STRIDE RANDOM ERASING AUGMENTATION

Teerath Kumar, Rob Brennan and Malika Bendechache

CRT AI and ADAPT, School of Computing, Dublin City University, Ireland

ABSTRACT

This paper presents a new method for data augmentation called Stride Random Erasing Augmentation (SREA) to improve classification performance. In SREA, probability based strides of one image are pasted onto another image and also labels of both images are mixed with the same probability as the image mixing, to generate a new augmented image and augmented label. Stride augmentation overcomes limitations of the popular random erasing data augmentation method, where a random portion of an image is erased with 0 or 255 or the mean of a dataset without considering the location of the important feature(s) within the image. A variety of experiments have been performed using different network flavours and the popular datasets including fashion-MNIST, CIFAR10, CIFAR100 and STL10. The experiments showed that SREA is more generalized than both the baseline and random erasing method. Furthermore, the effect of stride size in SREA was investigated by performing experiments with different stride sizes. Random stride size showed better performance. SREA outperforms the baseline and random erasing especially on the fashion-MNIST dataset. To enable the reuse, reproduction and extension of SREA, the source code is provided in a public git repository https://github.com/kmr2017/stride-aug.

KEYWORDS

Data Augmentation, Image Classification, Erasing Augmentation.

1. INTRODUCTION

Since the advent of deep learning, it has improved classification performance in a wide variety of domains including image classification [1, 2, 3], audio classification [4,5,6] and text classification [8,9,10]. The performance of deep learning algorithms is evaluated by model generalization. To prevent overfitting, two popular techniques of model generalization are used: model regularization i.e. batch normalization [11], dropout [12] and data augmentation [14, 15, 16]. There are many state-of-the-art techniques for data augmentation and random erasing data augmentation [14] is one of them. In random erasing, a randomly sized patch in a random position in an image is erased with 0 or 255 or the mean of the dataset. Though it is effective, there is a high probability that significant features of the image can be erased which deteriorates model performance. The effect of this deterioration is shown in Figure 1, where a random part of the image is erased, consequently the augmented image has lost many significant features of the original input. Thus, this augmented image when used as training data leads to bad model generalization rather than improving the performance. To overcome this issue, this paper proposes a new data augmentation named Stride Random Erasing Augmentation (SREA), where random size strides (or slices) of one image are pasted onto another image with a random probability. We investigate if SREA provides the benefits of random erasing augmentation while preserving the good features. In this work, we use the terms model and network interchangeably. Our work has the following contributions:

David C. Wyld et al. (Eds): AISCA, NET, DNLP - 2022
pp. 01-09, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120201

- We propose a novel augmentation approach, named Stride Random Erasing Augmentation (SREA), it does not only provide random erasing (as images are mixed in random stride way) but also preserves the significant features
- Unlike conventional augmentation techniques, features are not lost as in random erasing
- We perform a series of image classification experiments on standard datasets using our proposed approach and it outperforms both baseline and random erasing-based classification.
- We investigate the effect of different stride sizes (small, random and large) and the effect of different augmentation probability values.
- We provide full source code for SREA in an open repository: https://github.com/kmr2017/stride-aug

The rest of the paper is structured as follows: Section 2 describes the closely related work, Section 3 describes the algorithm of proposed SREA method, Section 4 explains the experimental setup and results, and finally, Section 5 provides conclusions and ideas for future work.

2. RELATED WORK

The objective of model generalization is to prevent the model from overfitting. The two main techniques used for model generalization are: regularization [11, 12, 21, 22, 23] and data augmentation [13, 15, 14, 16, 17, 18, 20].

2.1. Regularization

Dropout [12] is a regularization technique, where hidden and visible neural network neuron probabilities are randomly set to zero and are dropped. In Ba, J. [21] an adaptive dropout is proposed where the probability of a hidden neuron, that is to be discarded, is calculated using a binary belief network. DropConnect [22] randomly selects the subsets of weights and sets them to zero instead of disconnecting the neurons. In the stochastic pooling [23], parameter free activations are selected during training from a multinomial distribution and used with state-of-the-art regularization techniques.

