WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 8, Number 1, Year 2026
DOI: https://doi.org/10.62051/ijcsit.v8n1.04
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Research on Sensitive Information Recognition Algorithm Based on Deep Neural Network

Shuaina Huang [1, *], Karpovich Dmitry Semyonovich [2]

[1] College of Mechanical and Electrical Engineering, Luoyang Polytechnic, Luoyang, China
[2] College of Electronics Technology, Belarusian State Technological University, Belarusian
*Corresponding Author: Shuaina Huang

ABSTRACT

The spread of pornographic, violent, and politically sensitive images on social networks poses serious risks to youth well-being and social stability, making accurate detection essential for public safety. However, existing models often compromise representational capacity to meet computational constraints, while challenges such as illumination variation, changes in sensitive-region scale, and background interference further limit recognition accuracy. To address these issues, this study proposes a lightweight sensitive-image recognition framework. A high-order convolutional module is introduced to extract fine-grained features and suppress irrelevant background noise. A feature-guided fusion module is further designed to integrate texture and deep semantic features, improving robustness against lighting fluctuations and noise. The overall architecture builds on a compact high-resolution network enhanced with depthwise separable convolutions, MBConv layers, and Ghost Modules to significantly reduce parameters while maintaining strong performance. Experiments on a multi-class sensitive-image dataset verify the model's efficiency and effectiveness.

KEYWORDS

Sensitivity information recognition; Feature fusion; Deep neural network

1. INTRODUCTION

In recent years, the rapid development of social networks has led to a significant increase in pornographic, violent, and other sensitive content. Such information poses serious threats to the physical and mental well-being of adolescents, undermines social stability, and creates substantial regulatory pressure. Traditional sensitive-image recognition methods largely rely on handcrafted features. These approaches construct local or global descriptors to extract spatiotemporal patterns and then classify behaviors using support vector machines or random forests after removing redundant information. However, these methods are unable to represent complex high-level features, exhibit low recognition accuracy, and require extensive preprocessing. Moreover, they tend to be highly specialized, capable of recognizing only specific categories or features with limited generalization ability, and are easily affected by noise and background clutter—making real-time online detection particularly challenging. Achieving efficient and stable recognition of multi-category sensitive images has therefore become an urgent issue.

By contrast, deep-learning-based methods can automatically learn high-level feature representations and perform more accurate and effective classification of sensitive images. Conventional deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), adopt various optimization strategies—including architectural improvements and temporal feature fusion—to enhance recognition performance and robustness. Although many high-performing

models have emerged in recent years, such as MobileNet, ShuffleNet, SqueezeNet, and EfficientNet, applying them directly to sensitive-image recognition still presents challenges, and their effective integration with sensitive-content scenarios remains underexplored. Real-time sensitive-content detection further requires improved utilization of sensitive features while suppressing background interference, placing higher demands on model representational capacity and inference speed.

This chapter aims to explore new methods for balancing inference efficiency and recognition accuracy in sensitive-image detection. We propose a CNN-based sensitive-content recognition algorithm that incorporates HO deep convolution and a multi-scale attention mechanism. By weighting and integrating deep and shallow features, the model applies attention across both channel and spatial dimensions, enabling prioritized processing of sensitive patterns and reducing the impact of noise and irrelevant backgrounds. The HO deep convolution module performs fine-grained fusion of texture and semantic features, effectively mitigating the negative effects of illumination variation and noise. Furthermore, depthwise separable convolutions and Ghost Modules are introduced into the overall architecture to reduce parameters and computational cost. The proposed lightweight end-to-end deep learning framework enhances the feature extraction capability of CNNs, obtains richer representations even at lower network depths, and demonstrates strong trainability—achieving a favorable balance between accuracy and model compactness.

