BOANN: Bayesian-Optimized Attentive Neural Network for Classification

Luoyao He , Xinggi Wang , Yuzhen Lin , Xinjin Li * , Yu Ma , Zhenglin Li

Posted Date: 4 February 2025

doi: 10.20944/preprints202409.2367.v2

Keywords: Deep learning; Image classification

Article

BOANN: Bayesian-Optimized Attentive Neural Network for Classification

Luoyao He [1], Xingqi Wang [2], Yuzhen Lin [3], Xinjin Li [4,*], Yu Ma [5] and Zhenglin Li [5]

1. University College London, London, UK; zcbelhe@ucl.ac.uk
2. Johns Hopkins University, Baltimore, MD, USA; wxq19991001@gmail.com
3. Carnegie Mellon University, Pittsburgh, PA, USA; yuzhenl@alumni.cmu.edu
4. Columbia University, New York, NY, USA
5. Texas A&M University, College Station, TX, USA; yuma13926@gmail.com; lizhenglin2001@gmail.com

* Correspondence: i.xinjin@columbia.edu

***Abstract*—This** study presents the Bayesian-Optimized Attentive Neural Network (BOANN), a novel approach enhancing image classification performance by integrating Bayesian optimization with channel and spatial attention mechanisms. Traditional image classification struggles with the extensive data in today's big data era. Bayesian optimization has been integrated into neural networks in recent years to enhance model generalization, while channel and spatial attention mechanisms improve feature extraction capabilities. This paper introduces a model combining Bayesian optimization with these attention mechanisms to boost image classification performance. Bayesian optimization optimizes hyperparameter selection, accelerating model convergence and accuracy; the attention mechanisms augment feature extraction. Compared to traditional deep learning models, our model utilizes attention mechanisms for initial feature extraction, followed by a Bayesian-optimized neural network. On the CIFAR-100 dataset, our model outperforms classical models in metrics such as accuracy, loss, precision, recall, and F1 score, achieving an accuracy of 77.6%. These technologies have potential for broader application in image classification and other computer vision domains.

Keywords—Deep learning; Image classification; Convolutional neural network; Bayesian optimization

Introduction

Image classification is one of the basic tasks of computer vision, that is, given an input image, a certain classification algorithm is used to determine the category to which the image belongs. There are many ways to classify images, and different classification results will be obtained based on different classification standards. The main processes of image classification include image preprocessing, image feature description and extraction, and classifier design. Preprocessing includes image filtering (such as median filtering, mean filtering, Gaussian filtering, etc.) and normalization operations, whose purpose is to facilitate the subsequent processing of the target image. Image features are descriptions of its salient attributes, and each image has unique characteristics [1-6]. Feature extraction is to select and extract appropriate features according to the established classification method based on the characteristics of the image itself. A classifier is an algorithm that classifies target images based on selected features.

Traditional image classification methods are processed according to the above process. Their performance differences mainly depend on feature extraction and classifier selection. The features in traditional image classification algorithms are all manually selected. Commonly used image features include low-level visual features such as shape, texture, and color, as well as local invariant features such as scale-invariant feature transforms , local binary pattern, and oriented gradient histograms [7-9]. Although these features have a certain degree of universality, they are not very targeted to specific

 doi:10.20944/preprints202409.2367.v2

images and specific division methods. In addition, for images of some complex scenes, it is very difficult to find artificial features that can accurately describe the target image. Traditional classifiers include K nearest neighbors and support vector machines [10-11]. For some simple image classification tasks, these classifiers are simple to implement and have good results. However, when the category differences are subtle or the image interference is serious, their classification accuracy drops significantly. Therefore, traditional classifiers are not suitable for the classification of complex images.

