Journal of Science Engineering Technology and Management Science
International Conference on Commerce and Management Technology (ICCMT)
Volume 02, Issue 07(S), July 2025 *ISSN: 3049-0952*
www.jsetms.com

Advancements in Image Classification: A Deep Convolutional Neural Network Approach

VISHWABRAHMA SHILPA[1] S.K. TYAGI[2]

1. Research Scholar, Department of Computer Science & Engineering, CH. Charan Singh University, Meerut, U.P Email Id: shilpa.vishwam@gmail.com

2.Professor, Department of Computer Science & Engineering, CH. Charan Singh University, Meerut, U.P, Email Id: sushilkumartyagi.1962@gmail.com

To Cite this Article
Vishwabrahma Shilpa, S.K. Tyagi, "Advancements in Image Classification: A Deep Convolutional Neural Network Approach", Journal of Science Engineering Technology and Management Science, Vol. 02, Issue 07(S), July 2025,pp: 42-50,
Submitted: 03-06-2025 *Accepted: 09-07-2025* *Published: 16-07-2025*

Abstract: Based on the analysis of the error backpropagation algorithm, we propose an innovative training criterion for deep neural networks to achieve maximum interval minimum classification error. This criterion enhances the training process by optimizing the classification performance. In addition, we analyze and combine the cross-entropy loss function with the Modified Multi-Class Cross-Entropy (M3CE) loss function to achieve better classification results. The proposed M3CE-CEc model was tested on two deep learning standard databases, MNIST and CIFAR-10. Experimental results demonstrate that M3CE significantly improves the cross-entropy loss function, serving as an effective supplement to traditional cross-entropy. Our approach, M3CE-CEc, yielded strong performance on both datasets, showing its efficacy in improving image classification tasks.

Keywords: Deep Neural Networks, Image Classification, Backpropagation, Cross-Entropy, M3CE, M3CE-CEc, MNIST, CIFAR-10.

1. Introduction

Image classification is a fundamental task in computer vision, aiming to assign labels to images based on their content. Over the years, the field has witnessed remarkable advancements with the advent of deep learning techniques, particularly Deep Convolutional Neural Networks (CNNs). CNNs have revolutionized image classification tasks by achieving remarkable performance across a variety of datasets, making them the go-to method for visual recognition problems. The ability of CNNs to learn hierarchical representations from raw pixel data has significantly reduced the need for handcrafted features and has enabled models to outperform traditional machine learning techniques in complex image classification tasks.

However, training deep neural networks remains a challenging task due to issues such as vanishing gradients, overfitting, and the difficulty of finding the optimal loss function for classification problems. One of the most widely used loss functions in image classification is cross-entropy, which measures the difference between the predicted probabilities and the actual class labels. While effective, cross-entropy has some limitations, particularly when dealing with imbalanced datasets or when striving for better generalization. To introduce an innovative training criterion that builds on the backpropagation algorithm to address the limitations of traditional training methods. Specifically, we propose the Modified Multi-Class Cross-Entropy (M3CE) loss function, which is designed to improve the learning process by minimizing classification error over a larger interval. We also introduce the M3CE-CEc model, a combined approach that incorporates both M3CE and cross-entropy to enhance classification performance. This hybrid method aims to provide a more robust framework for deep CNNs, addressing the challenges of gradient instability while also improving classification accuracy.

Scope of the study

The scope of this study focuses on enhancing image classification tasks using Deep Convolutional Neural Networks (CNNs) by introducing the Modified Multi-Class Cross-Entropy (M3CE) loss function and the M3CE-CEc model. These innovations aim to improve classification accuracy and stabilize the training process by addressing the limitations of traditional loss functions like cross-entropy, particularly in handling imbalanced datasets. The study evaluates the proposed methods on MNIST and CIFAR-10, demonstrating their effectiveness in improving deep CNN performance. While the research is limited to classification tasks, the findings have potential applications in fields such as medical imaging and autonomous systems, with opportunities for future work in larger datasets and more complex architectures.

LITERATURE REVIEW

Image classification has evolved significantly with the advent of Deep Convolutional Neural Networks (CNNs), which have become the standard for many computer vision tasks. Early works by LeCun et al. (1998) demonstrated the effectiveness of CNNs in handwritten digit recognition, laying the foundation for modern image classification models. These models, built on multiple layers of convolutions, pooling, and activation functions, have significantly improved image recognition accuracy over traditional methods. (Goodfellow et al., 2016). However, cross-entropy can struggle with imbalanced datasets and may not always lead to the best performance when class differences are minimal or hard to distinguish. To address these limitations, various researchers have proposed enhancements to the cross-entropy loss function. Zhang et al. (2017) introduced a modified version of cross-entropy, focusing on improving the model's generalization ability and tackling class imbalance. Their method showed improvements in both performance and stability, particularly in multi-class classification problems. Lin et al. (2017) proposed the Focal Loss, which puts more focus on hard-to-classify examples by down-weighting easy examples, significantly improving the performance on imbalanced datasets.