2. RELATED WORKS

Sensitive image recognition is a subfield of image analysis that focuses on identifying specific categories of sensitive visual content. It relies heavily on machine learning and, more recently, on rapidly evolving artificial intelligence technologies, making it an important interdisciplinary research area. Sensitive visual features act as a crucial bridge between data and models, and appropriate feature representations enable more accurate and efficient model construction. Early research on pornography-related image detection primarily relied on features such as skin color, texture, and shape [1], combined with threshold-based classification. More complex features were later introduced, including the number of skin regions, the area of the largest skin region, and the count of connected components [2]. Jiao et al. [3] used skin-color models to detect exposed regions and applied Sobel operators and Gabor filters to suppress non-exposed areas. Jones and Rehg [4] estimated the distribution of skin and non-skin regions in color space using labeled training data, while Jedynak et al. [5] constructed a skin detection model based on pre-labeled images and maximum entropy principles. Lee et al. [6] proposed a pornography detection method based on skin-tone distribution and texture features to determine whether an image contains pornographic content.

Convolutional neural networks (CNNs) have achieved remarkable success in image processing due to their ability to automatically extract abstract features from raw pixels [7–8], significantly improving tasks such as image classification, localization, geometric transformation, and semantic segmentation. Deep convolutional neural network (DCNN)-based methods perform exceptionally well in image classification and object detection, motivating researchers to adopt DCNNs for recognizing pornography, violence, and other sensitive images [9]. Moustafa et al. [10] proposed AGNet, a DCNN architecture combining AlexNet and GoogleNet to extract discriminative global features. Their results showed that DCNN-derived features outperform most handcrafted descriptors. Ou et al. [11] argued that global features alone are insufficient and that local information—such as sensitive organs or private body parts—can significantly improve detection accuracy. They proposed DMCNet, which extracts global features using deep convolutional layers, detects local regions with Faster R-CNN, and performs weighted fusion of global, local, and concatenated features through a voting scheme. Although DMCNet integrates global and local information, its separate training processes limit the exploitation of their intrinsic correlations and increase computational cost. Integrating local and global contexts is highly beneficial for improving recognition performance in many applications [12]. Cheng et al. [13] proposed a DCNN-based adult image classification method

that categorizes images into pornographic, erotic, and benign. Their model incorporates both global and local cues: the global network includes hierarchical residual blocks capturing foreground–background relationships, while the local network detects and extracts features from specific sensitive body parts. The final decision fuses both sources of information. Wang et al. [14] introduced LocoaNet, a unified end-to-end multitask network that learns discriminative features from global and local contexts, using local information to guide attention toward sensitive regions rather than irrelevant backgrounds. Although LocoaNet achieved high accuracy, its local contextual information was not fully exploited, indicating room for improvement.

With the rapid advancement of deep neural networks, a growing number of AI algorithms are being applied to image recognition tasks. CNNs are particularly advantageous due to their automatic feature extraction capabilities, reducing the risk of feature selection errors. However, as model parameters increase, training and inference become more computationally expensive, limiting real-world applicability. This has led to rising interest in lightweight CNNs, which aim to reduce parameter counts and simplify architectures to meet the constraints of devices with limited memory and processing power, such as surveillance systems and UAVs. EfficientNet, for example, reduces network depth and resolution while achieving a strong balance between accuracy, computational cost, and model size. Its neural architecture, built via the AutoML MNAS framework, automatically searches for optimal width and resolution coefficients to maximize performance. How to effectively apply lightweight networks to sensitive-content recognition—and how to balance accuracy and efficiency—remains an important and unresolved challenge.

3. PROPOSED METHOD

To enhance the performance of sensitive-content recognition tasks, this study introduces a lightweight sensitive-image recognition network that integrates a High-Order (HO) convolution module with a multi-scale attention feature architecture. The network is designed to more effectively capture multi-scale information and improve the quality of feature representation.

In the proposed approach, the HO convolution module is embedded with depthwise separable convolutions and MBConv blocks, which jointly boost recognition accuracy while significantly reducing the number of parameters, thus achieving a balanced trade-off between precision and computational efficiency. Depthwise separable convolution decomposes the standard convolution into two operations: a depthwise convolution that applies a single-channel filter to each input channel, and a pointwise convolution that uses a 1×1 kernel to fuse information across channels.

The MBConv block consists of a 1×1 expansion convolution, a depthwise separable convolution, an SE (Squeeze-and-Excitation) module, a 1×1 projection convolution for dimensionality reduction, and a dropout layer. These components collaboratively enrich feature extraction, strengthen representation capability, and maintain a lightweight architecture suitable for real-time sensitive image recognition.