With the advent of the intelligent information age, deep learning has emerged. As a branch of machine learning, deep learning aims to simulate the human neural network system, build a deep artificial neural network, analyze and interpret the input data, and combine the underlying features of the data into abstract high-level features. This technology has played an irreplaceable role in artificial intelligence fields such as computer vision and natural language processing. As a typical representative of deep learning, the Deep Convolutional Neural Network (DCNN) performs well in computer vision tasks. Compared with traditional image classification algorithms that rely on manual feature extraction, convolutional neural networks extract features from input images through convolution operations, and can effectively learn feature expressions from a large number of samples, thereby enhancing the generalization ability of the model. For example, in autonomous driving, CNNs have improved tasks like recognizing images, understanding the environment, and planning paths [33]. Methods like Class Probability Space Regularization (CPSR), Multiple Distributions Representation Learning (MDRL), and advanced U-Net models with SimAM and CBAM have significantly improved image segmentation accuracy, including pixel-level precision, complex scene analysis, and multimodal healthcare applications [34][35][36]. Deep learning has also improved medical imaging by automatically identifying features like tumors, enhancing diagnostic accuracy [37]. In cybersecurity, AI and ML boost data security through faster threat detection, risk prediction, and smarter decision-making [38]. Dynasmile, a video-based smile analysis software that uses AI, make smile analysis faster and more accurate in orthodontics, advancing esthetic rehabilitation dentistry [39].

Figure 1 shows a classic neural network structure with three levels. The number of nodes in the input layer and the output layer is often fixed, and the number of nodes in the middle layer can be freely specified. The topology and arrows in the neural network structure diagram represent the data flow in the prediction process; the key in the structure diagram is not the circle (representing the neuron), but the connecting line (representing the connection between neurons), each connecting line corresponds to a different weight (its value is called weight), which needs to be trained.

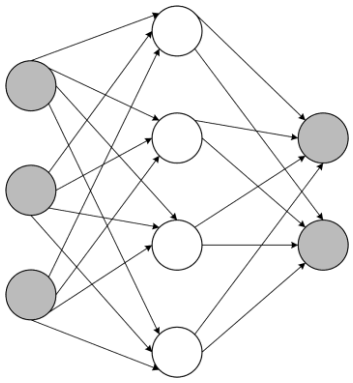

Figure 1. Classical neural network structure.

Among the many neural network models, AlexNet, GoogleNet, VGG16 and ResNet are classic representative architectures that have achieved breakthrough results in large-scale image recognition tasks [12-15]. The introduction of these models not only promoted the development of deep learning research, but also demonstrated strong performance in practical applications [30-32].

These models have excelled in real-world applications, such as optimizing GPU partitioning for better control in autonomous systems [40]. Bayesian optimization has been successful in black-box model tasks, using Neural Processes to handle large parameter spaces [41]. V2F-Net is great at

 doi:10.20944/preprints202409.2367.v2

pedestrian detection, even in occluded situations, by separating visible and full-body estimation [42]. Duality-based methods have achieved sublinear regret in decision-making within Markov decision processes [43]. Confident learning techniques have helped visual defect detection in noisy, imbalanced industrial data [44]. Multi-modal deep learning has improved classification of repairable defects by combining tabular and image data [45]. K-means clustering with SVMs has boosted classification in robotics tasks [46].

Although these classic deep learning models have performed well in image classification tasks, as application requirements increase and data scale expands, how to further improve the performance of neural networks remains an important research direction. In recent years, Bayesian optimization, channel attention mechanism, and spatial attention mechanism have received widespread attention as effective means to improve the performance of neural networks. Bayesian optimization efficiently searches for optimal hyperparameters by constructing a proxy model, while channel attention mechanism and spatial attention mechanism enhance the representation ability of the model by adaptively adjusting important information in the feature map [28-29].

This paper aims to reproduce the classic neural network models AlexNet, GoogleNet, VGG16 and ResNet, and on this basis, design and implement an improved model that combines Bayesian optimization, channel attention and spatial attention. Through experimental comparison, this paper will analyze and verify the performance advantages of the improved model in image classification tasks.

Methods

In this chapter, we will introduce our proposed method in detail. The network mainly integrates the spatial attention mechanism, channel attention mechanism and Bayesian neural network based on Bayesian optimization. Figure 1 shows the classification network structure proposed in this paper.

