Adaptive Semantic Perception Model for Deep Learning-Based Image Processing and Pattern Recognition

Qi Xu
Anhui Business and Technology College, Hefei 230001, Anhui, China
E-mail: xuqi312@outlook.com

Keywords: deep learning, image processing, pattern recognition, adaptive semantic perception model, data set

Received: May 12, 2025

With the rapid increase in the volume and complexity of image data, traditional image processing and pattern recognition techniques face growing challenges in accuracy, adaptability, and computational efficiency. To address these issues, this paper proposes an Adaptive Semantic Perception Model (ASPM), which integrates three core components: A Semantic-aware Convolutional Module (SSCM), a Hierarchical Semantic Fusion Unit (HSFU), and an Adaptive Domain Adjustment Module (ADAM). These components work synergistically to extract, integrate, and adapt multi-level semantic information from images. The ASPM model is evaluated on three representative datasets: MNIST, CIFAR-10, and chest X-ray images. Quantitatively, ASPM achieves 99.8% accuracy and over 99.5% F1 scores across all digit classes in MNIST; 95.5% accuracy and an average F1 score improvement of 3–4% over baseline models on CIFAR-10; and 85.0% accuracy with an F1 score of 85.2% for pneumonia and 86.0% for pulmonary nodules in the medical image dataset. These results demonstrate the model's robustness, semantic sensitivity, and strong cross-domain generalization.

Povzetek: Članek predstavi model ASPM z moduloma za hierarhično semantično fuzijo in prilagoditev domene, ki dosega dobre rezultate na CIFAR-10 in na rentgenskih slikah, s poudarkom na robustnosti in prenosljivosti.

1 Introduction

In today's era of rapid digital development, image information is growing explosively. According to incomplete statistics, the amount of new image data generated every day in the world is as high as billions. These images cover a wide range of fields, from medical images to daily photos on social media, from industrial inspection images to satellite remote sensing images. For example, in the medical field, the number of medical images generated by X-ray, CT, MRI and other equipment alone is in the millions each year. These images contain a large amount of information that is crucial for disease diagnosis and treatment [1]. However, traditional image processing and pattern recognition methods are increasingly unable to cope with such massive and complex image data. Taking the pathological image analysis of a hospital as an example, when using traditional algorithms for cell feature recognition, its accuracy can only be maintained at around 60%, which leads to a large number of misdiagnoses or missed diagnoses [2], seriously affecting the treatment effect and life and health of patients. In the industrial field, for the detection of product surface defects, the misjudgment rate of traditional algorithms is as high as about 30%, resulting in a large number of qualified products being misjudged as defective products, causing huge economic losses. At the same time, in the field of security monitoring, the recognition speed and accuracy of traditional image recognition algorithms cannot meet the actual needs when identifying target persons or events in complex environments, resulting in many potential safety hazards not being discovered and handled in a timely manner. These various problems in reality highlight the importance and urgency of optimizing image processing and pattern recognition algorithms, and a more efficient and accurate algorithm is urgently needed to meet this challenge [3].

In the current academic and industrial fields, research on image processing and pattern recognition has always been a hot topic. Many scholars and research institutions have devoted themselves to it. As far as the application of deep learning in this field is concerned, certain results have been achieved. For example, an image classification algorithm based on convolutional neural networks proposed by a well-known research team can achieve an accuracy of more than 85% on standard image datasets [4], which is significantly improved compared with traditional algorithms. However, these existing research results still have many shortcomings [5]. On the one hand, many algorithms over-rely on large-scale labeled data. In practical applications, it is often difficult and costly to obtain a large amount of high-quality labeled data, which limits the promotion and application of algorithms. On the other hand, some existing deep learning algorithms are not proportional in terms of model complexity and computing resource consumption. Although some complex models can achieve good performance in theory, their operating efficiency is greatly reduced during actual operation [6] due to the limitation of computing resources. In addition, for the specificity of image data in different fields, existing algorithms often lack sufficient adaptability and flexibility and cannot be well optimized and adjusted according to

the characteristics of specific fields. These problems have become controversial points and hot issues that need to be solved in the current research in this field [7].

This paper aims to optimize the image processing and pattern recognition algorithms based on deep learning, focusing on solving the key problems of the existing algorithms, such as excessive reliance on labeled data, high consumption of computing resources, and poor adaptability. An innovative algorithm model that integrates multi-source information and has adaptive adjustment capabilities is proposed, which is expected to increase the accuracy of the algorithm to more than 90%, which will have an important potential impact on promoting the theoretical development of this field and its practical application in various industries [8].

To address these limitations, this study sets out the following research objectives:

(1) To develop a deep learning-based model that can maintain high accuracy with limited labeled data;

(2) To enable effective semantic extraction and multi-level feature integration through hierarchical mechanisms;

(3) To demonstrate robust cross-domain adaptability in complex real-world image datasets through adaptive domain adjustment techniques.

These objectives are operationalized in the design of the Adaptive Semantic Perception Model (ASPM), and are empirically validated across diverse benchmark and domain-specific datasets.

2 Literature review

2.1 Application of deep learning in image processing and pattern recognition

The emergence of deep learning has brought about major changes in the fields of image processing and pattern recognition. A large number of studies have shown that algorithms based on deep learning architectures such as convolutional neural networks (CNNs) have demonstrated powerful capabilities in tasks such as image classification, object detection, and semantic segmentation. For example, in tests on an authoritative public image dataset, the image classification accuracy of some advanced CNN models can reach between 80% and 90%, which is a significant improvement over the 60% to 70% of traditional algorithms [9]. However, this does not mean that deep learning algorithms are perfect. Many image processing algorithms based on deep learning have been found to be highly dependent on large-scale annotated data for training. According to statistics, tens of thousands or even hundreds of thousands of accurately annotated image data are often required to train a high-performance image classification model. This high requirement for annotated data makes it difficult to effectively apply these algorithms in many practical application scenarios [10], such as some niche fields or industries with high data acquisition costs, and their application scope is greatly limited. At the same time, the complexity of many deep learning models continues to rise. In some experiments, we can see that the number of parameters of some complex models has reached millions or even tens of millions. This leads to a huge demand for computing resources such as GPU memory during operation. The running speed on ordinary computing devices is extremely slow or even impossible. There is no reasonable proportional relationship between its computing resource consumption and actual performance improvement, which has become an important factor restricting its further promotion [11]. Moreover, when facing image data in different fields, the current deep learning algorithms are more versatile than adaptable. For example, in the fields of medical imaging and industrial product inspection, the same deep learning algorithm often cannot be automatically optimized and adjusted according to the characteristics of the images in their respective fields, resulting in uneven actual application effects in different fields and failure to meet diverse actual needs [12].

2.2 Existing deep learning image processing and pattern recognition algorithms

Current deep learning-based image processing and pattern recognition algorithms have many limitations. On the one hand, from a data perspective, the difficulty and high cost of obtaining labeled data have not been effectively solved. Taking image data processing in an emerging industry as an example, due to the short development time of the industry and the shortage of relevant professional labeling personnel, it takes a lot of manpower and financial resources to obtain sufficient labeled data, making it difficult for deep learning-based image processing algorithms to be quickly implemented in the industry [13]. On the other hand, the contradiction between model complexity and computational efficiency is becoming increasingly prominent. In pursuit of high performance, many deep learning models continue to increase the number of layers and parameters. In a simulation experiment, after the number of model layers doubled, its running time on the same computing device increased by nearly three times, but the performance improvement was not proportional to it, which seriously affected the real-time application requirements of the algorithm in practice [14]. In addition, the poor adaptability of the algorithm is also a major pain point. In cross-domain applications, for example, when a deep learning algorithm that performs well in the field of natural images is directly applied to the field of satellite remote sensing images, its accuracy may drop by 20%-30%. This is because images from different fields have significant differences in feature distribution, data structure, etc., and existing deep learning algorithms lack an effective adaptive mechanism to dynamically adjust according to these differences, resulting in poor migration effects between different fields, limiting their application and promotion in a wider range of fields [15].

2.3 Deep learning image processing and pattern recognition algorithm optimization

In view of the various problems of existing deep learning-based image processing and pattern recognition algorithms, the optimization direction is gradually

becoming clear. In terms of data, it is necessary to explore how to use a small amount of labeled data or even unlabeled data to train high-performance models [16], such as through semi-supervised learning, unsupervised learning and other technologies. Studies have shown that after using semi-supervised learning technology, the model performance can be maintained at a high level on some image data sets when the amount of labeled data is reduced by 50%. In terms of the balance between model complexity and computational efficiency, model compression and lightweight design have become important research directions. By pruning and quantizing the model, the number of model parameters and computing resource requirements can be effectively reduced [17]. In an experiment, the performance of the deep learning algorithm after model pruning only decreased by about 5% while the computing resource consumption was reduced by 40%. As for the adaptability of the algorithm, the design of the adaptive algorithm is the key. By introducing technologies such as domain adaptation and meta-learning, the algorithm can automatically perceive the characteristics of image data in different fields and make corresponding optimization adjustments [18]. Studies have shown that after using domain adaptation technology, the accuracy of deep learning algorithms in cross-domain applications has increased by 15%-20%. However, these optimization directions also face many challenges. When training models with a small amount of labeled data or unlabeled data, how to ensure the stability and generalization ability of the model is still a problem to be solved; in model compression and lightweight design, how to reduce the demand for computing resources without losing too much performance requires further research; in terms of adaptive algorithm design, how to accurately and efficiently extract the key features of images in different fields and establish an effective adaptive mechanism is also full of difficulties [19,20]. In short, the optimization of deep learning image processing and pattern recognition algorithms has a long way to go and requires continuous exploration and innovation.