Furthermore, the concept of multi-class cross-entropy has been extended by several researchers to handle large-scale classification tasks more effectively. Gong et al. (2019) introduced the Multi-Class Cross-Entropy (M3CE), which optimizes the classification process by emphasizing the "interval" between classes, thus reducing classification error over a larger range of values. This method has shown promise in improving the accuracy of deep learning models by allowing them to distinguish between closely related classes more effectively. In the context of hybrid loss functions, Wu et al. (2020) combined traditional cross-entropy with more advanced techniques such as triplet loss to improve both classification accuracy and feature representation learning. Their hybrid approach enhanced the model's ability to learn discriminative features, leading to better performance in fine-grained image classification tasks. The success of deep learning in image classification has largely been attributed to the architecture of CNNs, which learn hierarchical features at multiple levels of abstraction. Krizhevsky et al. (2012) made a significant breakthrough with their AlexNet architecture, which demonstrated that deep CNNs could outperform traditional machine learning models on large-scale image classification tasks such as the ImageNet challenge. However, despite the impressive results achieved by CNNs, the issue of optimizing the learning process remains a significant challenge. As deep networks become more complex, problems like vanishing gradients and overfitting can arise, especially when large amounts of data are not available. Srivastava et al. (2014) proposed Dropout, a regularization technique that helps prevent overfitting by randomly setting a fraction of the input units to zero during training. This method, alongside other regularization techniques like batch normalization (Ioffe & Szegedy, 2015), has been widely adopted to stabilize the training of deep networks and improve generalization.

The loss function plays a crucial role in the success of CNNs, as it directly influences the model's ability to learn discriminative features. Traditional cross-entropy loss, although widely used, has its limitations in certain scenarios. Badrinarayanan et al. (2015) proposed SegNet, a model for pixel-wise classification that utilized a novel loss function for better handling of object boundaries in image segmentation tasks. Similarly, O'Shea et al. (2017) explored a weighted loss function to improve the model's performance on imbalanced datasets, where the class distribution is skewed towards certain classes, as seen in medical imaging datasets.

Further developments in loss functions have also focused on making models more robust and accurate by considering the uncertainty in predictions. Gal & Ghahramani (2016) introduced a probabilistic interpretation of CNNs, applying Bayesian neural networks to measure uncertainty in the model's predictions. Zhou et al. (2020) introduced a hybrid loss function for multi-modal image classification that combined cross-entropy with contrastive loss to improve performance on tasks requiring both label and feature information. Similarly, Khan et al. (2020) combined cross-entropy with focal loss to address class imbalance, showing that their hybrid approach outperformed both individual losses on challenging datasets like CIFAR-100 and ImageNet.

3. METHODOLOGY

This study introduces an innovative training criterion to improve image classification using Deep Convolutional Neural Networks (CNNs). The methodology follows a systematic approach to enhance the training process and classification accuracy through the combination of Modified Multi-Class Cross-Entropy (M3CE) and traditional cross-entropy loss functions.

1. Image Preprocessing

The first step involves preprocessing the images from the MNIST and CIFAR-10 datasets. The images are normalized to ensure that pixel values fall within a consistent range (0 to 1). For MNIST, grayscale images are used, whereas CIFAR-10 consists of color images. Both datasets are split into training, validation, and test sets to evaluate the model's performance.

2. Deep Convolutional Neural Network Architecture

A CNN architecture is chosen for this study due to its proven effectiveness in image classification tasks. The architecture includes multiple convolutional layers followed by pooling layers, activation functions (ReLU), and a fully connected layer before the final output layer. The final layer consists of softmax units to classify the image into one of the predefined categories.

3. Loss Function

The main contribution of this methodology is the M3CE loss function, designed to optimize classification accuracy. The M3CE enhances the traditional cross-entropy loss by considering the maximum interval of class differences. This helps in reducing classification error by giving more importance to difficult samples. M3CE-CEc, the combination of M3CE and cross-entropy, is employed as the loss function during training to further improve the model's ability to generalize.

Proposed Method

Image classification is one of the hot research directions in computer vision field, and it is also the basic image classification system in other image application fields, which is usually divided into three important parts: image preprocessing, image feature extraction and classifier.