A. Bayesian Optimization

Bayesian optimization is a commonly used hyperparameter selection method **Error! Reference source not found.**. It is a global optimization algorithm based on Bayes' theorem and is usually used when the objective function is difficult to calculate or the calculation cost is high. Specifically, in Bayesian optimization, the selection of hyperparameters is expressed as:

$$\theta_{n+1} = \arg\max_{\theta \in \Theta} \alpha(\theta) \qquad (1)$$

Among them, $\alpha(\theta)$ is the acquisition function, and the next set of hyperparameters is selected by maximizing $\alpha(\theta)$. Through Bayesian optimization, the neural network model can find the optimal hyperparameter combination in a shorter time, thereby improving the performance and training efficiency of the model. In the model of this article, Bayesian optimization is combined with Bayesian neural network in order to search and select the best hyperparameter configuration more efficiently. Specifically, some adjustable hyperparameters in the model are used as search objects, and Bayesian optimization is performed in each training. Bayesian optimization updates the mean and variance functions of the proxy model in each iteration, and re-evaluates and adjusts each hyperparameter by selecting the acquisition function. This adjustment process makes the model's hyperparameter configuration more reasonable, effectively improves model performance and reduces the possibility

 doi:10.20944/preprints202409.2367.v2

of overfitting.

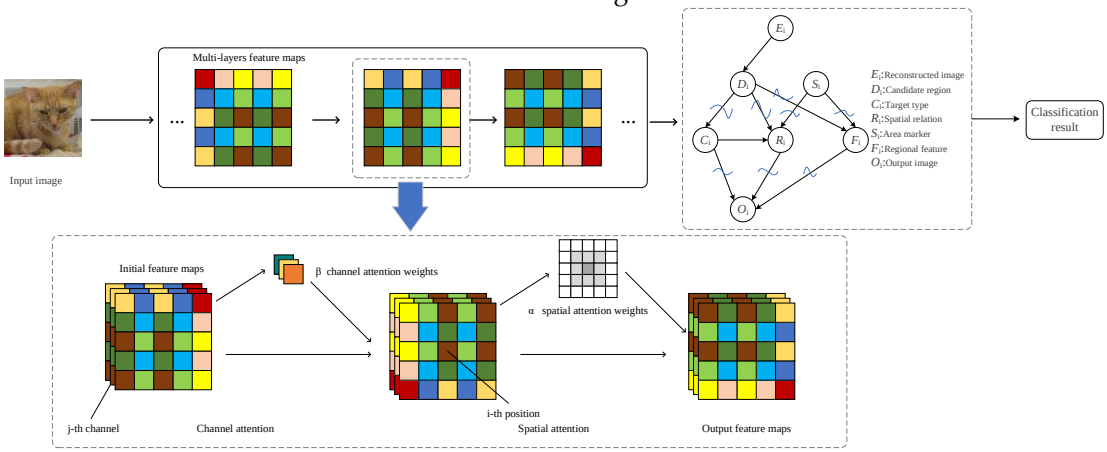

Figure 1. The classification network structure is proposedFusion Attention Mechanism.

The channel attention mechanism highlights the information of specific channels in the data by assigning different weights, thereby enhancing the expressive power of the model **Error! Reference source not found.**. The spatial attention mechanism pays more attention to the spatial position of the feature map. It expresses the importance of different positions in the form of weights, making the model pay more attention to useful spatial areas**Error! Reference source not found.**. In the fusion process, the feature map is first input into the channel attention module. Then, the global information of each channel is obtained by global average pooling and maximum pooling operations. Then, the pooled features computed from the shared multi-layer perceptron are fused to generate channel attention weights. These weights enhance the key information between channels by performing channel-wise weighting on the original feature maps. The channel attention is defined as follows:

$$V = AvgPool(X) + MaxPool(X)$$
$$\beta = MLP(V) \quad (2)$$

Among them, V is the feature of the previous layer, maxpool is the maximum pooling, avgpool is the average pooling, and MLP is the multi-layer perceptron model.The channel-weighted feature map is then input into the spatial attention module. In this part, the feature map first generates a spatial attention map through channel aggregation operations, such as average pooling and maximum pooling in the channel dimension. This image is used to represent the importance of each spatial position in those feature map, and then the feature map is weighted in the spatial dimension to retain important spatial information. The spatial attention is defined as follows:

$$\alpha = W_{ij}$$
$$X' = \alpha X_{ij} + b_{ij} \quad (3)$$

Among them, i and j represent the position. X' represents the output of the spatial attention module. The weights and biases are multiplied by the features position-wise. The order of channel first and then space can first extract the global information of important channels, and then further select key spatial regions from them, effectively focusing on the most useful features **Error! Reference source not found.**.The comprehensive attention process can be encapsulated as:

$$F' = F + \beta \otimes F \oplus b_c,$$
$$F'' = F' + \alpha \otimes F' \oplus b_s, \quad (4)$$

Where $\otimes$ denotes that the elements at each position are multiplied. β represents the channel attention weight, and α represents the spatial attention weight. 'b' represents the bias of each layer. β performs weighting on each channel and performs the first feature extraction. α performs weighting on the region in space and performs the second feature extraction. The extracted features are sent to the next stage.

Compared with multi-head attention mechanism, Fusion attention mechanism has higher computational efficiency and lower parameter number, because it relies on simple pooling and convolution operations to achieve attention allocation, which is suitable for embedding in lightweight

 doi:10.20944/preprints202409.2367.v2

devices. In addition, the Fusion attention mechanism has a simple structure and is easy to integrate directly into existing convolutional neural networks without significant changes to the backbone network. At the same time, fusing channel attention and spatial attention together can pay more attention to global and local features, which are mainly used for image classification tasks, while the multi-head attention mechanism is better at capturing long-distance dependent information, which often tends to ignore local information in the image **Error! Reference source not found.**.

Experimental Results

A. Experimental Design

In order to verify the effectiveness of the improved model in this paper, the CIFAR-100 dataset is used as the training and validation set. The dataset contains 60,000 color images, of which 50,000 are used for training and 10,000 are used for testing. The resolution of each image is 32x32. The dataset contains 20 superclasses. Each superclass has 5 subclasses, totaling 100 subclasses. Compared with the CIFAR-10 dataset with only 10 subclasses, this has stricter requirements on the generalization ability of the model. For the dataset, the pixel value of each image is standardized to a distribution with a mean of 0 and a standard deviation of 1. During the input process, random cropping and flipping are performed to increase data diversity. To accelerate the experiment, we use NVIDIA 3080 GPU as the experimental hardware, and use CUDA environment and Pytorch1.12 with CUDA environment to accelerate the experiment.

In the experiment, the proposed method is compared with classic models such as AlexNet, GoogleNet, VGG16 and ResNet. In the hyperparameter selection, the learning rate of these classic models is 0.01, the loss function momentum is 0.5, and the Dropout rate is 0 **Error! Reference source not found.**.

In order to comprehensively demonstrate the overall performance of the model, this article uses four commonly used indicators, including accuracy, precision, recall, and F1 score. Accuracy measures the classification accuracy of the model for the overall sample. Precision measures the prediction accuracy of the model for the positive sample. Recall evaluates the recognition ability of the model in the positive sample. The F1 score is calculated by precision and recall, which can comprehensively reflect the performance of the model on positive samples. By comprehensively evaluating the model through the above indicators, we can provide more comprehensive feedback on the classification performance of the model on the CIFAR-100 dataset.