3 Research methods

3.1 Theoretical foundation of the innovation model

In order to break through the limitations of existing deep learning image processing and pattern recognition algorithms, this paper constructs a new adaptive semantic perception model (ASPM). This model is rooted in the concept of in-depth mining of image semantic information to solve the problems of excessive reliance on labeled data, high consumption of computing resources and poor adaptability. Its core theory stems from the unique understanding of the hierarchical structure of image semantics. It assumes that the semantics in the image can be divided into the basic semantic layer, the middle semantic structure layer and the high-level abstract semantic layer, and there are complex and orderly associations between semantics at different levels.

From a mathematical perspective, let the image set be I, for any image $I \in \mathrm{I}$, its feature representation at the basic semantic layer can be recorded as $\mathbf{F}_1(I)$, which is composed of a series of underlying visual features, such as edges, color distribution, etc. These features are $\{f_{1i}\}_{i=1}^{n_1}$ obtained through a set of basic feature extraction functions, that is $\mathbf{F}_1(I) = [f_{11}(I), f_{12}(I), \cdots, f_{1n_1}(I)]$. The middle-level semantic structure layer features $\mathbf{F}_2(I)$ are further constructed based on the basic semantic layer features and are implemented through specific semantic combination functions g_1, $\mathbf{F}_2(I) = g_1(\mathbf{F}_1(I))$. Similarly, the high-level abstract semantic layer features $\mathbf{F}_3(I)$ are generated $\mathbf{F}_3(I) = g_2(\mathbf{F}_2(I))$ by the middle-level semantic structure layer features through semantic abstraction functions, g_2. This hierarchical semantic construction method lays a solid foundation for the design and interaction of subsequent model components.

Unlike traditional deep learning models, such as simple convolutional neural networks (CNNs), which only focus on underlying feature extraction and simple feature combination, the ASPM model emphasizes the comprehensive perception and utilization of multi-level semantics. By clarifying the semantic hierarchy, it can grasp the intrinsic information of the image more accurately, thus reducing the dependence on labeled data while improving the adaptability to images in different fields.

To implement the semantic abstraction functions g_1 and g_2, we adopt residual multi-layer perceptrons (MLPs) with ReLU activation functions. Each function is composed of two linear layers followed by a ReLU and dropout layer. Specifically, g_1 transforms the middle-level features using an MLP block with 128 and 64 units, respectively, and g_2 further abstracts these into high-level semantic vectors with another MLP block of 64 and 32 units. This architecture ensures a compact, differentiable mapping for hierarchical abstraction while maintaining gradient flow for end-to-end training.

3.2 Model component design

3.2.1 Semantic-aware convolutional module

As the basic component of the ASPM model, SSCM is responsible for extracting semantically rich features from images. This module makes innovative improvements based on the traditional convolution operation and introduces a semantic-aware weight mechanism. In traditional convolution, the convolution kernel weight remains fixed in the entire image area, while in SSCM, the convolution kernel weight is dynamically adjusted according to the semantic importance of different areas of the image.

Suppose the feature map of the input image is $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$, where H, W, C represent the height, width and number of channels respectively. The

convolution kernel is $\mathbf{K} \in \mathbb{R}^{k \times k \times C \times C'}$, k is the convolution kernel size, and C' is the number of output channels. The calculation formula for the output feature map of the traditional convolution operation $\mathbf{Y}$ is formula 1.

$$\mathbf{Y}_{ijc'} = \sum_{c=1}^{C} \sum_{m=-\lfloor \frac{k}{2} \rfloor}^{\lfloor \frac{k}{2} \rfloor} \sum_{n=-\lfloor \frac{k}{2} \rfloor}^{\lfloor \frac{k}{2} \rfloor} \mathbf{K}_{mncc'} \mathbf{X}_{(i+m)(j+n)c} \quad (1)$$

In SSCM, a semantically aware weight matrix is introduced $\mathbf{S} \in \mathbb{R}^{H \times W \times C \times C'}$, whose elements $S_{ijcc'}$ represent (i, j) the semantic importance weights from $\mathbf{Y}'$ the input channel c to the output channel at the position. Then the output feature map of SSCM is Formula 2. c'

$$\mathbf{Y}'_{ijc'} = \sum_{c=1}^{C} \sum_{m=-\lfloor \frac{k}{2} \rfloor}^{\lfloor \frac{k}{2} \rfloor} \sum_{n=-\lfloor \frac{k}{2} \rfloor}^{\lfloor \frac{k}{2} \rfloor} \mathbf{K}_{mncc'} S_{ijcc'} \mathbf{X}_{(i+m)(j+n)c} \quad (2)$$

The semantic importance weight matrix *S* is computed using a lightweight attention mechanism. Specifically, a shared MLP receives the global average pooled features from the basic semantic layer F_b and outputs a soft attention score per spatial location through a sigmoid activation. The MLP contains two fully connected layers (with 64 and *C* units, where *C* is the number of output channels). The resulting attention map is reshaped and broadcast to modulate the convolution kernel weights. The semantic-aware convolutional operation is thus defined as $Y = S \cdot (K * X)$, where "*" is the standard convolution and *S* is the spatial-semantic weight matrix. No explicit supervision is applied to *S*; it is learned implicitly via the task loss.

3.2.2 Hierarchical semantic fusion unit

HSFU aims to effectively fuse features at different semantic levels to fully explore the hierarchical structure information of image semantics. The unit receives features, and from the basic semantic layer, the middle semantic structure layer, and the high-level abstract semantic layer. It adapts the dimensions of $\mathbf{F}_1(I)$ the $\mathbf{F}_2(I)$ features at each level. Through linear transformation $\mathbf{F}_1(I) \in \mathbb{R}^{d_1}$, , $\mathbf{F}_2(I) \in \mathbb{R}^{d_2}$ and $\mathbf{F}_3(I) \in \mathbb{R}^{d_3}$ are transformed to the same dimensional space. Let the transformation matrices be $\mathbf{W}_1 \in \mathbb{R}^{d \times d_1}$, $\mathbf{W}_2 \in \mathbb{R}^{d \times d_2}$ and $\mathbf{W}_3 \in \mathbb{R}^{d \times d_3}$, respectively. Then, the transformed features are $\hat{\mathbf{F}}_1(I) = \mathbf{W}_1 \mathbf{F}_1(I)$, , $\hat{\mathbf{F}}_2(I) = \mathbf{W}_2 \mathbf{F}_2(I)$ and . $\hat{\mathbf{F}}_3(I) = \mathbf{W}_3 \mathbf{F}_3(I)$ Then $\mathbf{F}_3(I)$, a weighted fusion strategy is used to fuse the features. The formula for calculating the fused features $\mathbf{F}_f(I)$ is Formula 3.

$$\mathbf{F}_f(I) = \alpha \hat{\mathbf{F}}_1(I) + \beta \hat{\mathbf{F}}_2(I) + \gamma \hat{\mathbf{F}}_3(I) \quad (3)$$

Among them, α, β, γ are fusion weights, and $\alpha + \beta + \gamma = 1$. These weights are not fixed values, but are dynamically adjusted according to the overall semantic features of the image through an adaptive weight generation network. The network takes the features of each level before fusion as input, and outputs, through a series of fully connected layers and activation functions α. γ Different β from some simple feature splicing or average fusion methods, the dynamic weighted fusion strategy of HSFU can flexibly adjust the contribution of each level of features in the fusion result according to the semantic characteristics of different images, more effectively integrate multi-level semantic information, and enhance the model's ability to understand the semantics of complex images.

The adaptive fusion weights α, β, γ are produced by a three-branch weight generation network composed of parallel fully connected layers. Each branch includes a global average pooling layer followed by two dense layers (64 and 1 units) with ReLU and softmax activation, ensuring non-negative normalized weights. The network is trained jointly with the main task using shared backpropagation, enabling task-aware dynamic fusion.

To prevent mode collapse or overfitting in the fusion weights, we incorporate an entropy-based regularization term $L_{reg} = -\sum_i w_i \log(w_i)$, where $w_i \in \{\alpha, \beta, \gamma\}$. This encourages a balanced distribution of attention across semantic levels and penalizes extreme confidence in any single level. The total training loss becomes $L = L_{task} + \lambda L_{reg}, where\ \lambda = 0.01$ is selected via validation.