2.2. Data Augmentation

Data augmentation is one of the prominent techniques used for regularization [14]. Data augmentation is used to increase training dataset size and thereby increase classification test accuracy with less original data. There are many techniques for data augmentations, i.e., translation, rotation and addition of salt-and-pepper noise, etc. Among them, the three most popular and close to the proposed approach are flipping [15], random cropping [13] and random erasing [14]. Flipping is simply a manipulation where the object is flipped horizontally or vertically or both. Random cropping selects a random patch from an image and resizes it to the original image size. In random erasing [14], a random part of an image is erased during the training. In random image cropping and patching [16], patches from four images are extracted and mixed to create a new image and the labels are mixed correspondingly. This work [17] analyzes traditional data augmentation techniques i.e., rotating, cropping, zooming, histogram based methods and others. Recently a new perspective of data augmentation named mathematical framework was proposed in [18]. It explains data augmentation benefits and the authors proved that data augmentation is equivalent to performing the average operation on a certain group that does not vary in data distribution. The proposed SREA does not only provide random erasing (as images are mixed in random stride way) but also preserves the significant features (features are not lost as in random erasing [14]). So it is useful for models to learn these features, resulting in a good regularization effect.

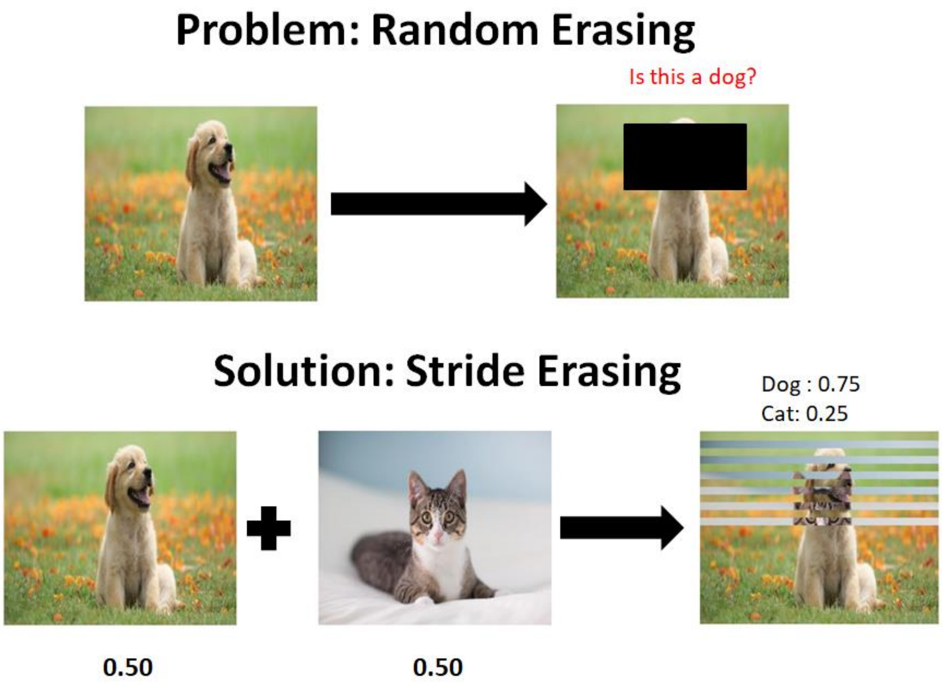

Fig. 1. The first row highlights the problem of important features removal with random erasing, the second row represents the proposed solution

3. Proposed Method

In this section, we explain our proposed approach stride random erasing data augmentation (SREA) method. During training, there is a probability P of performing SREA. In SREA, W and $P_s/2$ represent the width of image and the striding probability, respectively. There are n strides, calculated by $\lfloor W \times P_s/2 \rfloor$ and with random stride size S, of image X_1 and X_2 are pasted alternatively to generate a new augmented image X_a. As for images X_1 and X_2, the stride probability is $P_s/2$ and $1 - P_s/2$, respectively, so, with the same probability, L_1 and L_2 are labels of image X_1 and X_2, respectively, are mixed to generate an augmented label L_a. The newly augmented image X_a and augmented label L_a are used for training the model. The reason for halving the Ps is, in an augmented image, strides of images X_1 and X_2 are pasted alternatively i.e. one stride of X_1, then stride of X_2, process continues till n strides are done, consequently half strides of X_1 are pasted and place of half strides taken by strides of X_2, logically the probability of X_1 is also halved. For further clarification, for example, although the dog and cat have a initially mixing probability of 0.5 each before mixing, but in the augmented image, half strides of the cat are taken by strides of the dog, so the cat contributes half (0.25) of the original probability (0.5) as it is shown in Figure 1. We define the SREA mathematical combination operation as below:

$$X_a = X_2 \oplus [X_1 \otimes n * S] \quad \textit{Eq. 1}$$

In the above equation, $\oplus$ and $\otimes$ represent pasting and striding operations, respectively. $n * S$ represents n strides of size S each. In the same way, labels are also mixed as follows:

$$L_a = L_1 . P_s/2 + L_2 . (1 - P_s/2) \quad \textit{Eq. 2}$$

The labels are mixed in the same ratio as images are mixed, consequently this provides a strong regularization effect and makes the model more generalized.