A. Model performance verification

The results of the baseline model and the model proposed in this paper are shown in Table 1. As can be seen from Table 1, the improved model in this paper outperforms other classic models in all indicators. The accuracy of the improved model is 77.6%, which is higher than 65.2% of AlexNet, 70.8% of GoogleNet, 68.3% of VGG16 and 74.5% of ResNet, indicating that the improved model has better classification performance on the CIFAR-100 dataset and can more accurately identify image categories. The accuracy of the improved model is 78.3%, which is also significantly higher than other models, indicating that the improved model has a higher prediction accuracy for positive samples and fewer samples are misclassified as positive. The recall rate of the improved model is 76.8%, which is also significantly higher than other models, indicating that the improved model is more sensitive to positive samples and fewer samples are missed as negative samples. The F1 score of the improved model is 77.5%, which achieves a good balance between accuracy and recall and has the best overall performance. From the above results, it can be seen that the introduction of the fusion attention mechanism in the improved model plays a key role in feature extraction. The ability of the channel attention mechanism to dynamically assign weights to different channels strengthens the important features in the channel, while the spatial attention mechanism focuses on the significant areas in the image space. Both improve the accuracy of feature extraction. At the same time, the addition of Bayesian optimization enables the model to explore a deeper hyperparameter space, giving the model more parameter options and improving the generalization ability of the model. The addition of these changes has enabled the model to achieve better stability and performance, and also avoided overfitting.

 doi:10.20944/preprints202409.2367.v2

Table 1. Performance comparison of different models.

Model	Accuracy↑	Precision↑	Recall↑	F1 Score↑
AlexNet	65.20%	66.00%	64.50%	65.20%
GoogleNet	70.80%	71.50%	69.80%	70.60%
VGG16	68.30%	68.90%	67.40%	68.10%
ResNet	74.50%	75.20%	73.80%	74.50%
BNNAM	**77.60%**	**78.30%**	**76.80%**	**77.50%**

Table 2. Ablation experiment.

Model	Accuracy↑	Precision↑	Recall↑	F1 Score↑	Training Time(h)↓	Inference Time(s/images)↓
Baseline	74.50%	75.20%	73.80%	74.50%	5	0.05
without Bayesian	75.20%	75.80%	74.30%	75.00%	4.8	0.045
without Channel	76.00%	76.50%	75.20%	75.80%	4.6	0.04
without Spatial	75.50%	76.00%	74.70%	75.30%	4.7	0.042
without All	75.00%	75.60%	74.20%	74.90%	4.8	0.032
BNNAM	**77.60%**	**78.30%**	**76.80%**	**77.50%**	5.1	0.041
with multi-	**74.4%**	**74.8%**	**73.2%**	**73.2%**	5.3	0.06

A. Ablation Experiments

In order to evaluate the impact of each improvement technique on model performance, this paper trains and evaluates the above models on the CIFAR-100 dataset and records their accuracy, loss value, precision, recall, F1 score and other indicators. The results of the ablation experiment are shown in the table 2.

From the ablation experiment results in Table 2, it can be seen that removing different improvement techniques has different effects on the model performance. The improved model is better than the basic model in all evaluation indicators, especially the accuracy rate is increased by 3.1 percentage points. It shows that the introduced improvement technology significantly improves the model performance. After removing Bayesian optimization, the accuracy and other metrics of the model decreased, but were still better than the baseline model. This shows that Bayesian optimization plays an important role in hyperparameter tuning and can help the model find a better parameter combination. After removing the channel attention mechanism, the model

Table 3. Model metrics for different loss functions.

Model	Accuracy↑	Precision↑	Recall↑	F1 Score↑
Base Model	75.00%	73.00%	72.00%	72.50%

 doi:10.20944/preprints202409.2367.v2

w Cross-Entropy	76.00%	74.00%	73.00%	73.50%
w KL Divergence	75.00%	73.00%	72.00%	72.50%
Proposed Model	**77.60%**	**78.30%**	**76.80%**	**77.50%**

performance declined, especially in accuracy and F1 score, indicating that the channel attention mechanism plays an important role in enhancing the model's feature extraction capabilities. After removing the spatial attention mechanism, the model performance also declined, but the decline was slightly smaller than that of the channel attention. This shows that the spatial attention plays a key role in improving the model's attention to key areas. After removing all attention mechanisms, the model performance degrades further, but is still better than the base model. This shows that although the attention mechanism significantly contributes to performance improvement, Bayesian optimization also plays an important role in improving model performance. Through the above ablation experiments, we verified the effectiveness of Bayesian optimization, channel attention, and spatial attention in improving model performance. By using these improved techniques, the classification performance of the basic model on the CIFAR-100 data set is significantly improved.