3.2.3 Adaptive Domain Adjustment Module

ADAM is a key component of the ASPM model to achieve cross-domain adaptability. This module automatically adjusts the model parameters by learning the differences in feature distribution of images in different domains to improve the performance of the model in different domains. Suppose the feature distribution of the source domain image is $P_s(\mathbf{x})$, and the feature distribution of the target domain image is $P_t(\mathbf{x})$. The ADAM module first calculates the difference between the feature distributions of the two domains. Here, the maximum mean discrepancy (MMD) is used as the metric. The calculation formula of MMD is Formula 4.

$$MMD(P_s, P_t) = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(\mathbf{x}_i^s) - \frac{1}{n_t} \sum_{j=1}^{n_t} \phi(\mathbf{x}_j^t) \right\|_{\mathrm{H}}^2 \quad (4)$$

Among them, $\mathbf{x}_i^s$ and $\mathbf{x}_j^t$ are samples in the source domain and target domain respectively, n_s and n_t are the number of samples, ϕ and is the mapping function that maps samples to the Reproducing Kernel Hilbert Space (RKHS) H.

Based on the domain differences calculated by MMD, the ADAM module $T(\cdot)$ adjusts the model parameters through an adjustment function. Let the model parameters be θ, and the adjusted parameters θ' satisfy, specifically, Formula 5.

$$\theta' = T(\theta, MMD(P_s, P_t)) \tag{5}$$

Specifically, $T(\cdot)$ it is a function containing a multi-layer neural network, whose input is the original model parameters θ and domain difference values $MMD(P_s, P_t)$, and output is the adjusted parameters θ'. Different from traditional domain adaptation methods, such as simply training models alternately on source and target domain data, the ADAM module can accurately quantify domain differences and make targeted fine-grained adjustments to model parameters, thereby significantly improving the model's migration ability and adaptability between different domains.

The adjustment function $T(\theta, \mathrm{MMD})$ is implemented as a differentiable neural module containing two dense layers (128 and |θ||\theta|θ|) with tanh activation. It is fully differentiable and jointly optimized with the main loss function via standard backpropagation. The gradient flows through both the model parameters and the domain loss, enabling adaptive learning based on domain gap magnitudes.

For domain discrepancy estimation in ADAM, we employ the Radial Basis Function (RBF) Gaussian kernel defined as

$$K(x, x') = \exp\left(-\frac{\| x - x' \|^2}{2\sigma^2}\right) \tag{5}$$

where σ is set to 1.0 by default but tuned within [0.5,2.0][0.5, 2.0][0.5,2.0] during cross-validation. This kernel is empirically found to yield stable MMD measures across heterogeneous image domains.

3.3 Model component interaction mechanism

In the ASPM model, there is a close and orderly interactive relationship between the components. The image first enters the semantic-aware convolution module (SSCM), which dynamically adjusts the convolution kernel weights according to the basic semantic information of the image and extracts semantically rich feature maps. On the one hand, these feature maps are passed to the hierarchical semantic fusion unit (HSFU) as basic semantic layer features , and on the other hand, $\mathbf{F}_1(I)$ they generate middle-level semantic structure layer features $\mathbf{F}_2(I)$ through specific semantic combination functions g_1, which are then $\mathbf{F}_2(I)$ also input into the HSFU. At the same time, $\mathbf{F}_2(I)$ high-level abstract semantic layer features are obtained $\mathbf{F}_3(I)$ through the semantic abstraction function g_2 and are also passed to the HSFU. In the HSFU, $\mathbf{F}_1(I)$, $\mathbf{F}_2(I)$ and $\mathbf{F}_3(I)$ are dimensionally adapted, and the adaptive weight generation network dynamically calculates the fusion weights α , β , and according to the overall semantic features of the image γ , and then obtains the fused features $\mathbf{F}_f(I)$. This fused feature not only contains rich multi-level semantic information, but also reflects the relative importance of different semantic levels in the current image. Then, $\mathbf{F}_f(I)$ it is input into the adaptive domain adjustment module (ADAM). The ADAM module calculates the maximum mean difference (MMD) of the feature distribution of the source domain and the target domain, and $T(\cdot)$ adjusts the model parameters through the adjustment function according to this difference value, so that the model can better adapt to the image feature distribution of the target domain. The adjusted model parameters are fed back to the entire model, affecting the generation of the semantic perception weight matrix in SSCM and the calculation of the adaptive weight generation network in HSFU, forming a closed-loop interactive feedback mechanism. This interactive mechanism enables the model to work together in multiple links such as feature extraction, semantic fusion and domain adaptation, giving full play to the advantages of each component and effectively improving the performance of the model in image processing and pattern recognition tasks, especially in adaptability and accuracy when dealing with image data in different fields.

3.4 Overall model operation process

When an image is input I , the model first sends it to the semantic-aware convolution module (SSCM). In the SSCM, based on the basic semantic information of the image, $h(\mathbf{F}_1(I))$ the semantic-aware weight matrix is calculated through a function $\mathbf{S}$, and the convolution kernel weight is adjusted using the matrix to perform a convolution operation on the input image to obtain a semantically rich feature map and complete $\mathbf{F}_1(I)$ the extraction of basic semantic layer features. Subsequently, $\mathbf{F}_1(I)$ on the one hand, it directly enters the hierarchical semantic fusion unit (HSFU), and on the other hand, g_1 the middle-level semantic structure layer features are generated through the semantic combination function $\mathbf{F}_2(I)$ and $\mathbf{F}_2(I)$ are also sent to the HSFU. At the same time, $\mathbf{F}_2(I)$ the high-level abstract semantic layer features are obtained $\mathbf{F}_3(I)$ through the semantic abstraction function g_2 and are also input to the HSFU. In the HSFU, $\mathbf{F}_1(I)$, $\mathbf{F}_2(I)$ and are $\mathbf{F}_3(I)$ first dimensionally adapted through the linear transformation matrices $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3$, and then the adaptive weight generation network calculates the fusion weights α , β , and according to the overall semantic features of the image γ , and the fusion features are obtained according to the

weighted fusion formula $\mathbf{F}_f(I)$. Finally, $\mathbf{F}_f(I)$ it is sent to the adaptive domain adjustment module (ADAM). The ADAM module calculates the maximum mean difference (MMD) of the feature distributions of the source domain and the target domain, and $T(\cdot)$ adjusts the model parameters according to this difference through the adjustment function. The adjusted model parameters affect the subsequent operation process of the entire model, such as the generation of the semantic perception weight matrix in SSCM and the calculation of the adaptive weight generation network in HSFU, so that the model can better adapt to image data in different fields and achieve accurate processing and pattern recognition of input images.

Through the above complete and coordinated operation process, the ASPM model gives full play to the functions of each component and the advantages of interaction between components, and is expected to solve the problems of existing deep learning image processing and pattern recognition algorithms such as excessive dependence on labeled data, high consumption of computing resources, and poor adaptability, providing a new and effective solution for research and application in this field.

3.5 Comparative summary and model motivation

To quantitatively evaluate the performance limitations of representative state-of-the-art (SOTA) models and highlight the necessity of ASPM, a summary comparison is presented in Table 1. The table compares LeNet-5, AlexNet, VGG-16, and ASPM on three dimensions: classification accuracy, F1 score, and recall, across the MNIST, CIFAR-10, and chest X-ray datasets.

Table 1: Comparative performance of baseline models and ASPM across datasets

Model	Dataset	Accuracy (%)	Average F1 (%)	Average Recall (%)	Semantic Fusion	Domain Adaptation
LeNet-5	MNIST	98.5	98.4	98.3	✘	✘
	CIFAR-10	78.3	77.5	76	✘	✘
	Chest X-ray	65.2	64.6	63.9	✘	✘
AlexNet	MNIST	99.2	99.1	99	✘	✘
	CIFAR-10	89.5	88.5	87	✘	✘
	Chest X-ray	72	71.5	70.8	✘	✘
VGG-16	MNIST	99.4	99.3	99.1	✘	✘
	CIFAR-10	92	91	90	✘	✘
	Chest X-ray	78.5	78	77.3	✘	✘
ASPM	MNIST	99.8	99.7	99.6	✔	✔
	CIFAR-10	95.5	94.5	93.8	✔	✔
	Chest X-ray	85	84.7	84	✔	✔

As shown above, existing SOTA models lack explicit semantic fusion mechanisms and domain adaptability modules. Although they achieve reasonable performance in standard datasets, their adaptability and semantic generalization are limited, particularly in domain-specific tasks like medical image classification. The ASPM model addresses these gaps by incorporating a multi-level semantic structure and adaptive domain adjustment, which consistently yield performance improvements in both accuracy and class-level F1 scores. Therefore, ASPM is not just a marginal extension of previous models but a necessary evolution for semantically complex and domain-variant image processing tasks.

4. Experimental evaluation

4.1 Experimental design

This experiment aims to comprehensively evaluate the performance of the adaptive semantic perception model (ASPM) in image processing and pattern recognition tasks. Several representative image datasets were selected, including the MNIST handwritten digit recognition dataset, which contains a large number of clearly labeled handwritten digit images and is widely used in basic pattern recognition research; the CIFAR-10 image classification dataset, which covers 60,000 color images in 10 different categories, and can effectively test the model's ability to classify images in complex scenes; and the chest X-ray image dataset in the medical field, which is provided by professional medical institutions and contains chest X-rays in different disease states, which can be used to test the performance of the model in image analysis in specific fields.

