International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 40, No. 5, October, 2023, pp. 2251-2257
Journal homepage: http://iieta.org/journals/ts

Real-Time Recognition and Feature Extraction of Stratum Images Based on Deep Learning

Tong Wang[1,2*], Yu Yan[2], Lizhi Yuan[3], Yanhong Dong[2]

[1] School of Prospecting & Surveying Engineering, Changchun Institute of Technology, Changchun 130061, China
[2] China Northeast Municipal Engineering Design and Research Institute Co. Ltd., Changchun 130021, China
[3] Jilin Province Water Conservancy and Hydroelectric Engineering Bureau Grovpco, Co. Ltd., Changchun 130021, China

Corresponding Author Email: wangtong@ccit.edu.cn

https://doi.org/10.18280/ts.400542

Received: 3 May 2023
Revised: 12 August 2023
Accepted: 26 August 2023
Available online: 30 October 2023

Keywords:
stratum image, deep learning, pyramid model, feature extraction, real-time recognition, classification-regression network, anchor supervision

ABSTRACT

Accurate identification and feature extraction of stratum images play a crucial role in geological exploration, resource prospecting, and mining operations. Traditional methods of stratum image identification largely rely on human experience and manual operations, which are inefficient and prone to errors. In recent years, deep learning technology has provided new methods for the identification and feature extraction of stratum images, but existing deep learning models still face challenges in computational efficiency, multi-scale feature extraction, and uneven sample distribution. This paper proposes a stratum image feature extraction network based on the pyramid model and constructs a lightweight stratum identification model for real-time recognition. By introducing a classification-regression network structure and anchor-based sample supervision rules, this study aims to improve the accuracy and efficiency of the model, providing an effective solution for real-time recognition of stratum images.

1. INTRODUCTION

With the development of industry and science and technology, the role of stratum image recognition and feature extraction in geological exploration, resource prospecting, and mining has become increasingly prominent [1-4]. Especially in the extraction of oil and natural gas, accurately identifying and predicting the distribution and properties of strata are crucial for optimizing resource extraction strategies and improving safety [5, 6]. Traditional methods for stratum image recognition largely depend on human experience and inefficient manual operations, limiting work efficiency and potentially leading to high error rates [7-11].

In recent years, the rapid rise of deep learning technology has brought revolutionary changes to the field of image recognition and analysis, including the feature extraction and identification of stratum images [12, 13]. Using deep learning technology, key features can be automatically extracted from complex stratum images, realizing more accurate and efficient stratum classification and prediction. This not only helps improve resource utilization but also provides more scientific and reliable decision support for geological exploration and extraction activities [14, 15].

Although deep learning has achieved preliminary applications in stratum image recognition, existing methods still have some shortcomings. First, most models have a large number of parameters, leading to low computational efficiency, making them unsuitable for real-time applications. Secondly, traditional feature extraction networks find it challenging to effectively capture multi-scale information from images. Especially for pyramid-structured stratum images, high-level semantic information and low-level detail features are often difficult to consider simultaneously [16-18]. Additionally, the uneven distribution of samples in object detection also poses challenges to model training [19-23].

In response to these issues, this study proposes a stratum image feature extraction network based on the pyramid model. This network effectively integrates high-level semantic features with low-level feature maps, achieving comprehensive feature capture of pyramid-structured stratum images. Furthermore, this study constructs a lightweight stratum identification model for real-time recognition. On this basis, a classification-regression network structure is introduced, and an anchor-based sample supervision rule is proposed, ensuring high accuracy recognition even when facing uneven sample distribution. This research not only provides new methods and insights for real-time recognition of stratum images but also has significant application value and broad research prospects.

2. CONSTRUCTION OF STRATUM IMAGE FEATURE EXTRACTION NETWORK BASED ON PYRAMID MODEL

For the recognition and feature extraction of stratum images, the information they contain is both rich and complex, especially the structural information and relative positional information brought by their stratified structure. These pieces of information are crucial for accurately identifying the nature and types of strata. However, traditional fixed-scale feature extraction networks may be affected by the contextual relationship between local feature blocks when processing stratum images, leading to key information in the stratified structure of strata being disrupted by horizontal segmentation, thereby affecting the recognition performance of the model. Figure 1 shows the architecture of the traditional fixed-scale feature extraction network model. This paper introduces the pyramid model to attempt multi-scale feature extraction, avoiding the damage to this structural information caused by horizontal segmentation in traditional networks. At the same time, it ensures that the model can capture features of various scales, from large to small and from the whole to the local. This is crucial for enhancing the model's discriminative capability and generalization ability.