The proposed algorithm for this approach is defined in algorithm 1. The source code is available in a git repository.

Algorithm 1: Stride Augmentation(X_1,X_2,L_1, L_2)

```
Input: X1 : Image 1
       X2 : Image 2
       L1: label of X1,
       L2: label of X2,
Output: Augmented image, Augmented label
Probability = random() // random probability of mixing x2
strideSize= random(2,10) // random stride size in range of 2 to 10
width = width(X1) // getting width of image
totalStrides= int(width/strideSize) // Total number of strides
mixingStride = int(totalStrides*Probability) // number of strides use for
    mixing images
startPosition = random(1,int(width/2)) // startion position for mixing
    images
Xa = copy(X1) // Augmented image to be stored
for Choose i ∈ range(0, mixingStrides, 2) do
    Xa[startPosition+i*strideSize:startPosition+(i+1)*strideSize,:] =
        X2[startPosition+i*strideSize:startPosition+(i+1)*strideSize,:]
        // Mixing images with strides
Probability = Probability/2.0 // In loop, after each second interval
    stride of x2 is mixed, so probability is halved
La=L1*(1-Probability)+L2*Probability
return Xa, La
```

4. Experiment

In this section, we define the datasets used, the training set up and the classification results obtained for this initial evaluation of our SREA method, the random erasing method and a baseline with no data augmentation.

4.1. Datasets

We used four datasets for our experiments including Fashion-MNIST [24], CIFAR10 [25], CIFAR100 [25] and STL10 [26].

Fashion-MNIST

It consists of 70000 images including 60000 training and 10000 test images. Each image is gray scale and of size 28 × 28. There are 10 classes of clothing items e.g. t-shirt, shoe and dress. Before training, we normalized these images between 0 and 1.

CIFAR10 and CIFAR100

It consists of 60000 images, including 50000 training and 10000 test images. Each image is RGB color and of dimension 32 ×32 × 3. There are 10 classes in this dataset. These data images were normalized using the mean and standard deviation of the dataset. Similar to CIFAR10, CIFAR100 has the same number of images, same dimensions and everything except the number of classes are 100.

STL10

This dataset has a total of 8500 images including 500 training images and 8000 test images. Each image is RGB color and of dimension 96 x 96 x 3. There are 10 classes in this dataset. These images are acquired from the biggest imagenet dataset.