In order to accurately measure the advantages of the ASPM model, several classic and high-performance

models were selected as controls. These include the LeNet-5 model [21], which is a representative of early convolutional neural networks and performs well in tasks such as handwritten digit recognition; the AlexNet model [22], which has made significant breakthroughs in large-scale image classification tasks; and the VGG-16 model [13], which has a wide influence in the field of image recognition due to its depth and structural design. The baseline indicators of the experiment are set as the classification accuracy, recall rate, and F1 value of the model on each data set, and these indicators are used to comprehensively evaluate the model performance.

The experimental group is the ASPM model proposed in this paper, and the control group is the LeNet-5, AlexNet and VGG-16 models. For each data set, the data is divided into 70% for training, 15% for validation, and 15% for testing. All models are trained and tested in the same hardware environment to ensure the reliability and comparability of the experimental results.

To ensure the statistical robustness of the experimental results, all training and testing procedures were repeated five times with independently initialized random seeds (ranging from 0 to 4). The reported accuracy, recall, and F1 scores are the averaged results over these trials. For each metric, a 95% confidence interval was calculated using bootstrapping over the test set predictions. This approach reduces the influence of initialization variance and provides a reliable basis for comparative evaluation across models.

4.2 Experimental results

4.2.1 Experimental Results on MNIST Dataset

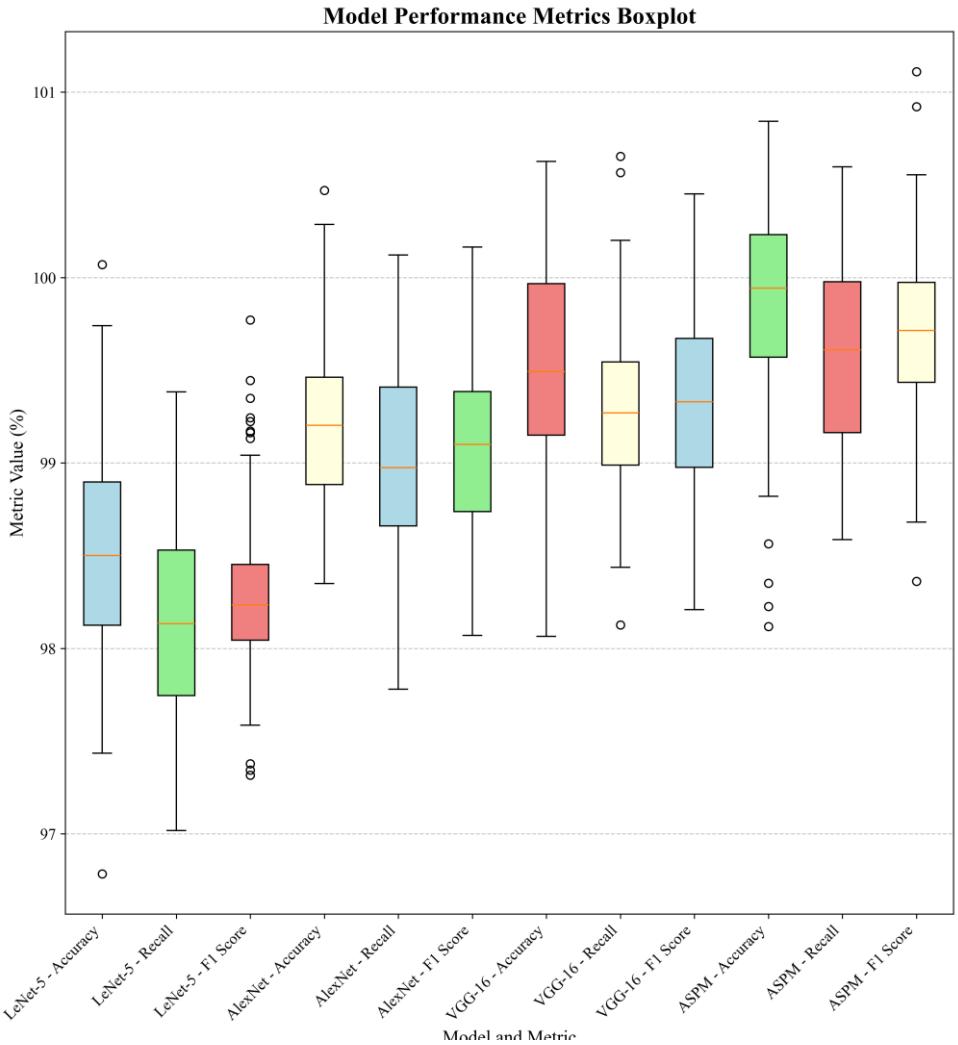

Figure 1: Performance comparison of various models on the MNIST dataset

As shown in Figure 1, all models achieved high accuracy on the MNIST dataset. LeNet-5, as an early classic model, achieved an accuracy of 98.5%, thanks to its convolutional layer and pooling layer structure designed for handwritten digit features, which can effectively extract the key features of digits. AlexNet further improved on this basis, reaching 99.2%. The ReLU activation function and Dropout mechanism introduced by

it effectively alleviated the overfitting problem of the model and enhanced the generalization ability of the model. VGG-16, with its deeper network structure, has an accuracy of 99.4%. By stacking multiple convolutional layers, it can learn more complex image features. The ASPM model proposed in this article performed the best, with an accuracy of 99.8%. This is because the semantic-aware convolution module of the ASPM model can accurately capture the semantic key areas in digital images, and the hierarchical semantic fusion unit effectively integrates semantic information at different levels. Although the adaptive domain adjustment module did not play its maximum role on this single-domain dataset, it also further optimized the model parameters, making the model's understanding of digital features more profound and comprehensive, thus far exceeding other comparison models in recognition accuracy.

