with other hybrid baselines, LMix showed significant improvements in generalization performance and improved model accuracy (Table III). With the same number of epochs trained, LMix achieves an accuracy of 64.06%, which is 0.08% higher than the strongest baseline.

### 4.1.3 Fashion-MNIST

We train the PreAct-ResNet18 network on the Fashion-MNIST dataset, a fashion product dataset containing 70,000 28×28 grayscale images in 10 categories with 7,000 images in each category. The training set has 60,000 images and the test set has 10,000 images. Fashion MNIST shares the same image size, data format, and training and test splitting structure as the original MNIST. We trained the PreAct-ResNet18 [19] model, where we trained the model with an initial learning rate of 0.1 for 200 epochs, and we used a decay learning rate of 0.1 at 150 and 180 epochs. we set the momentum to 0.9. in the case of mixup weights $\lambda$, for mixup, we set $\alpha = 1$ in mixup. Set the cropping area for Cutout and CutMix to $16 \times 16$.

**Table 4** Training PreAct-ResNet18 on the Fashion-MNIST dataset to evaluate LMix.

| Model | MaxAcc(% ) | Acc(% ) |
|---|---|---|
| Baseline | 95.70 | 95.52 |
| +CutMix | 96.02 | 95.93 |
| +Mixup | 96.26 | 96.20 |
| +LMix | 96.64 | 96.62 |

In the experiments using the Fashion-MNIST dataset, compared with other hybrid baselines, LMix showed significant improvements in generalization performance and improved model accuracy (Table IV). With the same number of epochs trained, LMix achieves an accuracy of 96.62%, which is 1.1% higher than the strongest baseline.

### 4.1.4 SVHN

We train multiple image classification network models on the SVHN dataset, a numerical classification benchmark dataset containing 600,000 32×32 RGB images of printed digits (from 0 to 9) cropped from door sign images. The cropped images are centered on the digit of interest, but nearby digits and other distractors are retained in the images.SVHN has three sets: a training set, a test set, and an additional set containing 530,000 images that are less difficult and can be used to aid in the training process. To evaluate the effect of LMix on the SVHN dataset, we applied LMix on PreAct-ResNet18,PreAct-ResNet34 and WideResNet-28-10, respectively. We set the epoch of the model to 300, the initial learning rate to 0.1, and decay at epochs of 75, 150, and 225 in multiples of 0.1, and set the batch size to 128. Also, we repeated the experiments several times to obtain the most reliable results.

LMix performance in SVHN: The results in Table V show that the same model is trained on the SVHN dataset and LMix provides significant improvements over other mixed-sample enhancement algorithms. For ResNet-18, LMix provides 0.48% higher accuracy than Mixup and 0.44% higher than Cutmix

**Table 5** Accuracy of the algorithm using PreAct-ResNet-18,PreAct-ResNet-34, WideResNet-28-10, and DenseNet-BC-190 models to test the algorithm for the image classification task in CIFAR-100.

| Data set | Model | Baseline | Mixup | CutMix | LMix |
|---|---|---|---|---|---|
| SVHN | ResNet-18 | 96.53 | 96.63 | 96.57 | 97.01 |
| | ResNet-34 | 97.04 | 97.21 | 97.44 | 97.66 |
| | WRN | 97.28 | 97.48 | 97.69 | 97.73 |

for the image classification task. Also when ResNet-34 and WideResNet-28-10 are applied, there is a good improvement in the accuracy and generalization of the model.

## 4.2 Combining MSDAs

We trained the PreAct-ResNet-18 network on the CIFAR-100 dataset and used it to evaluate the effect of the algorithm combination. We train 300 epoch models with an initial learning rate of 0.1, and we use a decay learning rate of 0.1 at 100, 150, and 225 epochs, with batch size set to 128. for Mixup, we set the hyperparameter $\alpha$ to 1. We also set the hyperparameters $\alpha$and $\delta$ of LMix to 1 and 3, respectively. we set the hyperparameters $\alpha$ and *delta* of LMix+ The hyperparameter *alpha* is set to 1 for the Mixup combination. As shown in Figure 4, the accuracy of LMix for the image classification task