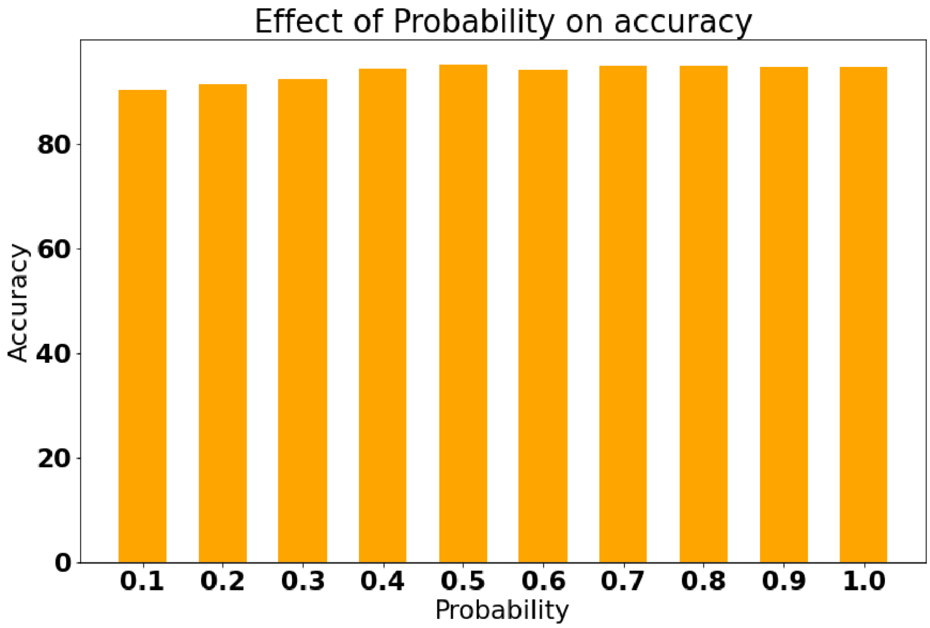

Fig. 2. On Fashion-MNIST dataset using Resnet20 network

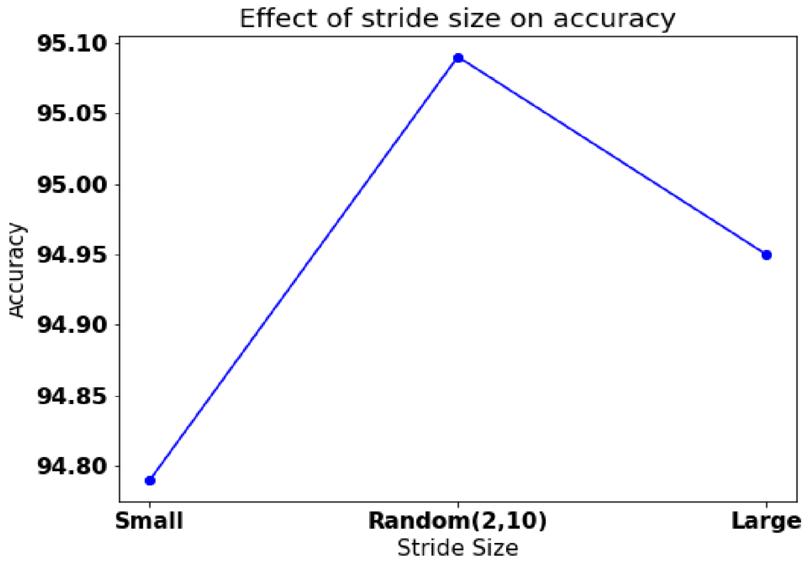

Fig. 3. On Fashion-MNIST dataset using Resnet20 network

4.2. Training setup

For training setup, we use multiple flavours of resnet [27] : resnet20, resnet32, resnet44, resnet56 , resnet100 and flavours of VGG [28] model i.e. VGG11, VGG13, VGG16 and VGG19. For the fair comparison with random erasing, the overall parametric settings are employed with the same setting as in [14]. We used 300 epochs for training, the learning rate was initially set to 0.1 and reduced by 10 times at epoch 100, 150, 175 and 190. The probability of performing SREA is set to 0.5 for the main experiments. This is because we initially investigated 10 different SREA probability settings with an interval of 0.1 starting from 0.1 on FashionMNIST using resnet20 model. In this test 0.5, SREA probability showed the best result, as shown in Figure 2. We re-performed all the Zhong et al.'s experiments for fashion-MNIST, because the original experiments performed in the random erasing paper [29] were on an old fashioned dataset, in which there was overlapping between test and training images (this issue is discussed in the Github repository of random erasing [14]). Each experiment is repeated three times and the mean error with standard deviation is reported in Table 1. Note that, boldface number shows the best performance.

4.3. Results

In this section, the results achieved with SREA are compared with the baseline and the standard random erasing augmentation method. Firstly, we investigated the effect of stride size. For this purpose, we used a fixed small stride size of 2, a fixed large stride size of 10 and a randomly generated stride size between 2 and 10 on the Fashion-MNIST dataset using resnet20. Out of all three sizes, the randomly generated stride size has shown better performance for this dataset as shown in Figure 3. Furthermore, with classification tasks, SREA also outperformed both baseline and random erasing in all flavours of the resnet model by showing better results in all categories, albeit sometimes within the margin of error. While in the case of CIFAR10 and CIFAR100, this initial implementation of SREA has shown competitive results with random erasing. In some resnet flavour cases it narrowly outperformed random erasing (again within the margin of error) and it showed impressive performance over baseline in all resnet flavours. For further evaluating the effectiveness of SREA, we use multiple flavours of VGG, it shows superior performance as compared to baseline and competitive performance with random erasing. For STL10 data, SREA outperformed both baseline and random erasing except the VGG19 network.

Table 1. Error rate performance comparison of the proposed SREA method with a baseline and random erasing.