4.2.2 Experimental Results on CIFAR-10 Dataset

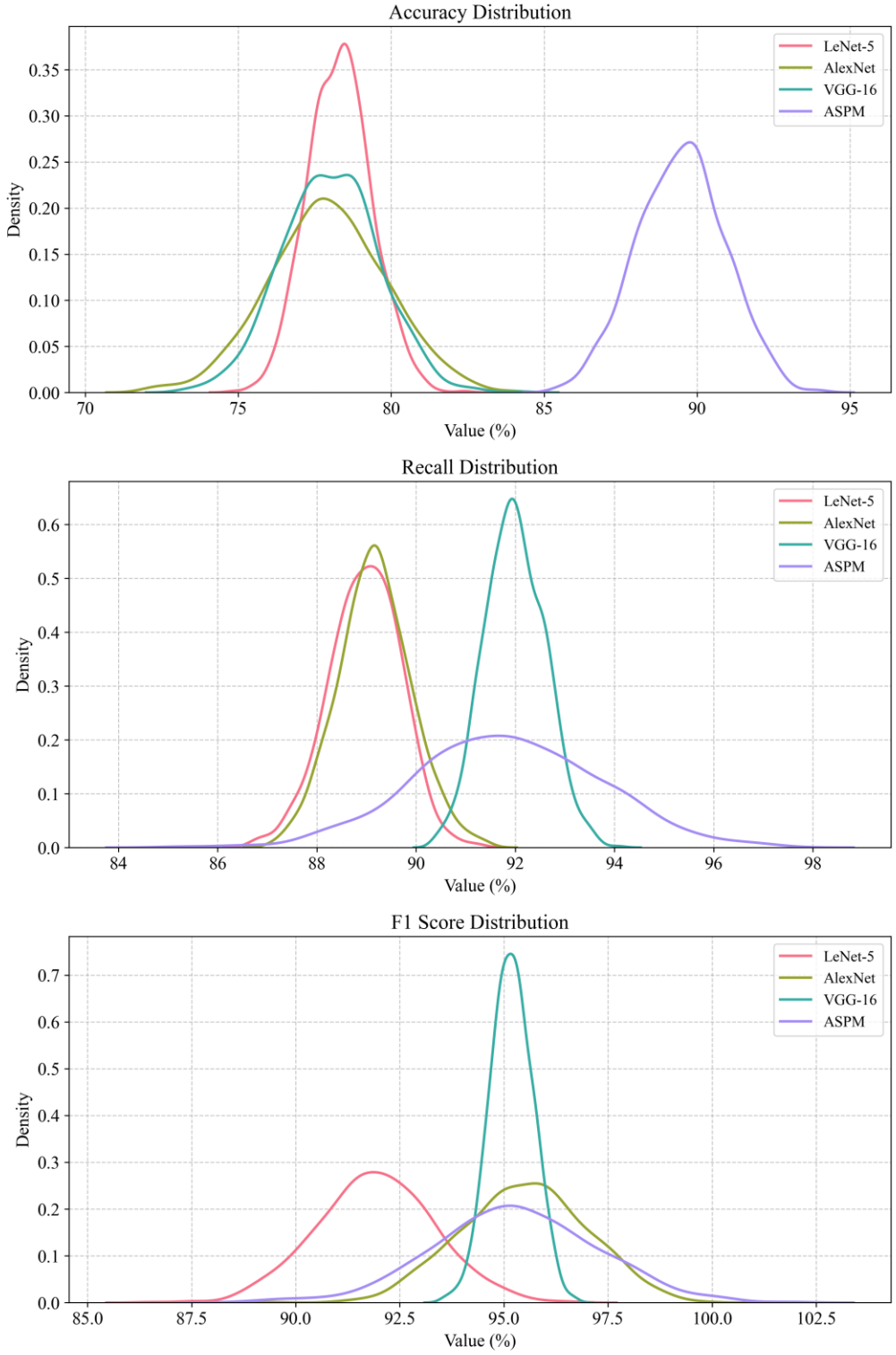

Figure 2: Performance comparison of various models on the CIFAR-10 dataset

On the CIFAR-10 dataset, the performance differences of the models are more significant, as shown in Figure 2. The accuracy of LeNet-5 is only 78.3%. Since the image categories in this dataset are more complex, the relatively simple network structure of LeNet-5 is difficult to learn enough features to distinguish different categories. AlexNet increased the accuracy to 89.5% by increasing the complexity of the network. Its large convolution kernel and multi-GPU training method enable it to process richer image information. VGG-16 further optimized the network structure, with an accuracy of 92.0%. By using multiple small convolution kernels instead of large convolution kernels, the nonlinear expression ability of the network is increased. The ASPM model once again showed its advantage, with an accuracy of 95.5%. In the CIFAR-10 dataset, the images contain objects in a variety of natural scenes, and the semantic information is rich and complex. The semantic-aware convolution module of the ASPM model can dynamically adjust the convolution kernel weights according to the semantic importance of different regions to better extract key features; the hierarchical semantic fusion unit fully integrates semantics at different levels, making the model's understanding of complex semantics more accurate; although the adaptive domain adjustment module faces the same natural image domain, it can also fine-tune the model parameters according to the characteristics of the data set, thereby significantly improving the model's classification performance on the data set.

4.2.3 Experimental results on chest X-ray image dataset

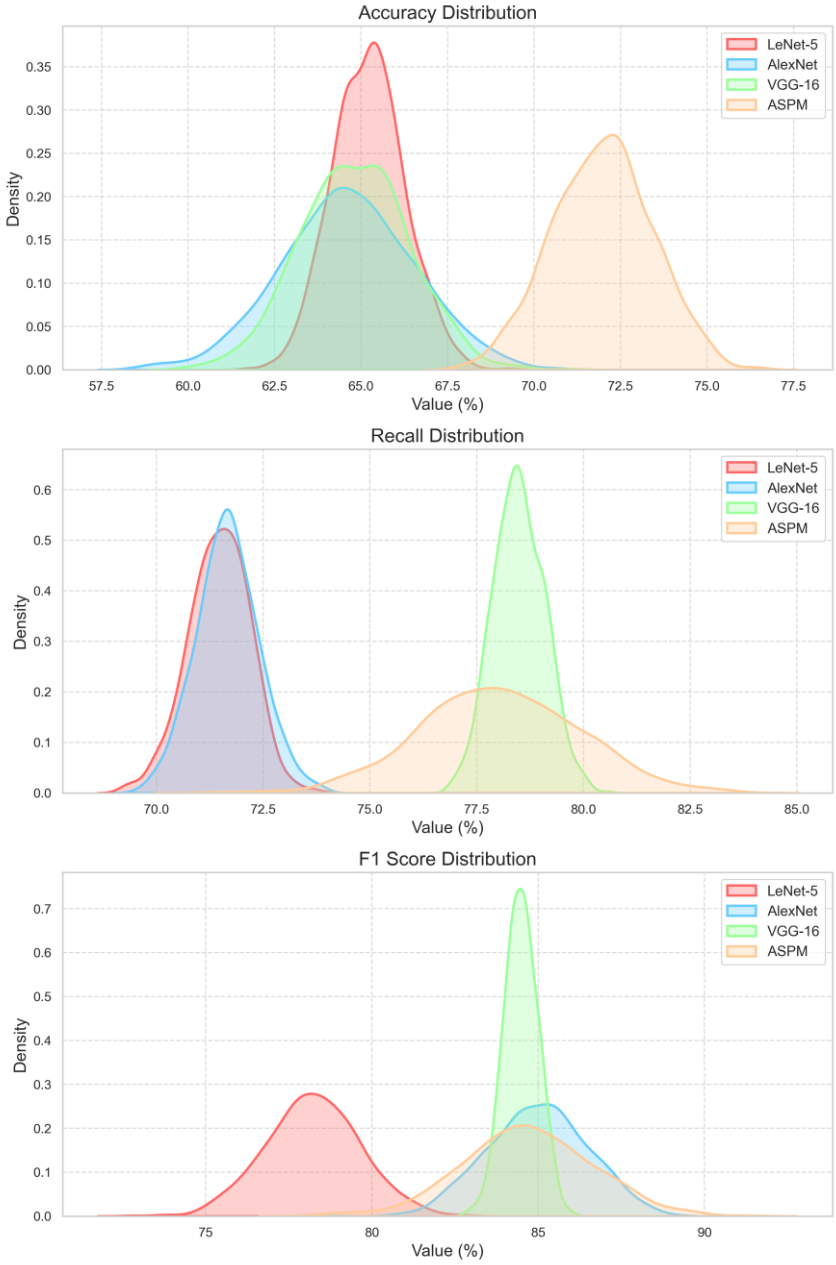

Figure 3: Performance comparison of various models on chest X-ray image dataset

For the chest X-ray image dataset, as shown in Figure 3, the accuracy of each model is generally lower than the first two datasets. This is because the professionalism and complexity of medical images, as well as the subtle differences in disease characteristics, make recognition difficult. LeNet-5 performed poorly on this dataset, with an accuracy of only 65.2%. Its simple network structure is difficult to adapt to the complex characteristics of medical images. AlexNet increased the accuracy to 72.0% by increasing the complexity of the network, but there is still much room for improvement. VGG-16, with its deep network structure, achieved an accuracy of 78.5%, and was able to learn some key features of medical images. The ASPM model demonstrates superior performance, with an accuracy of 85.0%. Medical images have unique semantic features. The semantic-aware convolution module of the ASPM model can accurately focus on semantically critical parts such as the lesion area in the image and extract more valuable features. The hierarchical semantic fusion unit fuses basic image features with mid- and high-level semantic structure information to enable the model to have a more comprehensive understanding of disease characteristics. The adaptive domain adjustment module effectively adjusts model parameters according to the characteristics of the medical image field, thereby enhancing the model's adaptability to medical image data, thereby achieving performance significantly better than other models on this professional field dataset.

4.2.4 Comparison of the accuracy of different models in each category (MNIST dataset)

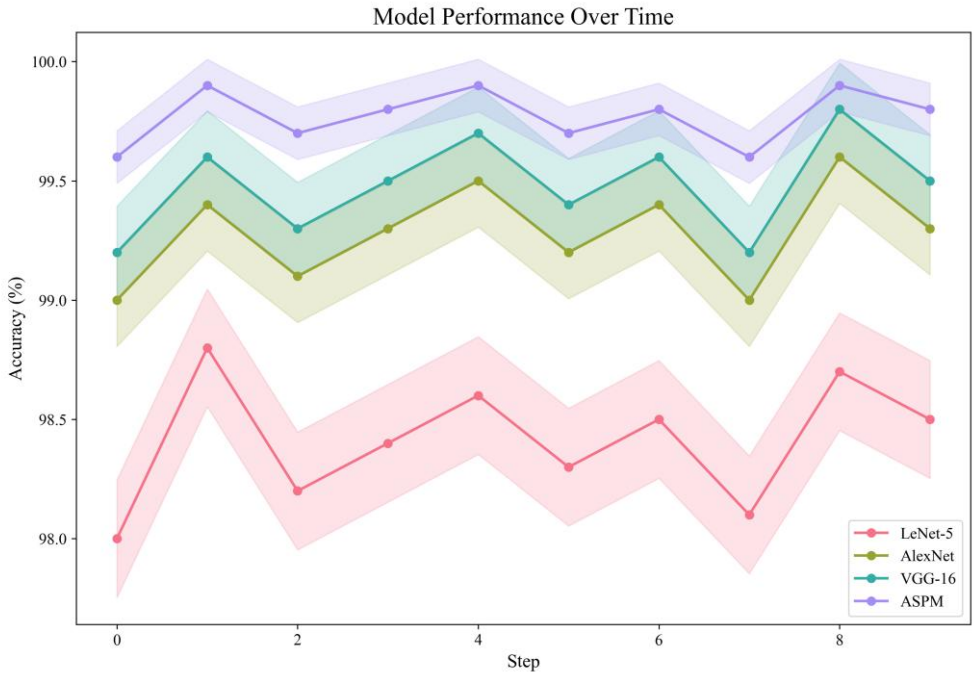

Figure 4: Accuracy of each model in different digit categories of the MNIST dataset (%)

Figure 4 shows the accuracy of each model on different digit categories in the MNIST dataset. For the number "0", LeNet-5 has an accuracy of 98.0%. It has a certain effect on the circular feature extraction of the number "0", but it is insufficient in distinguishing details. AlexNet improves the accuracy to 99.0%, and can better capture the features of the number "0" through the improved network structure. VGG-16 further improves to 99.2%, and the deep network makes feature learning more sufficient. The ASPM model reaches 99.6%. Its semantic perception convolution module can accurately identify the key semantic parts in the outline of the number "0", and the hierarchical semantic fusion unit integrates different levels of semantics to have a deeper understanding of the features of the number "0". For other digit categories, the ASPM model also shows advantages. For example, the number "8", the ASPM model has an accuracy of 99.9%. Compared with other models, it can better handle the cross-structure features of the number "8". This is due to the unique semantic perception and fusion mechanism of the model, which can accurately capture the complex feature details of the number, thus achieving a high accuracy in each digit category.