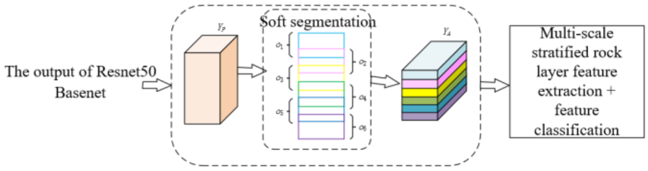
The output of Resnet50 Basenet
Soft segmentation
Y_F
Y_A
Multi-scale stratified rock layer feature extraction + feature classification

Figure 1. Architecture of traditional fixed-scale feature extraction network model

The multi-scale feature extraction network model constructed in this paper adopts a 6-layer pyramid structure. The core idea of this structure is to extract and integrate local features from different scales to ensure that key information can be captured at all scales. The input to the network model is the output tensor from the fixed-scale feature extraction network. This means that preliminary feature extraction has been completed, and the focus of the subsequent work is on further integrating and optimizing these features. These fixed-scale local feature tensors originate from some soft segmentation technique. Soft segmentation typically better preserves some of the key characteristics and contextual information of the original image. At each level of the pyramid, the horizontal local feature tensors are combined in different ways. In this manner, each layer can generate local features of a different scale.

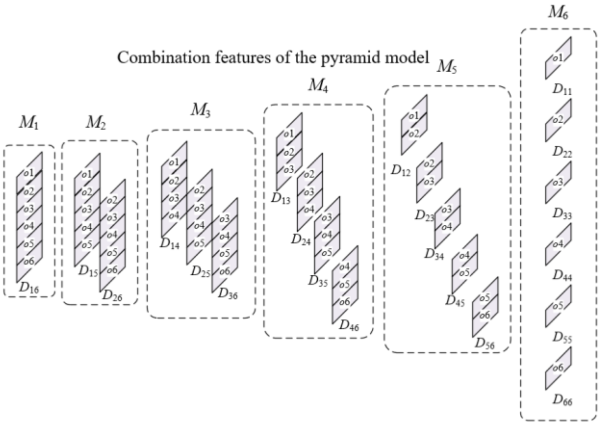

Figure 2. Features of the pyramid model

Figure 2 gives a schematic of the features of the pyramid model. Suppose tensor Y_A is the feature map D_{11} of the top layer M_1 of the pyramid model. The second layer M_2 of the pyramid model produces two feature maps: D_{21}, which consists of horizontal local feature tensors $o_1o_2o_3o_4o_5$, and D_{22}, which consists of $o_2o_3o_4o_5o_6$. The third layer M_3 of the pyramid model produces three feature maps: D_{31} formed by $o_1o_2o_3o_4$, D_{32} formed by $o_2o_3o_4o_5$, and D_{33} formed by $o_3o_4o_5o_6$. The fourth layer M_4 of the pyramid model produces four feature maps: D_{41} consisting of $o_1o_2o_3$, D_{42} consisting of $o_2o_3o_4$, D_{43} consisting of $o_3o_4o_5$, and D_{44} consisting of $o_4o_5o_6$. The fifth layer M_5 of the pyramid model produces five feature maps: FD_{51} formed by o_1o_2, D_{52} formed by o_2o_3, D_{53} formed by o_3o_4, D_{54} formed by o_4o_5, and D_{55} formed by o_5o_6. The sixth layer M_6 of the pyramid model produces six feature maps: D_{61} consisting of o_1, D_{62} consisting of o_2, D_{63} consisting of o_3, D_{64} consisting of o_4, D_{65} consisting of o_5, and D_{66} consisting of o_6. The following equations give the expression for the pyramid model:

$$PY=\begin{cases} M_1=[D_{11}]=[o_1o_2o_3o_4o_5o_6] \\ M_2=[D_{21},D_{22}]=[o_1o_2o_3o_4o_5,o_2o_3o_4o_5ro_6] \\ M_3=[D_{31},D_{32},D_{33}] \\ =[o_1o_2o_3o_4,o_2o_3o_4o_5,o_3o_4o_5o_6] \\ M_4=[D_{41},D_{42},D_{43},D_{44}] \\ =[o_1o_2o_3,o_2o_3o_4,o_3o_4o_5,o_4o_5o_6] \\ M_5=[D_{51},D_{52},D_{53},D_{54},D_{55}] \\ =[o_1o_2,o_2o_3,o_3o_4,o_4o_5,o_5o_6] \\ M_6=[D_{61},D_{62},D_{63},D_{64},D_{65},D_{66}] \\ =[o_1,o_2,o_3,o_4,o_5,o_6] \end{cases} \tag{1}$$

Given the 6-layer pyramid model produces 6 different scales of local feature maps, these feature maps might vary in size and characteristics. Adaptive pooling allows for dynamic pooling according to the specifics of each feature map, ensuring key information from each scale is effectively retained, thereby enhancing the robustness of the feature descriptor. Let the function *FL* denote rounding down, the following equations give the parameter settings in adaptive pooling:

$$\begin{cases} ST=FL(IP\_SI/OP\_SI) \\ K\_S=I\_S-(O\_S-1)\times ST \\ PA=0 \end{cases} \tag{2}$$