From the ablation experiment results in Table 3, it can be seen that different loss functions have different effects on the model performance. The cross entropy loss is better than other loss functions in all evaluation indicators. Compared with the lowest accuracy, the accuracy is The rate increased by 3.6 percentage points, indicating that the introduced loss function significantly improved the model performance. After using the mean square error loss, the model performance has declined, especially in accuracy and precision, indicating that the cross-entropy loss is important in accurately quantifying errors, effectively guiding parameter updates, and significantly improving the classification accuracy of the model. effect. As can be seen from the training and inference time data, models using channel and spatial attention mechanisms ("w All Improve") demonstrate excellent performance without significantly increasing training and inference times. Compared with the baseline model, the training time of this model was reduced from 5 hours to 4.8 hours, and the inference time was reduced to 0.032 seconds/image. Compared with the 0.05 seconds of the baseline model and 0.06 seconds of the multi-head attention mechanism, it saved about 36 seconds. % to 47% of the time. This time cost saving is very important in practical applications, especially for scenarios that require real-time response such as edge computing and embedded systems.In addition, compared with the multi-head attention mechanism, channel and spatial attention not only optimize time efficiency, but also maintain or even exceed the baseline level in accuracy. Therefore, the combination of channel and spatial attention mechanisms can not only significantly reduce computational complexity while ensuring model performance, but also significantly reduce time and resource costs, making it an effective method to achieve a balance between performance and efficiency.Through the above ablation experiments, we verified the effectiveness of different loss functions in improving model performance. By using these improved techniques, the classification performance of the basic model on the CIFAR-100 data set is significantly improved.

Conclusions

This paper conducts detailed research and experiments on the performance improvement of classic neural network models in image classification tasks. This article adopts Bayesian optimization method in hyperparameter tuning, which significantly improves the convergence speed and final performance of the model. At the same time, this article introduces a channel attention mechanism into the model, allowing the model to adaptively adjust the importance of different channels. At the same time, the model also introduces a spatial attention mechanism so that the model can better capture the spatial relationships in the image and focus on key areas. Through experimental verification on the CIFAR-100 data set, the improved model is significantly better than the classic model in terms of accuracy, loss value, precision, recall and F1 score, especially the accuracy rate reached 77.6%, proving The effectiveness of the proposed improvement method is demonstrated. The

 doi:10.20944/preprints202409.2367.v2

successful application of the improved model not only demonstrates the potential of Bayesian optimization and attention mechanisms in improving model performance, but also provides new ideas for the design of future deep learning models. In the future, these improved technologies may be applied and promoted in a wider range of image classification tasks and other computer vision fields, further promoting the development of deep learning technology.