Models	**Baselines**	**Random Erasing**	**SREA**
Fashion-MNIST			
ResNet20	6.21± 0.11	5.04 ± 0.10	**4.91 ± 0.12**
Resnet32	6.04 ± 0.13	4.84 ± 0.12	**4.81 ± 0.17**
Resnet44	6.08 ± 0.16	4.87 ± 0.1	**4.07 ± 0.14**
Resnet56	6.78 ± 0.16	5.02 ± 0.11	**5.00 ± 0.19**
CIFAR10			
Resnet20	7.21 ± 0.17	**6.73 ± 0.09**	7.18 ± 0.13
Resnet32	6.41 ± 0.06	**5.66 ± 0.10**	6.31 ± 0.14
Resnet44	5.53 ± 0.0	5.13 ± 0.09	**5.09 ± 0.10**
Resnet56	5.31 ± 0.07	**4.89 ± 0.0**	5.02 ± 0.11
VGG11	7.88±0.76	7.82±0.65	**7.80±0.65**
VGG13	6.33±0.23	6.22±0.63	**6.18±0.54**
VGG16	6.42±0.34	6.21±0.76	**6.20±0.34**
VGG19	6.88±0.65	6.85±0.65	**6.75±0.55**
CIFAR100			
Resnet20	30.84 ± 0.19	**29.97 ± 0.11**	30.18 ± 0.27
Resnet32	28.50 ± 0.37	27.18 ± 0.32	**27.08 ± 0.34**
Resnet44	25.27 ± 0.21	**24.29 ± 0.16**	24.49 ± 0.23
Resnet56	24.82 ± 0.27	23.69 ± 0.33	**23.35 ± 0.26**
VGG11	28.97±0.76	28.73±0.67	**28.26±0.75**
VGG13	25.73±0.67	**25.71±0.54**	25.71±0.56
VGG16	26.64±0.56	26.63±0.75	**26.61±0.65**
VGG19	**28.65±0.23**	28.69±0.76	28.75±0.76
STL10			
VGG11	22.29±0.13	22.27±0.21	**20.68±0.23**
VGG13	20.64±0.26	20.18±0.23	**19.91±0.92**
VGG16	20.62±0.34	20.12±0.65	**20.09±0.23**
VGG19	**19.15±0.32**	19.22±0.45	19.35±0.11

5. CONCLUSION

This paper addressed the issues of random erasing, where good features are lost due to randomly erasing a random size of patch, which deteriorates the model performance. To cope up with this issue, we proposed a new data augmentation method named Stride Random Erasing data augmentation, that not only provides random erasing but also preserves significant features. We investigated the effect of different probability values and stride sizes parameters on our approach. Furthermore, our approach outperformed baseline and random erasing on a wide variety of datasets using different flavour of resnet and vgg. In future, we will extend our work by including column-wise strides, both row-wise and column-wise strides and test SREA on audio datasets. Nonetheless this first implementation of the approach shows promise for building a new family of stride-based data augmentation techniques.

ACKNOWLEDGEMENTS

This publication has emanated from research [conducted with the financial support of/supported in part by a grant from] Science Foundation Ireland under Grant number 18/CRT/6223 and is supported by the ADAPT Centre for Digital Content Technology which is funded under the SFI Research Centres Programme (Grant 13/RC/2106/_P2), Lero SFI Centre for Software (Grant 13/RC/2094/_P2) and is co-funded under the European Regional Development Fund. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

REFERENCES

[1] Kumar, J., Bedi, P., Goyal, S. B., Shrivastava, A., & Kumar, S. (2021, March). Novel Algorithm for Image Classification Using Cross Deep Learning Technique. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1099, No. 1, p. 012033). IOP Publishing.

[2] Liu, J. E., & An, F. P. (2020). Image classification algorithm based on deep learning-kernel function. *Scientific programming, 2020*.

[3] Wang, H., & Meng, F. (2019). Research on power equipment recognition method based on image processing. *EURASIP Journal on Image and Video Processing*, *2019*(1), 1-11.

[4] Nanni, L., Maguolo, G., Brahnam, S., & Paci, M. (2021). An ensemble of convolutional neural networks for audio classification. *Applied Sciences*, *11*(13), 5796.

[5] Hershey, S., Chaudhuri, S., Ellis, D. P., Gemmeke, J. F., Jansen, A., Moore, R. C., ... &Wilson, K. (2017, March). CNN architectures for large-scale audio classification. In *2017 ieee international conference on acoustics, speech and signal processing (icassp)* (pp. 131-135). IEEE.