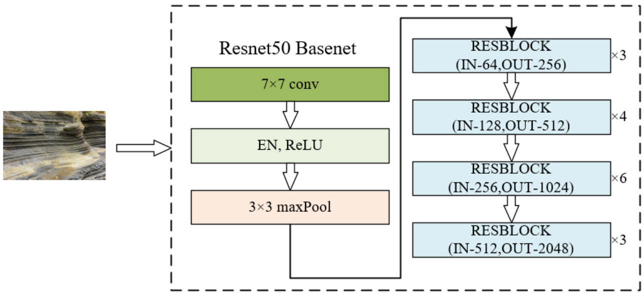
Resnet50 Basenet
7×7 conv
EN, ReLU
3×3 maxPool
RESBLOCK (IN-64,OUT-256)
×3
RESBLOCK (IN-128,OUT-512)
×4
RESBLOCK (IN-256,OUT-1024)
×6
RESBLOCK (IN-512,OUT-2048)
×3

Figure 3. The adopted *ResNet50* network architecture

Figure 3 illustrates the *ResNet50* network architecture adopted in this study. The original *ResNet50*, as a base network, outputs a feature map with dimensions of 1×1×2048. This is a relatively high dimension, and directly using such features for further processing and learning might complicate the model and increase computational burdens. In this paper, during the rock layer image feature classification phase, a classifier composed of a fully connected layer and a *Softmax* function were primarily employed. One of the main advantages of the *Softmax* function is its ability to transform input features into a probability distribution, meaning that the model can predict not only which class of rock layer the image belongs to but also provide the confidence level of that prediction. This is instrumental in interpreting the model's prediction and offering additional decision-making bases.

Following this phase, we can obtain the probability distribution of input features belonging to different rock layers. Subsequently, the *Identity* function was used to finalize the feature classification for the input image. The *Identity* function ensures that the output feature description from the model retains its continuity, which might be beneficial for feature interpretability and subsequent processing.

Assuming the convolution operation is represented by $\otimes$, the parameters in the convolution layer are represented by ϕ_I, the feature vector is represented by d, the predicted value for the input image u is represented by $\hat{o}_u$, the real probability is represented by o_u, the image label is represented by y, the following equations depict the feature classification process:

$$\hat{o} = \mathrm{softmax}\left(\phi_U \otimes d\right) \tag{3}$$

$$ID\left(d, y, \phi_U\right) = \sum_{u=1}^{J} -o_u \log\left(\hat{o}_u\right) \tag{4}$$

3. CONSTRUCTION OF THE REAL-TIME LIGHTWEIGHT ROCK LAYER RECOGNITION MODEL

Real-time rock layer image recognition models are not only crucial for scientific research and industrial applications but also have immense potential value and necessity in many practical working conditions. They can bring us higher efficiency, safety, and economic benefits. For example, in the exploration of oil and natural gas, real-time rock layer recognition can quickly determine whether the current drilling depth has reached the target layer or is nearing a potential oil and gas reservoir. This can not only reduce unnecessary drilling time and costs but also minimize potential damage to the subterranean environment. In situations like tunnel excavation or foundational construction, real-time rock layer recognition can provide immediate information about the underground materials. This aids in choosing the appropriate excavation strategy and preventing potential risks, such as cemented rock and flowing sand layers.