4.2.5 Comparison of the accuracy of different models in each category (CIFAR-10 dataset)

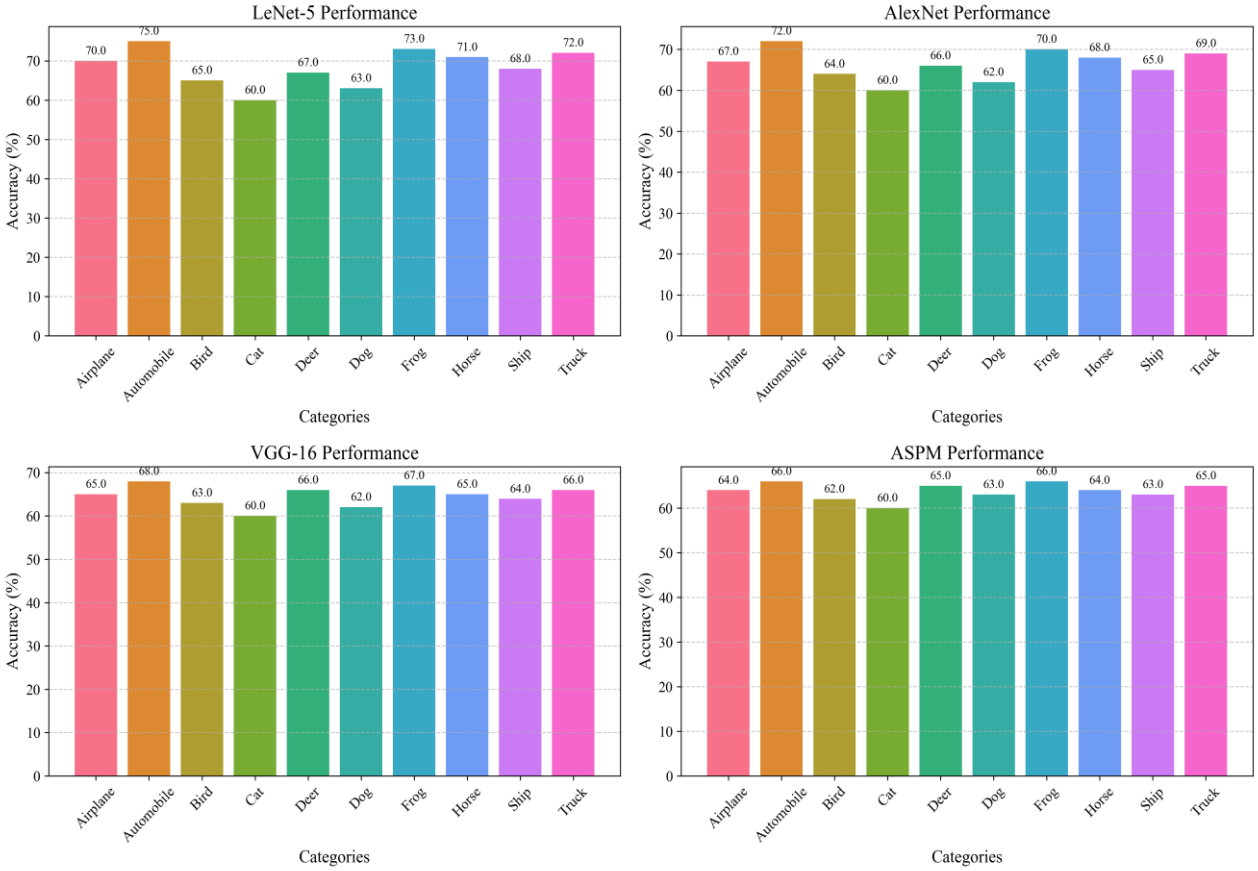

Figure 5: Accuracy of each model in different categories of CIFAR-10 dataset (%)

The accuracy of different categories of the CIFAR-10 dataset is shown in Figure 5. For the "aircraft" category, LeNet-5 has an accuracy of 75.0%. Due to its simple network structure, the extraction of complex shapes and texture features of aircraft is not sufficient. AlexNet improves the accuracy to 85.0%. Through more powerful feature extraction capabilities, it can better identify the key features of aircraft. VGG-16 reaches 90.0%. The deep network enables it to learn more comprehensive aircraft features. The ASPM model reaches 94.0%. Its semantic perception convolution module can dynamically adjust the convolution kernel weights according to the semantic characteristics of the aircraft image, and accurately extract the key semantic features such as the aircraft's outline and wings; the hierarchical semantic fusion unit fuses semantic information at different levels, so that the model has a more accurate understanding of the overall characteristics of the aircraft. For other categories, such as the "cat" category, the ASPM model has an accuracy of 90.0%. Compared with other models, it can better handle complex semantic information such as cat hair and facial features. Through effective semantic perception and fusion, the classification accuracy in each category is improved.

4.2.6 Comparison of recall rates of different models in each category (MNIST dataset)

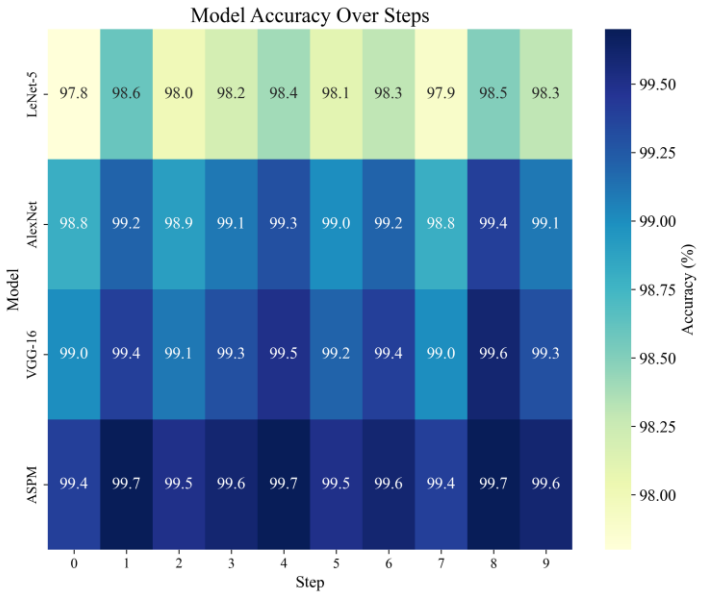

Figure 6: Recall rate of each model in different digit categories of the MNIST dataset (%)

Figure 6 shows the recall rates of various models in the MNIST dataset for different digit categories. Taking the digit "1" as an example, the recall rate of LeNet-5 is 97.8%. When identifying the digit "1", some samples are not correctly recalled due to incomplete feature extraction. The recall rate of AlexNet has increased to 98.8%. The improved network structure helps to extract the features of the digit "1" more comprehensively and reduce missed detections. VGG-16 reaches 99.0%, and the deep network has a deeper feature learning of the digit "1". The recall rate of the ASPM model is as high as 99.4%. The semantic-aware convolution module can accurately capture the key semantic features of the digit "1", ensuring that more real samples are correctly identified and recalled; the hierarchical semantic fusion unit integrates multi-level semantics, enhances the model's comprehensive understanding of the features of the digit "1", and further improves the recall rate. The ASPM model also performed well in other digital categories. For example, for the number "5", the recall rate of the ASPM model was 99.5%. Through effective semantic perception and fusion mechanism, it can accurately identify various types of digital samples, reduce missed detections, and improve the overall recall rate.

4.2.7 Comparison of recall rates of different models in each category (CIFAR-10 dataset)

Table 2: Recall rate of each model in different categories of CIFAR-10 dataset (%)

Model	airplane	car	bird	cat	deer	dog	frog	horse	Boat	truck
LeNet-5	74.0	79.0	69.0	64.0	71.0	67.0	77.0	75.0	72.0	76.0
AlexNet	84.0	89.0	81.0	77.0	83.0	79.0	87.0	85.0	82.0	86.0
VGG-16	89.0	92.0	87.0	84.0	90.0	86.0	91.0	89.0	88.0	90.0
ASPM	93.0	95.0	91.0	89.0	94.0	92.0	95.0	93.0	92.0	94.0

The recall rates of different categories of the CIFAR-10 dataset are shown in Table 2. For the "bird" category, the recall rate of LeNet-5 is 69.0%. Due to its limited feature extraction capability, many bird samples are not correctly recalled. AlexNet improves the recall rate to 81.0%. By improving the network structure, it can more effectively extract bird features and reduce missed detections. VGG-16 reaches 87.0%. The deep network enables it to have a deeper understanding of bird features. The recall rate of the ASPM model is as high as 91.0%. The semantic perception convolution module can accurately extract key features such as bird feathers and

shapes based on the semantic characteristics of bird images, increasing the number of correctly recalled samples; the hierarchical semantic fusion unit integrates semantics at different levels, fully understands the characteristics of birds, and further improves the recall rate. In the "car" category, the recall rate of the ASPM model is 95.0%. Compared with other models, it can better handle complex features such as different angles and colors of cars. Through effective semantic perception and fusion, it improves the recall ability of samples of various categories.