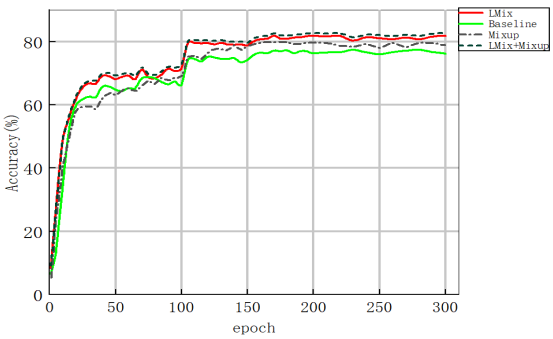


**Fig. 4** Training PreAcResNet-34 on the CIFAR-100 dataset

with the PreAct-ResNet-34 model trained under the CIFAR-100 dataset is significantly higher than the baseline performance of Mixup and the original model, while the combined approach further improves the accuracy of the model after combining LMix with Mixup.

### 4.3 Robustness

When performing image classification tasks, the neural network is first trained and then minimized concerning the error on the training sample, a learning rule called empirical risk minimization. This learning rule is called empirical risk minimization. Small changes in the data samples can have a substantial influence on the model's performance. To achieve quicker computing speed, most contemporary neural networks build the model in a linear model, resulting in a relatively poor battle against disturbing data. Certain data-dependent regularization strategies, such as interpolating the data to train the model, might lessen the fragility of the adversarial cases. Therefore, the robustness of the regularization model to adversarial instances can be used as a criterion for comparison. For FSGM [23], it is necessary to perform adversarial sample production of the original image $x$, its label $y$, a good classification model $M$, the parameters x of the classification model $M$, and also to generate an attack noise $\eta$ using FGSM.

$$\eta = \epsilon sign(\bigtriangledown_x J(\theta, x, y)) \tag{7}$$

where $J(\theta, x, y)$ denotes the loss function of the model with parameter $\theta$and $\epsilon$ denotes the control perturbation. The symbolic function $sign()$ denotes the direction of the extracted gradient, $x$ denotes the original sample and $y$ denotes the true label of $M$. Subsequently, the original image is added with the attack noise $\eta$ to obtain the adversarial sample $\hat{x}$ of the original image $x$

$$\hat{x} = x + \eta \tag{8}$$

To evaluate the robustness of LMix against adversarial attacks, we compare

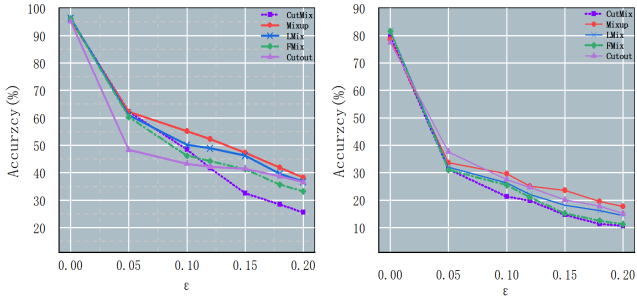


**Fig. 5** Robust FGSM attack, (a) comparison at CIFAR-10 using the PreAc-tResNet-18 model, (b) comparison at CIFAR-100 using the PreAct-ResNet-34 model.

the performance of PreAct-ResNet-18 and PreAct-ResNet-34 on CIFAR-10 and CIFAR-100 with the adversarial examples generated by the FGSM attack described in. Our experiments show that LMix is effective against attacks

in most cases. Figure 5 shows a comparison of the impact of state-of-the-art regularization techniques on the robustness of the model against FGSM attacks.

# 5 Conclusion

In this paper, we introduce LMix, Mixed Sample Data Augmentation which improves the classification performance and generalization ability of a model. The model is improved by preserving the local consistency of the images and then maximizing the number of masks. We run a series of analyses to ensure the feasibility of the idea and then design preliminary experiments and find that LMix performs very well on the classification task. It is 1.51% above the baseline on the CIFAR dataset, yielding an optimal result of 96.35% while outperforming other regularization methods in the same situation in terms of classification accuracy. Our experimental results show that LMix excels in both generalization performance and robustness against interference.

**Acknowledgments.** This work is funded by the National Natural Science Foundation of China under Grant No. 61772180, the Key R D plan of Hubei Province (2020BHB004, 2020BAB012).

# References

[1] Foret P , Kleiner A , Mobahi H , et al. Sharpness-Aware Minimization for Efficiently Improving Generalization[J]. 2020.