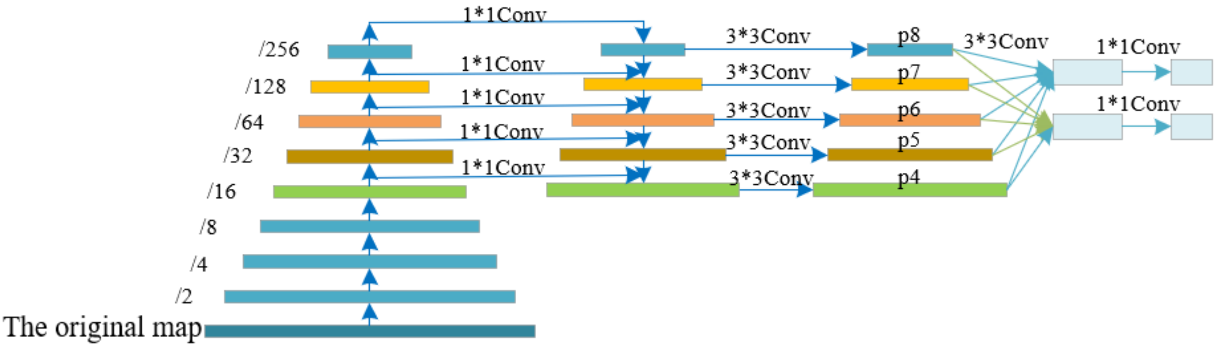

Figure 4. Improved *SSD* classification regression network structure

The lightweight rock layer recognition model for real-time recognition tasks built in this paper integrates the high-level semantic hierarchical features of the rock layer into the low-level feature map, using the fused feature map for lightweight rock layer recognition. High-level semantic features typically contain global and contextual information about the image, while low-level feature maps focus more on the details of the image. Integrating the two ensures that the model can not only capture the overall structure of the rock layer during real-time recognition but also accurately recognize specific rock layer details. Furthermore, this paper introduces the *SSD* (Single Shot multi-box Detector) classification regression network and proposes an anchor-based sample supervision rule in response to the problem of the sample supervision method in the network model suppressing real object features. *SSD* is an efficient network structure that performs excellently in object detection tasks. Introducing *SSD* helps the model to more accurately locate and recognize various parts of the rock layer, thereby enhancing the accuracy of real-time recognition. Figure 4 shows the improved *SSD* classification regression network structure.

The adopted *SSD* classification regression network employs a 3×3 convolution kernel for convolution operations, ensuring that the size of the feature map remains unchanged, while extracting deep features of the input image. Such deep feature extraction is crucial for rock layer recognition, as the complexity and diversity of rock layers require the model to have robust feature extraction capabilities. The feature map after deep feature extraction is connected to two 1x1 convolutional layers, namely the classification layer and the position regression layer. The classification layer is primarily responsible for predicting whether each anchor box contains a rock layer and its type. The position regression layer mainly predicts the specific position of the rock layer in the image, such as the coordinates of the bounding box. Each unit on the feature map corresponds to an anchor point on the original image, and these anchor points are mapped based on the height and width of the feature map. Each anchor point will have multiple anchor boxes, with the centers of these boxes being the corresponding anchor points. Anchor boxes are responsible for predicting rock layers of different shapes and sizes, ensuring that the model can recognize rock layers of various sizes and forms. Let the feature map of the *u*-th layer produced when mapping the feature map unit to the pixel of the original image anchor point be represented by $O_u \in E^{G\times Q\times V}$. The total span from the input original image to this layer is represented by a. Mapping each unit (z, t) of O_u back to the pixel of the input original image anchor point can be obtained through the following formula:

$$\left(z_p, t_p\right) = \left(\left[\frac{A}{2}\right] + za, \left[\frac{A}{2}\right] + ta\right) \tag{5}$$

The anchor-based sample supervision rule proposed in this paper focuses on whether the anchor points fall within the real object box. This allows the model to more accurately determine whether a given area truly contains the target (rock layers in the context of this paper). This helps to reduce false positives (incorrectly identifying non-target areas as the target) and false negatives (failing to identify genuine target areas). By judging the relative position of the anchor points to the real object box, the model's spatial localization ability is enhanced. This is especially important in rock layer image recognition, as the structure and hierarchical information of the rock layers play a crucial role in the accuracy of the recognition results. Figure 5 shows a schematic diagram of the anchor-based sample supervision rule.

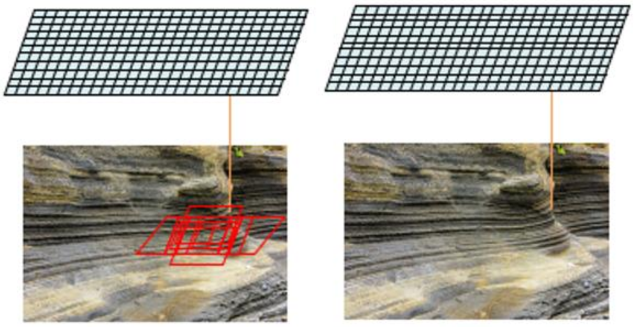

Figure 5. Schematic diagram of the anchor-based sample supervision rule

The feature fusion network outputs multi-level feature maps, each of which is followed by the *SSD* classification and regression network. Given that each unit of the feature map corresponds to multiple anchor boxes, a large number of candidate regions will be produced during the model prediction phase. This setup can increase the diversity of detection, improving the detection accuracy of objects of different sizes, shapes, and positions. However, this can also lead to a lot of overlapping and redundant detection boxes, which might detect the same object multiple times. Hence, this paper introduces non-maximum suppression (*NMS*) to help select the detection box with the highest confidence while suppressing other highly overlapping boxes, enhancing detection accuracy and reducing false positives.

Here are the specific steps for *NMS*:

(1) Sort all candidate regions in descending order based on class confidence scores. This means the highest-scoring candidate region is considered first.

(2) From the sorted list, select the candidate region with the highest confidence score and mark it as "retained". For the marked candidate region, calculate its overlap with all other unmarked candidate regions in the list. This is typically done by calculating the Intersection over Union (*IoU*) between the two regions. Assuming the candidate region's area is represented by A_o and the real object box's area by A_y, then the formula for calculating the overlap is:

$$IoU = \frac{A_o \cap A_y}{A_o \cup A_y} \tag{6}$$

(3) If an unmarked candidate region overlaps with the currently marked region beyond a predetermined threshold, discard the unmarked region since it likely represents the same object as the marked region. Return to step 2, select the highest-scoring unmarked candidate region from the list and mark it.