References

1. Bhattacharyya S. A brief survey of color image preprocessing and segmentation techniques[J]. Journal of Pattern Recognition Research, 2011, 1(1): 120-129.
2. Vega-Rodriguez M A. Feature extraction and image processing[J]. The Computer Journal, 2004, 47(2): 271-272.
3. Perreault, S., & Hébert, P. (2007). Median filtering in constant time. IEEE transactions on image processing, 16(9), 2389-2394.
4. Ślot, K., Kowalski, J., Napieralski, A., & Kacprzak, T. (1999). Analogue median/average image filter based on cellular neural network paradigm. Electronics Letters, 35(19), 1619-1620.
5. Direkoglu C, Nixon M S. Image-based multiscale shape description using Gaussian filter[C]//2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing. IEEE, 2008: 673-678.
6. Zhang, D., Liu, B., Sun, C., & Wang, X. (2011). Learning the Classifier Combination for Image Classification. J. Comput., 6(8), 1756-1763.
7. Tao, Y., Jia, Y., Wang, N., & Wang, H. (2019, July). The fact: Taming latent factor models for explainability with factorization trees. In Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval (pp. 295-304).
8. Amato G, Falchi F. Local feature based image similarity functions for knn classification[C]//International Conference on Agents and Artificial Intelligence. SCITEPRESS, 2011, 2: 157-166.
9. Joachims, T. (1999). Making large-scale svm learning practical. advances in kernel methods-support vector learning. http://svmlight. joachims. org/.
10. Xu, Y., Cai, Y., & Song, L. (2023). Latent fault detection and diagnosis for control rods drive mechanisms in nuclear power reactor based on GRU-AE. IEEE Sensors Journal, 23(6), 6018-6026.
11. Zhang, J., Wang, X., Ren, W., Jiang, L., Wang, D., & Liu, K. (2024). RATT: AThought Structure for Coherent and Correct LLMReasoning. arXiv preprint arXiv:2406.02746.
12. Lyu, W., Zheng, S., Ma, T., & Chen, C. (2022). A study of the attention abnormality in trojaned berts. arXiv preprint arXiv:2205.08305.
13. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).
14. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
15. Xu H, Yuan Y, Ma R, et al. Lithography hotspot detection through multi-scale feature fusion utilizing feature pyramid network and dense block[J]. Journal of Micro/Nanopatterning, Materials, and Metrology, 2024, 23(1): 013202-013202.
16. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
17. Dan, H. C., Yan, P., Tan, J., Zhou, Y., & Lu, B. (2024). Multiple distresses detection for Asphalt Pavement using improved you Only Look Once Algorithm based on convolutional neural network. International Journal of Pavement Engineering, 25(1), 2308169.
18. Tao, Y. (2023, August). Meta Learning Enabled Adversarial Defense. In 2023 IEEE International Conference on Sensors, Electronics and Computer Engineering (ICSECE) (pp. 1326-1330). IEEE.
19. Yu, L., Cao, M., Cheung, J. C. K., & Dong, Y. (2024). Mechanisms of non-factual hallucinations in language models. arXiv preprint arXiv:2403.18167.