[2] D Mahajan, Girshick R , Ramanathan V , et al. Exploring the Limits of Weakly Supervised Pretraining[J]. Springer, Cham, 2018.

[3] Tan M , Le Q V . EfficientNetV2: Smaller Models and Faster Training[J]. 2021.

[4] Lopes R G , Yin D , Poole B , et al. Improving Robustness Without Sacrificing Accuracy with Patch Gaussian Augmentation[J]. 2019.

[5] He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition[J]. IEEE, 2016.

[6] Devries T , Taylor G W . Improved Regularization of Convolutional Neural Networks with Cutout[J]. 2017.

[7] Taylor L , Nitschke G . Improving Deep Learning using Generic Data Augmentation[J]. 2017.

[8] Zhang H , Cisse M , Dauphin Y N , et al. mixup: Beyond Empirical Risk Minimization[J]. 2017.

[9] Yun S , Han D , Chun S , et al. CutMix: Regularization Strategy to Train Strong Classifiers With Localizable Features[C]// International Conference on Computer Vision. 0.

[10] Verma V , Lamb A , Beckham C , et al. Manifold Mixup: Better Representations by Interpolating Hidden States[J]. 2018.

[11] Kim J H , Choo W , Song H O . Puzzle Mix: Exploiting Saliency and Local Statistics for Optimal Mixup[J]. 2020.

[12] Harris E , Marcu A , Painter M , et al. FMix: Enhancing Mixed Sample Data Augmentation[J]. 2020.

[13] Chapelle O , Weston J , Léon Bottou, et al. Vicinal risk minimization. 2000.

[14] Yuan Y , Chen X , Wang J . Object-Contextual Representations for Semantic Segmentation[C]// European Conference on Computer Vision. Springer, Cham, 2020.

[15] Dong C , Loy C C , He K , et al. Image Super-Resolution Using Deep Convolutional Networks[J]. IEEE Trans Pattern Anal Mach Intell, 2016, 38(2):295-307.

[16] Vapnik V . Statistical Learning Theory[M]. DBLP, 1998.

[17] Krizhevsky A , Hinton G . Learning multiple layers of features from tiny images[J]. Handbook of Systemic Autoimmune Diseases, 2009, 1(4).

[18] H Pouransari, S Ghili. Tiny ImageNet Visual Recognition Challenge.

[19] Leibe B , Matas J , N Sebe, et al. [Lecture Notes in Computer Science] Computer Vision – ECCV 2016 Volume 9908 Identity Mappings in Deep Residual Networks[J]. 2016, 10.1007/978-3-319-46493-0(Chapter 38):630-645.

[20] Zagoruyko S , Komodakis N . Wide Residual Networks[J]. 2016.

[21] Huang G , Liu Z , Laurens V , et al. Densely Connected Convolutional Networks[J]. IEEE Computer Society, 2016.

[22] Han D , Kim J , Kim J . Deep Pyramidal Residual Networks[J]. 2016.

[23] Goodfellow I J , Shlens J , Szegedy C . Explaining and harnessing adversarial examples[C]// ICML. 2015.

[24] Technicolor T , Related S , Technicolor T , et al. ImageNet Classification with Deep Convolutional Neural Networks [50].

[25] Srivastava N , Hinton G , Krizhevsky A , et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J]. Journal of Machine Learning Research, 2014, 15(1):1929-1958.

[26] Zhang C , Be Ngio S , Hardt M , et al. Understanding deep learning requires rethinking generalization[C]// 2016.

[27] Shorten C , Khoshgoftaar T M . A survey on Image Data Augmentation for Deep Learning[J]. Journal of Big Data, 2019, 6(1).

[28] Touvron H , Vedaldi A , Douze M , et al. Fixing the train-test resolution discrepancy: FixEfficientNet[J]. 2020.

[29] He K , Girshick R , Dollar P . Rethinking ImageNet Pre-Training[C]// International Conference on Computer Vision. 0.

[30] Sun P , Zhang R , Jiang Y , et al. Sparse R-CNN: End-to-End Object Detection with Learnable Proposals[J]. 2020.

[31] Lee S H , Lee S , Song B C . Vision Transformer for Small-Size Datasets[J]. arXiv e-prints, 2021.

[32] Bao H , Dong L , Wei F . BEiT: BERT Pre-Training of Image Transformers[J]. 2021.