(4) Repeat the above step, calculating the overlap of this region with all other unmarked regions, and discard those that exceed the threshold. When all candidate regions have been marked, the algorithm stops. In the end, only the candidate regions marked as "retained" will be considered as the detection result.

When using a target detection framework like *SSD*, anchor boxes are typically used as references to predict object positions. These anchor boxes provide an initial, fixed bounding box for object detection, uniformly distributed across the input image at different scales and aspect ratios. However, the actual position of the object might slightly deviate from the anchor box. Therefore, these anchor boxes need some fine-tuning to more accurately cover the target object. Assuming the predicted anchor box position correction parameter is represented by y^*, and the anchor box's positional correction parameter relative to the real object box is also represented by y, the box's center coordinates and its width and height are represented by z, t, q, and g respectively. The candidate region is denoted by variables z_o, z, and z^*. The formulas for calculating y and y^* are:

$$y_z = \log\frac{(z_o - z)}{q}, y_t = \log\frac{(t_o - t)}{g} \tag{7}$$

$$y_q = \log\frac{q_o}{q}, y_g = \log\frac{g_o}{g} \tag{8}$$

$$y_z^* = \log\frac{(z^* - z)}{q}, y_t^* = \log\frac{(t^* - t)}{g} \tag{9}$$

$$y_q^* = \log\frac{q^*}{q}, y_g^* = \log\frac{g^*}{g} \tag{10}$$

By fine-tuning the anchor boxes, the model can more accurately capture the position and shape of the target object, thereby enhancing the model's detection precision. Output offset and scale parameters are easier to learn compared with the object's absolute coordinates, and this is because the model only needs to focus on how to fine-tune a given anchor box, rather than predicting an entirely new bounding box from scratch. Training the position regression layer is essentially the process of making y approach y^*.

4. EXPERIMENTAL RESULTS AND ANALYSIS

This paper introduces a stratum image feature extraction network based on the pyramid model. The network effectively integrates high-level semantic features with low-level feature maps, realizing comprehensive feature capture for pyramid-structured stratum images. As seen from Figure 6, after an initial decline over some time, the training loss (*trainLoss*) tends to stabilize, indicating that the model has achieved some convergence on the training set. The validation loss (*valLoss*) also stabilizes after its initial drop and remains within a range similar to the training loss. This suggests that the model doesn't exhibit significant overfitting, meaning the degree to which the model fits the training data is consistent with its performance on validation data. In the initial few epochs, both the training and validation losses drop rapidly, implying that the model has learned many vital features about the data at the beginning stages. Subsequently, the rate of decline in the loss slows down but continues to decrease, indicating that the model is still learning, albeit at a reduced pace. In the subsequent epochs, especially after 30 epochs, both training and validation losses become relatively stable, suggesting that the model might have approached its optimal performance.

In conclusion, the stratum image feature extraction network based on the pyramid model exhibits good convergence and stability without apparent overfitting. This possibly indicates that the proposed network structure and method are effective in the task of stratum image feature extraction.

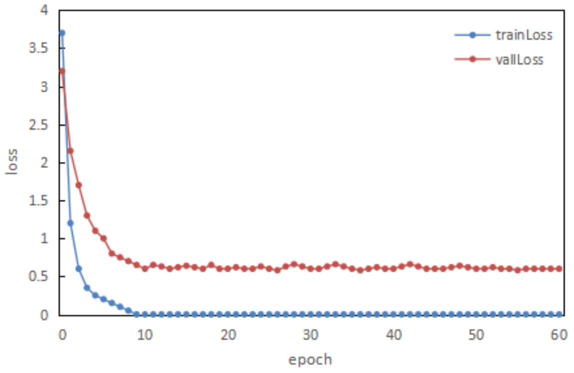

Figure 6. Loss curves of the stratum image feature extraction network

Table 1. Comparison of the proposed model and other models

Method	Base Network	Loss Function	*Rank*-1	*mAP*
Triplet Network	*CNN*	Triplet loss	75.6%	-
Siamese Networks	*DenseNet*121	Comparison loss	82.1%	62.3%
Autoencoders	*DenseNet*121	Reconstruction loss	88.31%	71.25%
Region-based CNN	*GoogleNet*	Classification loss	86.36%	72.13%
Attention Mechanism	*GoogleNet*	Classification loss	88.9%	72.56%
Transfer Learning	*ResNet*50	Classification loss	92.4%	78.56%
GAN	*ResNet*50	Classification loss	91.23%	76.23%
Feature Pyramid Networks	*ResNet*50	Classification loss	93.21%	76.42%
The proposed model	*ResNet*50	Classification loss	92.35%	78.94%