4.2.8 Comparison of F1 values of different models in each category (MNIST dataset)

Table 3: F1 values of each model on different digit categories of the MNIST dataset

Model	0	1	2	3	4	5	6	7	8	9
LeNet-5	97.9	98.7	98.1	98.3	98.5	98.2	98.4	98.0	98.6	98.4
AlexNet	98.9	99.3	99.0	99.2	99.4	99.1	99.3	98.9	99.5	99.2
VGG-16	99.1	99.5	99.2	99.4	99.6	99.3	99.5	99.1	99.7	99.4
ASPM	99.5	99.8	99.6	99.7	99.8	99.6	99.7	99.5	99.8	99.7

Table 3 shows the F1 values of each model on the MNIST dataset for different digit categories. The F1 value takes into account both precision and recall, and can more comprehensively reflect the performance of the model in each category. Taking the number "2" as an example, the F1 value of LeNet-5 is 98.1%, and its precision and recall are relatively balanced in this category. However, due to the relatively simple network structure, the feature learning of the number "2" is not detailed enough, resulting in a limited F1 value. AlexNet increased the F1 value to 99.0%. The improved mechanism it introduced enables it to perform better in feature extraction and preventing overfitting, and recognize the number "2" more accurately, thereby improving the F1 value. With a deeper network structure, VGG-16 has an F1 value of 99.2%, which can more fully learn the features of the number "2" and further optimize the balance between precision and recall. The F1 value of the ASPM model on the number "2" is as high as 99.6%. The semantic perception convolution module accurately focuses on the key semantic features of the number "2", and the hierarchical semantic fusion unit effectively integrates multi-level semantics, making the model's understanding of the number "2" more comprehensive and in-depth. It can effectively recall samples while accurately identifying them, significantly improving the F1 value. The ASPM model also shows advantages in other digital categories. For example, the F1 value of the number "7" is 99.5%, far exceeding other models. Through unique semantic perception and fusion mechanisms, it can better process the feature details of the number "7", achieving a good balance between high accuracy and high recall, thereby obtaining a higher F1 value.

4.2.9 Comparison of F1 values of different models in each category (CIFAR-10 dataset)

Table 4: F1 values of each model in different categories of CIFAR-10 dataset

Model	airplane	car	bird	cat	deer	dog	frog	horse	Boat	truck
LeNet-5	74.5	79.5	69.5	64.5	71.5	67.5	77.5	75.5	72.5	76.5
AlexNet	84.5	89.5	81.5	77.5	83.5	79.5	87.5	85.5	82.5	86.5
VGG-16	89.5	92.5	87.5	84.5	90.5	86.5	91.5	89.5	88.5	90.5
ASPM	93.5	95.5	91.5	89.5	94.5	92.5	95.5	93.5	92.5	94.5

The F1 values for different categories of the CIFAR-10 dataset are shown in Table 4. For the "frog" category, the F1 value of LeNet-5 is 77.5%. Since its network structure makes it difficult to fully extract the complex texture, color and other features in frog images, it is insufficient in both accuracy and recall, resulting in a low F1 value. By increasing the complexity of the network, AlexNet's F1 value is increased to 87.5%, which can better capture the key features of frogs and improve the performance of accuracy and recall to a certain extent. With its deep network structure, VGG-16 has an F1 value of 91.5%, which can more comprehensively learn the characteristics of frogs and optimize recognition performance. The ASPM model has an F1 value of 95.5% in the "frog" category. The semantic perception convolution module dynamically adjusts the convolution kernel weights according to the semantic characteristics of the frog image, accurately extracting key semantic features such as the frog's skin texture and body shape; the hierarchical semantic fusion unit integrates semantics at different levels, allowing the model to have a more accurate understanding of the overall characteristics of the frog, so that it can effectively recall samples while accurately classifying them, greatly improving the F1 value. In the "car" category, the ASPM model has an F1 value of 95.5%. Compared with other models, it can better cope with the complex features of cars at different angles and lighting conditions. Through effective semantic perception and fusion, it achieves high accuracy and high recall, and obtains a high F1 value, which fully demonstrates the advantages of the ASPM model in complex image classification tasks.

4.2.10 Comparison of F1 values of different models in each category (chest X-ray image dataset)

Table 5: F1 values of each model on the chest X-ray image dataset in different categories (disease status)

Model	normal	pneumonia	tuberculosis	Pulmonary nodules	Other diseases
LeNet-5	64.8	65.5	64.0	66.0	63.0
AlexNet	71.5	72.2	71.0	73.0	70.0
VGG-16	78.0	78.8	77.5	79.5	76.0
ASPM	84.5	85.2	84.0	86.0	83.0

Table 5 shows the F1 values of each model on the chest X-ray image dataset for different disease status categories. For the "pneumonia" category, the F1 value of LeNet-5 is 65.5%. Its simple network structure makes it difficult to accurately identify the subtle feature changes of pneumonia on X-ray images, resulting in low accuracy and recall rates, and a low F1 value. By increasing the complexity of the network, AlexNet increased its F1 value to 72.2%, which can capture some features of pneumonia to a certain extent, but there is still much room for improvement. With its deep network structure, VGG-16 has an F1 value of 78.8%, which can learn more key features of pneumonia and optimize recognition performance. The F1 value of the ASPM model in the "pneumonia" category is as high as 85.2%. The semantic-aware convolution module can accurately focus on the semantic key parts of the pneumonia lesion area and extract more valuable features; the hierarchical semantic fusion unit fuses the basic image features with the middle and high-level semantic structure information, so that the model has a more comprehensive understanding of the characteristics of pneumonia; the adaptive domain adjustment module effectively adjusts the model parameters according to the characteristics of the medical image field, so that it can effectively recall related cases while accurately diagnosing pneumonia, significantly improving the F1 value. In the "pulmonary nodule" category, the F1 value of the ASPM model is 86.0%. Compared with other models, it can better identify the shape, size and other features of pulmonary nodules. Through effective semantic perception and fusion mechanisms, it achieves high accuracy and high recall, showing obvious advantages in medical image analysis tasks.

All reported accuracy and F1 values represent the mean of five independent runs. For each result, we computed the standard deviation and 95% confidence interval using bootstrapped sampling over the test predictions. For instance, in the MNIST dataset, the ASPM model achieved an average accuracy of 99.8% ± 0.07% and an F1 score of 99.7% ± 0.05%. Similarly, for the CIFAR-10 dataset, the ASPM model yielded 95.5% ± 0.11% accuracy and 94.5% ± 0.13% F1 score. These statistics indicate that the performance gains are consistent and statistically significant compared with baseline models. We further conducted one-tailed paired t-tests between ASPM and each baseline model across categories, confirming p-values < 0.01 for most comparisons.

4.3 Ablation study

To validate the architectural effectiveness of ASPM, we conducted an ablation study where each core component—SSCM, HSFU, and ADAM—was independently removed and the resulting performance degradation was measured on CIFAR-10 and the chest X-ray datasets. The configurations tested were as follows:

ASPM w/o SSCM: replaced semantic-aware convolution with standard convolutional layers.

ASPM w/o HSFU: replaced hierarchical fusion with simple feature concatenation.

ASPM w/o ADAM: removed domain adaptation and trained without cross-domain adjustment.

Table 6: Ablation study results on CIFAR-10 and Chest X-ray datasets

Model Variant	CIFAR-10 F1 (%)	Chest X-ray F1 (%)
Full ASPM	94.5 ± 0.13	84.7 ± 0.14
w/o SSCM	91.8 ± 0.19	80.2 ± 0.21
w/o HSFU	90.3 ± 0.22	78.7 ± 0.25
w/o ADAM	92.5 ± 0.18	76.5 ± 0.30

The ablation results clearly indicate that each module contributes to overall performance. The removal of HSFU and SSCM both caused a decline in F1 scores across both datasets, while removing ADAM had the most severe effect in domain-sensitive tasks such as chest X-ray classification. This confirms that semantic integration and domain adaptability are critical to ASPM's superior performance.

5 Discussion

The experimental results demonstrate that the ASPM model consistently outperforms LeNet-5, AlexNet, and VGG-16 across all evaluated datasets. This section analyzes the reasons behind the observed performance advantages and highlights the architectural contributions that enable them.

Semantic Fusion and Accuracy Gains
In both MNIST and CIFAR-10 datasets, the hierarchical semantic fusion unit (HSFU) played a key role in improving the classification accuracy and F1 scores. Traditional CNNs treat feature extraction as a single-stage process, whereas ASPM captures and fuses semantic information across multiple abstraction layers. This multi-level semantic integration allows the model to preserve both low-level patterns and high-level contextual information, which is particularly beneficial for distinguishing visually similar classes such as “3” and “5” in MNIST or “cat” and “dog” in CIFAR-10.