20. Grabner M, Grabner H, Bischof H. Fast approximated SIFT[C]//Computer Vision–ACCV 2006: 7th Asian Conference on Computer Vision, Hyderabad, India, January 13-16, 2006. Proceedings, Part I 7. Springer Berlin Heidelberg, 2006: 918-927.
21. He L, Zou C, Zhao L, et al. An enhanced LBP feature based on facial expression recognition[C]//2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. IEEE, 2006: 3300-3303.
22. Déniz O, Bueno G, Salido J, et al. Face recognition using histograms of oriented gradients[J]. Pattern recognition letters, 2011, 32(12): 1598-1603.
23. Guan, R., Li, Z., Tu, W., Wang, J., Liu, Y., Li, X., ... & Feng, R. (2024). Contrastive multi-view subspace clustering of hyperspectral images based on graph convolutional networks. IEEE Transactions on Geoscience and Remote Sensing.
24. Guan, R., Li, Z., Li, X., & Tang, C. (2024, April). Pixel-superpixel contrastive learning and pseudo-label correction for hyperspectral image clustering. In ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 6795-6799). IEEE.
25. Xu, Y., Cai, Y. Z., & Song, L. (2023). Anomaly Detection for In-core Neutron Detectors Based on a Virtual Redundancy Model. IEEE Transactions on Instrumentation and Measurement.
26. Li, Y., Yu, X., Liu, Y., Chen, H., & Liu, C. (2023, July). Uncertainty-Aware Bootstrap Learning for Joint Extraction on Distantly-Supervised Data. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 1349-1358).
27. Li, C., Liu, X., Wang, C., Liu, Y., Yu, W., Shao, J., & Yuan, Y. (2024). GTP-4o: Modality-prompted Heterogeneous Graph Learning for Omni-modal Biomedical Representation. arXiv preprint arXiv:2407.05540.
28. Zhou, Y., Geng, X., Shen, T., Long, G., & Jiang, D. (2022, April). Eventbert: A pre-trained model for event correlation reasoning. In Proceedings of the ACM Web Conference 2022 (pp. 850-859).
29. Lyu, W., Zheng, S., Pang, L., Ling, H., & Chen, C. (2023). Attention-enhancing backdoor attacks against bert-based models. arXiv preprint arXiv:2310.14480.
30. Zhang, X., Wang, Z., Jiang, L., Gao, W., Wang, P., & Liu, K. (2024). TFWT: Tabular Feature Weighting with Transformer. arXiv preprint arXiv:2405.08403.
31. Sun, D., Liang, Y., Yang, Y., Ma, Y., Zhan, Q., & Gao, E. (2024). Research on Optimization of Natural Language Processing Model Based on Multimodal Deep Learning. arXiv preprint arXiv:2406.08838.
32. Liu, X., Dong, Z., & Zhang, P. (2024). Tackling data bias in music-avqa: Crafting a balanced dataset for unbiased question-answering. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 4478-4487).
33.
[1]. Zhang, Jingyu, et al. "Research on the Application of Computer Vision Based on Deep Learning in Autonomous Driving Technology." arXiv preprint arXiv:2406.00490 (2024).
34. Yin, Jianjian, et al. "Class Probability Space Regularization for semi-supervised semantic segmentation." Computer Vision and Image Understanding (2024): 104146.
35. Yin, Jianjian, et al. "Class-level multiple distributions representation are necessary for semantic segmentation." International Conference on Database Systems for Advanced Applications. Singapore: Springer Nature Singapore, 2024.
36. Yang, Qiming, et al. "Research on Improved U-net Based Remote Sensing Image Segmentation Algorithm." arXiv preprint arXiv:2408.12672 (2024).
37. Yukun, Song. "Deep Learning Applications in the Medical Image Recognition." American Journal of Computer Science and Technology 9.1 (2019): 22-26.
38. Weng, Yijie, and Jianhao Wu. "Leveraging Artificial Intelligence to Enhance Data Security and Combat Cyber Attacks." Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023 5.1 (2024): 392-399.
39. Chen, Ke, et al. "Dynasmile: Video-based smile analysis software in orthodontics." SoftwareX 29 (2025): 102004.
40. Xu, Shengjie, et al. "Neural Architecture Sizing for Autonomous Systems." 2024 ACM/IEEE 15th International Conference on Cyber-Physical Systems (ICCPS). IEEE, 2024.

41. Shangguan, Zhongkai, et al. "Neural process for black-box model optimization under bayesian framework." arXiv preprint arXiv:2104.02487 (2021).
42. Shang, Mingyang, et al. "V2F-Net: Explicit decomposition of occluded pedestrian detection." arXiv preprint arXiv:2104.03106 (2021).
43. Gong, Hao, and Mengdi Wang. "A duality approach for regret minimization in average-award ergodic markov decision processes." Learning for Dynamics and Control. PMLR, 2020.
44. Cheng, Qisen, Shuhui Qu, and Janghwan Lee. "72 - 3: Deep Learning Based Visual Defect Detection in Noisy and Imbalanced Data." SID Symposium Digest of Technical Papers. Vol. 53. No. 1. 2022.
45. Balakrishnan, Kaushik, et al. "6 - 4: Deep Learning for Classification of Repairable Defects in Display Panels Using Multi - Modal Data." SID Symposium Digest of Technical Papers. Vol. 54. No. 1. 2023.
46. Liu, Rui, et al. "Enhanced detection classification via clustering svm for various robot collaboration task." arXiv preprint arXiv:2405.03026 (2024).
47. Kang, Yixiao, et al. "6: Simultaneous Tracking, Tagging and Mapping for Augmented Reality." SID Symposium Digest of Technical Papers. Vol. 52. 2021.
48. Weng, Yijie. "Big data and machine learning in defence." International Journal of Computer Science and Information Technology 16.2 (2024): 25-35.
49. Ma, Danqing, et al. "Transformer-Based Classification Outcome Prediction for Multimodal Stroke Treatment." arXiv preprint arXiv:2404.12634 (2024).