From Table 1, it is evident that the method using *ResNet50* as the base network achieved favorable results on both the *Rank-1* and *mAP* (Mean Average Precision) evaluation metrics, with the *Rank-1* accuracy of the three methods exceeding 90%. *DenseNet121* and *GoogleNet*, when used as base networks, also performed well, but overall were not as effective as the methods using *ResNet50*. Except for the *Triplet Network* and *Siamese Network*s, all other methods adopted classification loss. These two unique loss methods performed slightly lower on the *Rank-1* metric compared to other methods, especially the *Triplet Network*. *Autoencoders*, using reconstruction loss, achieved 71.25% on the *mAP* metric, demonstrating their advantage in feature extraction. *Feature Pyramid Networks* achieved the best results on the *Rank-1* metric, reaching 93.21%. However, on the *mAP*, its performance was slightly below that of the model in this paper. The model in this paper achieved the best results on the *mAP* metric, reaching 78.94%, and also performed very well on *Rank-1*, reaching 92.35%. On the *Rank-1* metric, *Feature Pyramid Networks*, *Transfer Learning*, *GAN*, and the model from this paper all achieved accuracy rates over 90%. On the *mAP* metric, both *Transfer Learning* and the proposed model achieved results over 78%, indicating strong robustness.

In conclusion, the method using *ResNet50* as the base network excelled in extracting features from stratum images, especially those adopting classification loss. The model in this paper is comparable in overall performance to other top-tier methods, especially achieving the best results on the *mAP* metric, demonstrating the model's robustness and generalization capability.

Figure 7 presents the *CMC* (Cumulative Match Characteristic) curves for four base network models (*CNN*, *GoogleNet*, *DenseNet121*, and the proposed model). As evident from the chart, the *CNN* model has the lowest recognition rate at *Rank-1*. However, as the *Rank* increases, its recognition rate gradually rises and stabilizes, eventually nearing the rates of other models. At initial *Rank* values, *GoogleNet* outperforms *CNN* but is slightly inferior to *DenseNet121* and the proposed model. In subsequent *Rank*, *GoogleNet*'s growth trend is similar to that of *CNN*, ultimately reaching a comparable steady state. The performance of *DenseNet121* is relatively good across the entire *Rank* spectrum, especially in the initial *Rank*. It consistently maintains the second position, indicating its efficacy in extracting features from stratum images. Across all values of *Rank*, the proposed model consistently exhibits the best performance. From the figure, it is evident that its recognition rate is always higher than the other three models, particularly in the *Rank-1* to *Rank-10* range, where its advantage is even more pronounced.

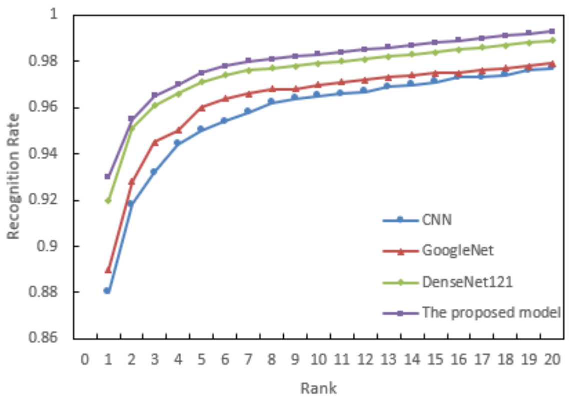

Figure 7. Comparison of *CMC* curves of four base network models

In conclusion, the *CMC* curve of the proposed model is superior, especially at the initial *Rank* values, highlighting its notable performance and robustness in feature extraction from stratum images. Among the four models, *DenseNet121* ranks second, also showcasing relatively stable and superior performance. *GoogleNet* and *CNN* have comparable recognition rates overall but are both outperformed by *DenseNet121* and the proposed model. For tasks related to feature extraction from stratum images, the proposed model offers a more optimal choice, especially in scenarios demanding high accuracy.

Table 2. Experimental results of lightweight rock layer recognition model

Model	Feature Fusion Network	Sample Supervision Method	***mAP***
Before *SSD* improvement	/	/	22.6%
		Anchor	24.5%
	With *NMS i*ntroduced	/	28.9%
		Anchor	31.4%
After *SSD* improvement	/	/	21.5%
		Anchor	22.3%
	With *NMS i*ntroduced	/	26.7%
		Anchor	28.8%

Further, a real-time lightweight stratum recognition model was developed with the classification-regression network structure introduced, and a sample supervision rule was proposed based on anchor points, allowing the model to still achieve high-precision recognition in case of imbalanced sample distribution. Table 2 lists the *mAP* results of two models (before and after *SSD* improvement) under different feature fusion networks and sample supervision methods. From the table, it's evident that regardless of whether it's before or after the *SSD* improvement, using *NMS* always enhances the *mAP*. This indicates that *NMS* effectively filters out redundant detection frames, improving model accuracy. For both pre-improvement and post-improvement *SSD* models, using anchor-based sample supervision boosts the *mAP*, aligning with the paper's statements. By using anchor-point-based sample supervision, the model maintains high recognition accuracy even with imbalanced sample distribution. Compared to the model before *SSD* improvement, the post-improvement *SSD* model has a slightly reduced *mAP* under the same conditions. This may suggest that some newly introduced structures or methods during the improvement might not have achieved the desired effect. However, it's also possible that in the pursuit of lightweight and real-time features, some recognition accuracy was sacrificed.