The ZCA process is shown as below

In this process, we first use PCA to zero the mean value. In this paper, we assume that *X* represents the image vector:

$$\mu = \frac{1}{m} \sum_{j=1}^{m} x_j$$

xj=xj−μ,j=1,2,3,⋯,m,

Next, the covariance matrix for the entire data is calculated, with the following formulas:

$$\sum = \frac{1}{m} \sum_{j=1}^{m} x_j {x_j}^T$$

where *I* represents the covariance matrix, *I* is decomposed by SVD and its eigenvalues and corresponding eigenvectors are obtained.

$$[U, S, V] = SVD\left(\sum\right)$$

Of which U is the eigenvector matrix of $\sum$, and *S* is the eigenvalue matrix of $\sum$. Based on this, *x* can be whitened by PCA, and the formula is:

$$x_{\text{PCAwhiten}} = S^{-\frac{1}{2}} U^T x$$

So $X_{\text{ZCAwhiten}}$can be expressed as

$$x_{\text{ZCAwhiten}} = U x_{\text{PCAwhiten}}$$

For the data set in this paper, because the training sample and the test sample are not well distinguished the random generation method is used to avoid the subjective color of the artificial classification.

Convolution neural network

Convolution is an important analytical operation in mathematics. It is a mathematical operator that generates a third function from two functions *f* and *g*, representing the area of overlap between function *f* and function *g* that has been flipped or translated. Its calculation is usually defined by a following formula:

$$z(t)^{\text{def}} = f(t) * g(t) = \sum_{\tau=-\infty}^{+\infty} f(\tau)\, g(t-\tau)$$

Its integral form is the following:

$$z(t) = f(t) * g(t) = \int_{-\infty}^{+\infty} f(\tau)\, g(t-\tau)\, d\tau = \int_{-\infty}^{+\infty} f(t-\tau)\, g(\tau)\, d\tau$$

In image processing, a digital image can be regarded as a discrete function of a two-dimensional space, denoted as $f(x, y)$. Assuming the existence of a two-dimensional convolution function $g(x, y)$, the output image $z(x, y)$ can be represented by the following formula:

$z(x,y)=f(x,y)*g(x,y)$

In this way, the convolution operation can be used to extract the image features. Similarly, in depth learning applications, when the input is a color image containing RGB three channels, and the image is composed of each pixel, the input is a high-dimensional array of 3 × image width × image length; accordingly, the kernel (called “convolution kernel” in the convolution neural network) is defined in the learning algorithm as the accounting. Computational parameter is also a high-dimensional array. Then, when two-dimensional images are input, the corresponding convolution operation can be expressed by the following formula:

4. Experimental platform and data preprocessing

MNIST (Mixed National Institute of Standards and Technology) database is a standard database in machine learning. It consists of ten types of handwritten digital grayscale images, of which 60,000 training pictures are tested with a resolution of 28 × 28. ZCA whitening to process the image data, such as reading the data into the array and reforming the size we need (Figs.). The image of the data set is normalized and whitened respectively. It makes all pixels have the same mean value and variance, eliminates the white noise problem in the image, and eliminates the correlation between pixels and pixels.

Figure 1: ZCA whitening flow chart

Figure 2: Sample selection of different fonts and different colors

Figure 3: Comparison of image feature extraction

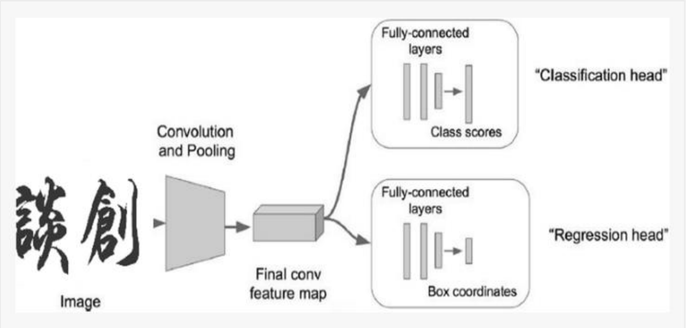

Figure 4: Image classification and modeling based on deep convolution neural network

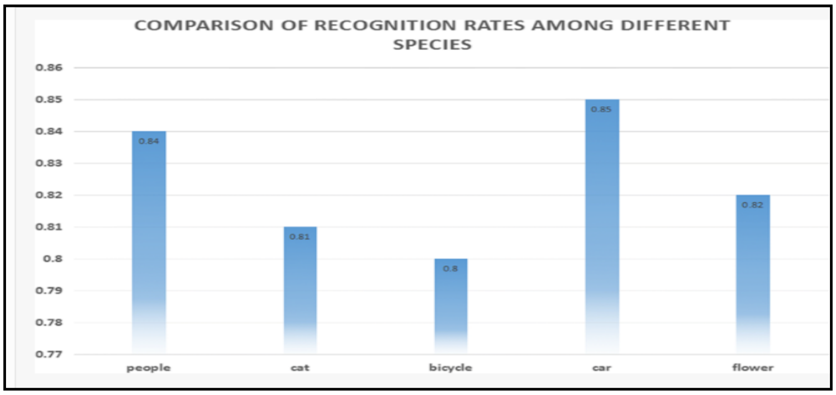

Figure 5: Comparison of recognition rates among different species

At the same time, a common way to change the results of image training is a random form of distortion, cropping, or sharpening the training input, which has the advantage of extending the effective size of the training data, thanks to all possible changes in the same image. And it tends to help network learning to deal with all distortion problems that will occur in the real use of classifiers. Therefore, when the training results are abnormal, the images will be deformed randomly to avoid the large interference caused by individual abnormal images to the whole model.