The multi-scale attention feature structure consists of three parallel branches, each operating on feature maps of different spatial resolutions. Channel attention is applied within every branch. The channel attention mechanism adaptively adjusts the weight of each feature channel and explicitly models inter-channel dependencies. Global average pooling is employed to reduce the dimensionality of the input feature map $U = G(y_{n+1}) \in R^{H \times W \times C}$, transforming its dimensions from H×W×C to 1×1×C. This operation aggregates global contextual information and provides a compact descriptor for subsequent attention weighting. Its mathematical expression is as follows:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j)$$

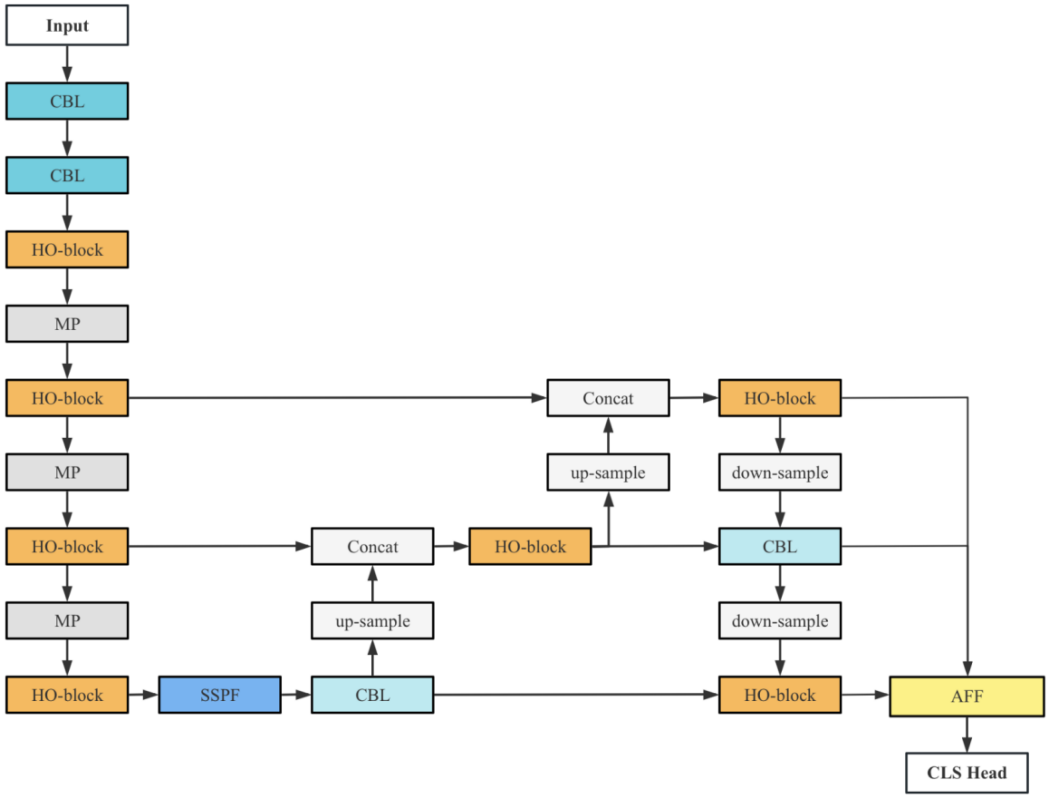

Figure 1. Structure diagram of sensitive information recognition algorithm

Here, z_c represents the aggregated information across the entire channel, while $u_c(i,j)$ corresponds to the value at position (i,j) within channel c. This attention mechanism converts each two-dimensional feature map into a single vector, compressing features along the spatial dimension while keeping the number of channels unchanged. The process functions similarly to a pooling operation with a global receptive field. The channel weights are determined using two fully connected layers followed by activation functions. After the first fully connected layer, a ReLU activation is applied, producing an intermediate output of dimension C/r. Finally, a five-layer convolutional network combined with a sigmoid activation function is used to perform sensitive-content recognition.