In conclusion, introducing *NMS* positively impacts the model's *mAP*. The anchor-based sample supervision method effectively enhances the accuracy of stratum recognition, especially in case of imbalanced samples. While the *SSD* model's enhancements might prioritize lightweight and real-time capabilities, there might be a trade-off in terms of accuracy. However, the specific decisions would depend on the actual application scenarios and requirements.

Table 3 demonstrates the performance of the pre-improvement and post-improvement real-time stratum recognition models in terms of CPU usage, startup time, and recognition time. As per the table, the CPU usage of the post-improvement model is slightly higher than the pre-improvement model, increasing from 5.74% to 6%. This suggests that the improved model might have incorporated more complex structures or algorithms, leading to an increase in CPU usage. The startup time for the post-improvement model is marginally shorter, decreasing from 2.16s to 2.14s. Though the difference is minimal, it indicates enhanced startup speed in the improved model. The recognition time of the post-improvement model reduced from 51ms to 41ms, implying that the improved model might require less computational time during the recognition process.

Table 3. Test results of real-time stratum recognition performance

Model	***CPU*** **Usage**	**Startup Time/*s***	**Recognition Time/*ms***
Before improvement	5.74%	2.16	51
After improvement	6%	2.14	41

In conclusion, the improved real-time stratum recognition model has slightly higher CPU usage than its predecessor, likely due to the introduction of more intricate structures or algorithms. In terms of startup speed, the improved model exhibits a slight edge. However, the improved model excels in recognition speed, completing the recognition task in a shorter span.

5. CONCLUSION

This paper presents a pyramid-based stratum image feature extraction network that successfully amalgamates high-level semantic features with low-level feature maps, ensuring comprehensive feature capture of pyramid-structured stratum images. Moreover, a lightweight stratum recognition model oriented for real-time recognition was devised. By integrating a classification-regression network structure and an anchor-based sample supervision rule, the model maintains high precision in recognition even with imbalanced sample distribution.

The *loss* curves indicate a declining trend for both training and validation losses as training epochs progress, signifying the model's learning and gradual optimization. The performance of this paper's model, in terms of *Rank-t* and *mAP* metrics, is on par with models based on different base networks, underscoring its efficacy. Through the introduction of anchor points and the NMS strategy, there's a marked enhancement in the model's *mAP*. While the lightweight improved model sees a minor increase in CPU usage and a slight reduction in startup time, recognition time is notably shortened.

Successfully, this paper has put forth and validated a pyramid-model-based stratum image feature extraction network capable of deeply capturing the features of stratum images. Additionally, to cater to real-time recognition requirements, a lightweight stratum recognition model was developed. By incorporating a classification-regression network structure and an anchor-based sample supervision rule, the model delivers outstanding performance even with imbalanced sample distribution. This research not only paves the way for novel methods and perspectives for real-time stratum image recognition but also holds significant application value and a vast research horizon.

REFERENCES

[1] Zhang, Q., Liu, J., Gu, J., Tian, Y. (2022). Study on coal-rock interface characteristics change law and recognition based on active thermal excitation. European Journal of Remote Sensing, 55(sup1): 35-45. https://doi.org/10.1080/22797254.2022.2031307

[2] Huiling, G., Xin, L. (2019). Coal-Rock Interface Recognition Method Based on Image Recognition. Nature Environment & Pollution Technology, 18(5): 1627-1633.

[3] Zhang, G., Cheng, D., Hou, Y., Li, Z., Zhong, L. (2020). Study on automatic recognition method of Continental Shale Sandy laminae based on electrical imaging image. In Journal of Physics: Conference Series, 1549(2): 022019. https://doi.org/10.1088/1742-6596/1549/2/022019

[4] Chen, J.Y., Huang, H.W., Zhang, D.M., Zhou, M.L., Qin, S.Y., Yang, T.J., Duan, Z.P. (2020). Deep learning based weak inter-layers segmentation and measurement of rock tunnel face. In ISRM International Symposium-EUROCK.