[6] Rong, F. Audio classification method based on machine learning. 2016 International Conference On Intelligent Transportation, Big Data & Smart City (ICITBS) pp.81-84 (2016)

[7] Aiman, A., Shen, Y., Bendechache, M., Inayat, I. & Kumar, T. AUDD: Audio Urdu Digits Dataset for Automatic Audio Urdu Digit Recognition. Applied Sciences. 11, 8842 (2021)

[8] Kolluri, J., Razia, D. S., & Nayak, S. R. (2019, June). Text classification using Machine Learning and Deep Learning Models. In *International Conference on Artificial Intelligence in Manufacturing & Renewable Energy (ICAIMRE)*.

[9] Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep Learning--based Text Classification: A Comprehensive Review. *ACM Computing Surveys (CSUR)*, *54*(3), 1-40.

[10] Nguyen, T. H., & Shirai, K. (2013, June). Text classification of technical papers based on text segmentation. In *International Conference on Application of Natural Language to Information Systems* (pp. 278-284). Springer, Berlin, Heidelberg.

[11] Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (pp. 448-456). PMLR.

[12] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*(1), 1929-1958.

[13] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*, 1097-1105.

[14] Zhong, Z., Zheng, ., Kang, G., Li, S., & Yang, Y. (2020, April). Random erasing data augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 07, pp. 13001-13008).

[15] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[16] Takahashi, R., Matsubara, T., & Uehara, K. (2019). Data augmentation using random image cropping and patching for deep CNNs. *IEEE Transactions on Circuits and Systems for Video Technology*, *30*(9), 2917-2931.

[17] Mikołajczyk, A., & Grochowski, M. (2018, May). Data augmentation for improving deep learning in image classification problem. In *2018 international interdisciplinary PhD workshop (IIPhDW)* (pp. 117-122). IEEE.

[18] Chen, S., Dobriban, E., & Lee, J. H. (2020). A group-theoretic framework for data augmentation. *Journal of Machine Learning Research*, *21*(245), 1-71.

[19] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1), 1-48

[20] Wei, J., & Zou, K. (2019). Eda: Easy data augmentation techniques for boosting performance on text classification tasks. *arXiv preprint arXiv:1901.11196*.

[21] Ba, J., & Frey, B. (2013). Adaptive dropout for training deep neural networks. *Advances in neural information processing systems*, *26*, 3084-3092.

[22] Wan, L., Zeiler, M., Zhang, S., Le Cun, Y., & Fergus, R. (2013, May). Regularization of neural networks using dropconnect. In *International conference on machine learning* (pp. 1058-1066). PMLR.

[23] Zeiler, M. D., & Fergus, R. (2013). Stochastic pooling for regularization of deep convolutional neural networks. *arXiv preprint arXiv:1301.3557*.

[24] Xiao, H., Rasul, K., & Vollgraf, R. (2017). Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*.

[25] Krizhevsky, A., & Hinton, G. (2009). Learning multiple layers of features from tiny images.

[26] A. Coates, A. Ng, and H. Lee, “An analysis of single-layer networks in unsupervised feature learning,” in Proceedings of the fourteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings, 2011, pp. 215–223.

[27] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[28] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[29] https://github.com/zhunzhong07/Random-Erasing/issues/9

AUTHORS

Teerath kumar received his Bachelor’s degree in Computer Science with distinction from National University of Computer and Emerging Science (NUCES), Islamabad, Pakistan, in 2018. Currently, he is pursuing PhD from Dublin City University, Ireland. His research interests include advanced data augmentation, deep learning for medical imaging, generative adversarial networks and semi-supervised learning.

R. Brennan is an Assistant Professor in the School of Computing, Dublin City University, Chair of the DCU MA in Data Protection and Privacy Law and a Funded investigator in the Science Foundation Ireland ADAPT Centre for Digital Content Technology which is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund, His main research interests are data protection, data value, data quality, data privacy, data/AI governance and semantics.

M. Bendechache is an Assistant Professor in the School of Computing at Dublin City University, Ireland. She obtained her Ph.D. degree from University College Dublin, Ireland in 2018. Malika’s research interests span the areas of Big data Analytics, Machine Learning, Data Governance, Cloud Computing, Blockchain, Security, and Privacy. She is an academic member and a Funded Investigator of ADAPT and Lero research centres.