4. EXPERIMENTAL RESULT AND DISCUSSION

To validate the performance of the proposed intelligent perception system, experiments were conducted using datasets containing pornographic, violent, and normal images. The experiments were performed on a workstation equipped with an RTX 4090 GPU running the Windows operating system, and the model was implemented using the PyTorch framework.

Three representative baseline methods—VGGNet, ResNeXt, and DenseNet—were selected for comparison. Evaluation metrics included accuracy, recall, F1-score, and the number of parameters. Experimental results demonstrate that the proposed model significantly outperforms traditional approaches in terms of accuracy, recall, and energy efficiency.

Table 1. Performance Comparison of Recognition Results

Method	Accuracy	Precision	F1-score	Parameters
VggNet	91.6	92.0	92.1	134.2M
Resnext	93.5	93.8	93.5	22.9M
DenseNet	93.0	93.2	93.0	6.96M
Our method	94.8	94.7	94.1	4.52M

5. SUMMARY

This study investigates a deep neural network–based approach for sensitive-content recognition and proposes an innovative architecture that integrates a High-Order (HO) neural network with a multi-scale attention mechanism. A deep learning model is constructed accordingly, and its effectiveness is validated through extensive experiments. The results show that the proposed algorithm not only significantly improves recognition accuracy but also reduces the number of parameters, demonstrating strong potential for practical deployment. Future research may further advance this field by exploring multimodal data fusion, collaborative decision-making mechanisms, and edge–cloud collaborative computing, thereby accelerating the real-world implementation and industrialization of intelligent sensitive-content recognition technologies.

ACKNOWLEDGEMENTS

The work is supported by Henan Province Science and Technology Research Project NO.242102210122

REFERENCES

[1] Mishra D, Panda S. A Comparative Analysis of Pornography Detection Models to Prevent Gender Violence [M]//Communication technology and gender violence. Cham: Springer International Publishing, 2023: 99-107.

[2] Garcia M B, Revano T F, Habal B G M, et al. A pornographic image and video filtering application using optimized nudity recognition and detection algorithm [C]//2018 IEEE 10th international conference on humanoid, nanotechnology, information technology, communication and control, environment and management (HNICEM). IEEE, 2018: 1-5.

[3] jiao F, Gao W, Duan L, et al. Detecting adult image using multiple features [C]//2001 International conferences on info-tech and info-net. proceedings (Cat. No. 01EX479). IEEE, 2001: 378-383.

[4] Jones M J, Rehg J M. Statistical color models with application to skin detection [J]. International journal of computer vision, 2002, 46(1): 81-96.

[5] Jedynak B, Zheng H, Daoudi M. Statistical models for skin detection [C]//2003 Conference on computer vision and pattern recognition workshop. IEEE, 2003: 92-92.

[6] Lee J S, Kuo Y M, Chung P C, et al. Naked image detection based on adaptive and extensible skin color model [J]. Pattern recognition, 2007, 40(8): 2261-2270.

[7] Chen L C, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs [J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(4): 834-848.

[8] Liu H, Guo R Y. Detection and identification of SAWH pipe weld defects based on X-ray image and CNN [J]. Chinese journal of scientific instrument, 2018, 39(4): 247-256.

[9] Wang Y H, Wang J, Tan X. Pornographic image recognition by strongly-supervised deep multiple instance learning [C]//2016 IEEE international conference on image processing (ICIP). IEEE, 2016: 4418-4422.

[10] Moustafa M. Applying deep learning to classify pornographic images and videos [J]. arxiv preprint arxiv:1511.08899, 2015.

[11] Ou X, Ling H, Yu H, et al. Adult image and video recognition by a deep multicontext network and fine-to-coarse strategy [J]. ACM transactions on intelligent systems and technology (TIST), 2017, 8(5): 1-25.

[12] Xiang T Z, Xia G S, Bai X, et al. Image stitching by line-guided local warping with global similarity constraint [J]. Pattern recognit, 2018, 77: 113-125.

[13] Cheng F, Wang S L, Wang X Z, et al. A global and local context integration DCNN for adult image classification [J]. Pattern recognition, 2019, 96: 106983.

[14] Wang X, Cheng F, Wang S, et al. Adult image classification by a local-context aware network [C]//2018 25th IEEE international conference on image processing (ICIP). IEEE, 2018: 2989-2993.