[5] Pascual, A.D.P., Shu, L., Szoke-Sieswerda, J., McIsaac, K., Osinski, G. (2019). Towards natural scene rock image classification with convolutional neural networks. In 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), Edmonton, AB, Canada, pp. 1-4. https://doi.org/10.1109/CCECE.2019.8861885

[6] Greenhalgh, S., Manukyan, E. (2013). Seismic reflection for hardrock mineral exploration: Lessons from numerical modeling. Journal of Environmental and Engineering Geophysics, 18(4): 281-296. https://doi.org/10.2113/JEEG18.4.281

[7] Wang, J., Xue, L., Gao, X. (2023). Identification method of volcanic rock slices based on a deep residual shrinkage network. In Fourth International Conference on Geoscience and Remote Sensing Mapping (GRSM 2022), 12551: 389-394. https://doi.org/10.1117/12.2668168

[8] Guo, K.F., Zhang, Y.P. (2023). Experimental study on the mechanism of gas accumulation and soil deformation in double-layered soils. Rock and Soil Mechanics, 44(1): 99-108. https://doi.org/10.16285/j.rsm.2022.5268

[9] Pang, X.J., Wang, G.W., Kuang, L.C., Lai, J., Gao, Y., Zhao, Y.D., Li, H.B., Wang, S., Mao, M., Liu, S.C., Liu, B.C. (2022). Prediction of multiscale laminae structure and reservoir quality in fine-grained sedimentary rocks: The Permian Lucaogou Formation in Jimusar Sag, Junggar Basin. Petroleum Science, 19(6): 2549-2571. https://doi.org/10.1016/j.petsci.2022.08.001

[10] Zhang, M.C., Zhao, L.J., Wang, Y.D. (2021). Recognition system of coal-rock cutting state based on CPS perception analysis. Meitan Xuebao/Journal of the China Coal Society, 46(12): 4071-4087.

[11] Liu, J., Du, W., Zhou, C., Qin, Z. (2021). Rock Image Intelligent Classification and Recognition Based on Resnet-50 Model. In Journal of Physics: Conference Series, 2076(1): 012011. https://doi.org/10.1088/1742-6596/2076/1/012011

[12] Zhao, L., Sun, X., Liu, F., Wang, P., Chang, L. (2022). Study on morphological identification of tight oil reservoir residual oil after water flooding in secondary oil layers based on convolution neural network. Energies, 15(15): 5367. https://doi.org/10.3390/en15155367

[13] Lai, J., Liu, B.C., Li, H.B., Pang, X.J., Liu, S.C., Bao, M., Wang, G.W. (2022). Bedding parallel fractures in fine-grained sedimentary rocks: Recognition, formation mechanisms, and prediction using well log. Petroleum Science, 19(2): 554-569. https://doi.org/10.1016/j.petsci.2021.10.017

[14] Yang, Z., Guo, N., Zhang, H. (2021). Study on microstructure characteristics of clay rock of Xigeda formation in Xichang city based on softening test and image recognition. In Hydraulic and Civil Engineering Technology VI, 73-78.

[15] Wei, W., Li, L., Shi, W. F., Liu, J.P. (2021). Ultrasonic imaging recognition of coal-rock interface based on the improved variational mode decomposition. Measurement, 170: 108728. https://doi.org/10.1016/j.measurement.2020.108728

[16] Li, X., Su, D., Chang, D., Liu, J., Wang, L., Tian, Z., Wang, S.X., Sun, W. (2023). Multi-scale feature extraction and fusion net: Research on UAVs image semantic segmentation technology. Journal of ICT Standardization, 11(1): 97-116. https://doi.org/10.13052/jicts2245-800X.1115

[17] Zou, P., Teng, Y., Niu, T. (2022). Multi-scale Feature Extraction and Fusion for Online Knowledge Distillation. In International Conference on Artificial Neural Networks, Bristol, UK, pp. 126-138. https://doi.org/10.1007/978-3-031-15937-4

[18] Hu, W., Wang, T., Wang, Y., Chen, Z., Huang, G. (2022). LE–MSFE–DDNet: A defect detection network based on low-light enhancement and multi-scale feature extraction. The Visual Computer, 38(11): 3731-3745. https://doi.org/10.1007/s00371-021-02210-6

[19] Jeon, B.U., Chung, K. (2022). CutPaste-based anomaly detection model using multi scale feature extraction in time series streaming data. KSII Transactions on Internet & Information Systems, 16(8): 2787-2800.

[20] Micó, V., García, J. (2010). Common-path phase-shifting lensless holographic microscopy. Optics Letters, 35(23): 3919-3921. https://doi.org/10.1364/OL.35.003919

[21] Raman, N., Shah, S., Veloso, M. (2022). Synthetic document generator for annotation-free layout recognition. Pattern Recognition, 128: 108660. https://doi.org/10.1016/j.patcog.2022.108660

[22] Chai, S., Zhuang, L., Yan, F. (2023). LayoutDM: Transformer-based diffusion model for layout generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18349-18358.

[23] Tan, Z., Chu, Q., Chai, M., et al. (2022). Semantic probability distribution modeling for diverse semantic image synthesis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(5): 6247-6264. https://doi.org/10.1109/TPAMI.2022.3210085