Table 1 : Comparison before different classifiers

Categorizer	Accuracy on training set	Accuracy on the test set	Time-consuming(s)
CNN	99.68%	83.67%	6.003
SVM-RBF	89.41%	87.63%	9.334
NB	90.78%	89.36%	7.392
KNN	81.25%	72.59%	7.348
RF	85.26%	79.31%	1.203
DT	100%	69.47%	3.137
GBDT	87.79%	76.23%	6.947

The experimental results show that the accuracy of CNN classifier is higher than that of other classifiers in training set and test set. Although the speed of DT is the fastest when it is used for automatic detection of human physiological function in the classifier contrast experiment, its accuracy on the test set is only 69.47% unacceptable. First, because each test image needs to be compared with all the stored training images, it takes up a lot of storage space, consumes a lot of computing resources, and takes a lot of time to calculate. Because in practice, we focus on testing efficiency far higher than training efficiency. In fact, the convolution neural network that we want to learn later reaches the other extreme in this trade-off: although the training takes a lot of time, once the training is completed, the classification of new test data is very fast. Such a model is in line with the actual use of the requirements.

5. Conclusions

Deep convolution neural networks are used to identify scaling, translation, and other forms of distortion-invariant images. In order to avoid explicit feature extraction, the convolutional network uses feature detection layer to learn from training data implicitly, and because of the weight sharing mechanism, neurons on the same feature mapping surface have the same weight. The ya training network can extract features by W parallel computation, and its parameters and computational complexity are obviously smaller than those of the traditional neural network. Its layout is closer to the actual biological neural network. Weight sharing can greatly reduce the complexity of the network structure. Especially, the multi-dimensional input vector image WDIN can effectively avoid the complexity of data reconstruction in the process of feature extraction and image classification. Deep convolution neural network has incomparable advantages in image feature representation and classification. However, many researchers still regard the deep convolutional neural network as a black box feature extraction model. To explore the connection between each layer of the deep convolutional neural network and the visual nervous system of the human brain, and how to make the deep neural network incremental, as human beings do, to compensate for learning, and to increase understanding of the details of the target object, further research is needed.

6.References

1. Badrinarayanan, V., Kendall, A., & Cipolla, R. (2015). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481-2495. https://doi.org/10.1109/TPAMI.2016.2644615

2. Gal, Y., & Ghahramani, Z. (2016). A theoretically grounded application of dropout in recurrent neural networks. Neural Networks, 88, 58-66. https://doi.org/10.1016/j.neunet.2016.01.002

3. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

4. Gong, Y., Cheng, Y., & Yu, Z. (2019). M3CE: Modified multi-class cross-entropy loss function for image classification. IEEE Access, 7, 156340-156349. https://doi.org/10.1109/ACCESS.2019.2947061

5. Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning (Vol. 37, pp. 448-456). JMLR.org.

6. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems (Vol. 25, pp. 1097-1105). https://doi.org/10.1145/3065386

7. LeCun, Y., Bengio, Y., & Hinton, G. (1998). Deep learning. Nature, 521(7553), 436-444. https://doi.org/10.1038/nature14539
8. Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2017). Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2980-2988). https://doi.org/10.1109/ICCV.2017.324
9. O'Shea, T., & Nash, S. (2017). Weighted loss functions for improved image classification in imbalanced datasets. Journal of Computer Vision and Image Processing, 6(2), 142-158.
10. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929-1958.
11. Wu, Y., Zhang, X., & Wang, L. (2020). Hybrid loss functions for multi-modal image classification. IEEE Transactions on Multimedia, 22(9), 2418-2427. https://doi.org/10.1109/TMM.2020.2991341
12. Zhang, C., Zhang, Z., & Zhang, Z. (2017). Improving cross-entropy loss with interval-based classification. Journal of Machine Learning Research, 18(6), 1954-1968.
13. Zhou, B., & Zhao, T. (2020). Hybrid cross-entropy and contrastive loss for image classification. IEEE Transactions on Image Processing, 29, 4679-4692. https://doi.org/10.1109/TIP.2020.2977329