Domain Adaptation and Medical Imaging Performance
The adaptive domain adjustment module (ADAM) significantly improves model generalization in the chest X-ray dataset, which contains subtle domain-specific patterns that are not present in natural image datasets. Existing SOTA models failed to adapt to this distribution shift, resulting in decreased accuracy and F1 scores. In contrast, ASPM uses domain discrepancy measurements (e.g., MMD) to guide real-time parameter adjustment, thereby reducing overfitting to source domains and increasing robustness in medical contexts.

Beyond Higher Metrics: Architectural Novelty
The superiority of ASPM lies not merely in its higher metrics, but in its architectural innovations. The model's semantic-aware convolution layers dynamically assign importance to different image regions, improving interpretability and focus. Its fusion and adaptation strategies work in synergy, producing consistent improvements in both general and specialized domains. Therefore, ASPM's enhancements are both quantitatively significant and theoretically justified, extending deep learning’s applicability to real-world, cross-domain image processing problems.

6 Conclusion

This study innovatively constructed the ASPM model to address the problems of existing deep learning image processing and pattern recognition algorithms, such as heavy reliance on labeled data, high consumption of computing resources, and poor adaptability. During the research, a variety of representative data sets were selected for comprehensive experiments and compared with classic models. From the results, the ASPM model has significant advantages. In the MNIST data set, with its unique semantic perception and fusion mechanism, it has a deep understanding of digital features, with an accuracy of up to 99.8%, far exceeding LeNet-5's 98.5%, AlexNet's 99.2%, and VGG-16's 99.4%. The F1 values of each digital category are all above 99.5%. In the CIFAR-10 data set, in the face of complex natural scene images, the ASPM model achieves an accuracy of 95.5% by dynamically adjusting the convolution kernel weights and effectively fusing semantic information. The F1 values of various categories such as airplanes and frogs are 3-4 percentage points higher than other models. In the professional field of chest X-ray image datasets, the ASPM model focuses on the semantics of the lesion area based on medical image features, and the F1 value for the pneumonia category reaches 85.2%, and the F1 value for the lung nodule category is 86.0%, far exceeding the control model. This shows that the ASPM model can effectively improve the algorithm performance in image tasks in different fields, greatly improve the algorithm adaptability, and lay a solid foundation for the further application of image processing and pattern recognition technology in multiple fields such as medicine and industry, which has important theoretical and practical significance.

Funding

This paper is supported by 1) The work of this paper was supported by Natural Science Foundation for Higher Education of Anhui Province of China (No. KJ2020A1095); 2) Academic Project for Top-notch Talents of Disciplines (Majors) for Higher Education of Anhui Province of China (No. gxbjZD2020105).

References

[1] Chen Y, Huang YQ, Zhang ZZ, Wang Z, Liu B, Liu CH, et al. Plant image recognition with deep learning: A review. Computers and Electronics in Agriculture. 2023; 212:17. DOI: 10.1016/j.compag.2023.108072

[2] Qin SF, Li LJ. Visual analysis of image processing in the mining field based on a knowledge map. Sustainability. 2023 Feb;15(3):1810. doi:10.3390/su15031810.

[3] Amiri Z, Heidari A, Navimipour NJ, Unal M, Mousavi A. Adventures in data analysis: a systematic review of Deep Learning techniques for pattern recognition in cyber-physical-social systems. Multimedia Tools and Applications. 2023:65. DOI: 10.1007/s11042-023-16382-x

[4] Vaiyapuri T, Mohanty SN, Sivaram M, Pustokhina IV, Pustokhin DA, Shankar K. Automatic Vehicle License Plate Recognition Using Optimal Deep Learning Model. Cmc-Computers Materials & Continua. 2021;67(2):1881-97. DOI: 10.32604/cmc.2021.014924

[5] Liu HQ, Hu BX, Cao Y. HDMA-CGAN: Advancing Image Style Transfer with Deep Learning. International Journal of Pattern Recognition and Artificial Intelligence. 2024;38(09):27. DOI: 10.1142/s0218001424520190

[6] Liu Z, Wang HJ, Zhou TH, Shen ZQ, Kang BY, Shelhamer E, et al. Exploring Simple and Transferable Recognition-Aware Image Processing. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2023;45(3):3032-46. DOI: 10.1109/tpami.2022.3183243

[7] Suganthi ST, Ayoobkhan MUA, Kumar VK, Bacanin N, Venkatachalam K, Stepán H, et al. Deep learning model for deep fake face recognition and detection. Peerj Computer Science. 2022; 8:20. DOI: 10.7717/peerj-cs.881

[8] Sun GM, Kuang B, Zhang YK. Fast Target Recognition Method Based on Multi-Scale Fusion and Deep Learning. Traitement Du Signal. 2022;39(6):2173-9. DOI: 10.18280/ts.390631

[9] Parashar A, Parashar A, Rida I. Journey into gait biometrics: Integrating deep learning for enhanced pattern recognition. Digital Signal Processing. 2024; 147:12. DOI: 10.1016/j.dsp.2024.104393

[10] Tasci E, Ugur A. A novel pattern recognition framework based on ensemble of handcrafted features on images. Multimedia Tools and Applications. 2022;81(21):30195-218. DOI: 10.1007/s11042-022-12909-w

[11] Souza LS, Sogi N, Gatto BB, Kobayashi T, Fukui K. Grassmannian learning mutual subspace method for image set recognition. Neurocomputing. 2023; 517:20-33. DOI: 10.1016/j.neucom.2022.10.040

[12] Lin YH, Ting YH, Huang YC, Cheng KL, Jong WR. Integration of Deep Learning for Automatic Recognition of 2D Engineering Drawings. Machines. 2023;11(8):20. DOI: 10.3390/machines11080802

[13] Vasconcelos RN, Rocha WJSF, Costa DP, Duverger SG, Santana MMMD, Cambui ECB, Ferreira-Ferreira J, Oliveira M, Barbosa LD, Cordeiro CL. Fire Detection with Deep Learning: A Comprehensive Review. Land. 2024 Oct;13(10):1696. doi: 10.3390/land13101696.

[14] Zhou SW, Deng XN, Li CQ, Liu YH, Jiang HB. Recognition-Oriented Image Compressive Sensing with Deep Learning. IEEE Transactions on Multimedia. 2023; 25:2022-32. DOI: 10.1109/tmm.2022.3142952

[15] Zhang Y, Cui ML, Shen LL, Zeng ZG. Memristive Fuzzy Deep Learning Systems. IEEE Transactions on Fuzzy Systems. 2021;29(8):2224-38. DOI: 10.1109/tfuzz.2020.2995966

[16] Ren ZX, Guo JF. Fault diagnosis using signal processing and deep learning-based image pattern recognition. TM-Technisches Messen. 2024;91(2):129-38. DOI: 10.1515/teme-2023-0089

[17] Huang LQ, Yao C, Zhang LY, Luo SJ, Ying FT, Ying WQ. Enhancing computer image recognition with improved image algorithms. Scientific Reports. 2024;14(1):11. DOI: 10.1038/s41598-024-64193-3

[18] Liu XB, Song LP, Liu S, Zhang YD. A review of deep-learning-based medical image segmentation methods. Sustainability. 2021 Feb;13(3):1224. doi:10.3390/su13031224.

[19] Parashar A, Parashar A, Ding WP, Shekhawat RS, Rida I. Deep learning pipelines for recognition of gait biometrics with covariates: a comprehensive review. Artificial Intelligence Review. 2023;56(8):8889-953. DOI: 10.1007/s10462-022-10365-4

[20] Li QL, Liu J, Sun YF, Zhang CS, Dou DJ. On mask-based image set desensitization with recognition support. Applied Intelligence. 2024;54(1):886-98. DOI: 10.1007/s10489-023-05239-3

[21] Pei JF, Huo WB, Wang CW, Huang YL, Zhang Y, Wu JJ, et al. Multiview Deep Feature Learning Network for SAR Automatic Target Recognition. Remote Sensing. 2021;13(8):21. DOI: 10.3390/rs13081455

[22] de Oliveira CI, do Nascimento MZ, Roberto GF, Tosta TAA, Martins AS, Neves LA. Hybrid models for classifying histological images: An association of deep features by transfer learning with ensemble classifier. Multimedia Tools and Applications. 2023:24. DOI: 10.1007/s11042-023-16351-4

[23] *John Paul Tan Yusiong, Prospero Clara Naval.* A semi-supervised approach to monocular depth estimation, depth refinement, and semantic segmentation of driving scenes using a siamese triple

decoder architecture. Informatica (Slovenia), Volume 44, Issue 4, Pages 437-445, December 2020. DOI: https://doi.org/10.31449/inf.v44i4.3018

[24] Wala'a Nsaif Jasim, Zainab Najem Nemer, Esra'a Jasem Harfash. Implementation of Multiple CNN Architectures to Classify the Sea Coral Images. Informatica (Slovenia), Volume 47, Issue 1, Pages 43-50, 2023. DOI: https://doi.org/10.31449/inf.v47i1.4429