

RESEARCH

Open Access



A lightweight convolutional neural network based on dense connection for open-pit coal mine service identification using the edge-cloud architecture

Yang Liu and Jin Zhang*

Abstract

Remote sensing is an important technical tool for rapid detection of illegal mining behavior. Due to the complex features of open-pit coal mines, there are few studies about automatic extraction of open-pit coal mines. Based on Convolutional Neural Network and Dense Block, we propose a lightweight densely connected network-AD-Net for the extraction of open-pit coal mining areas from Sentinel-2 remote sensing images, and construct three sample libraries of open-pit coal mining areas in north-central Xinzhou City, Shanxi Province. The AD-Net model consists of two convolutional layers, two pooling layers, a channel attention module, and a Dense Block. The two convolutional layers greatly reduce the complexity of the model, and the Dense Block enhances the feature propagation while reducing the parameter computation. The application is designed in different modules that runs independently on different machines and communicate with each other. Furthermore, we create and build a unique remote sensing image service system that connects a remote datacentre and its associated edge networks, employing the edge-cloud architecture. While the datacentre acts as the cloud platform and is in charge of storing and processing the original remote sensing images, the edge network is largely utilised for caching, predicting, and disseminating the processed images. First, we find out the optimal optimizer and the optimal size of the input image by extensive experiments, and then we compare the extraction effect of AD-Net with AlexNet, VGG-16, GoogLeNet, Xception, ResNet50, and DenseNet121 models in the study area. The experimental results show that the combination of NIR, red, green, and blue band synthesis is more suitable for the extraction of the open-pit coal mine, and the OA and Kappa of AD-Net reach 0.959 and 0.918 respectively, which is better than other models and well balances the classification accuracy and running speed. With this design of edge-cloud, the proposed system not only evenly distributes the strain of processing activities across the edges but also achieves data efficiency among them, reducing the cost of data transmission and improving the latency.

Keywords Cloud, Remote sensing images, Edge processing, Convolutional neural network, Dense block, Channel attention

Introduction

Modern computational technologies including cloud computing, internet of things, and faster networks have applications in all fields of real life. For example, the huge amount of collected data in all areas needs to be stored somewhere, getting processed when needed, and apply certain methods for monitoring

*Correspondence:

Jin Zhang
zjgps@163.com
School of Mining Engineering, Taiyuan University of Technology, Shanxi,
Taiyuan 030024, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

and management. Geographical information system is also an important research field where the application of technology is significant. Mineral resources are an important material basis for the survival of human society and an important guarantee for national security and economic development [1]. China is a vast country, rich in mineral resources, and some areas have shallow buried mineral resources, which are easy to develop. Some illegal miners, for the sake of immediate interests, do not follow the relevant regulations for the predatory mining of various mineral resources, and some even illegally occupy land for private mining without approval. These behaviors have caused the loss of national resources and serious damage to the ecological environment, and it is difficult for the relevant regulatory authorities to quickly detect some illegal mining behaviors [2]. Therefore, fast and accurate access to the land use and destruction of open pit mining areas is significantly important to identify illegal mining behaviors and thus to stop them in time [3].

The open-pit mining can cause dramatic changes to the original land form, however illegal mining is often rapid and violent in order to avoid supervision, and such high-intensity open-pit coal mining can cause serious damage to the surrounding environment. In order to better detect illegal mining activities in a timely manner, many scholars have combined remote sensing images to extract open-pit mine information. Since the traditional image element-based classification is subject to the "salt and pepper" phenomenon, most of the researchers adopt the object-oriented approach to extract the information of open-pit mines. Various research work, as described in Section 2, have shown the applications of deep learning technology for remote sensing images and their efficient processing. However, these methods are implemented over a single machine which might take quite long time for deep learning methods to train model, in particular, when the amount of remote sensing images is quite large.

The above research shows that deep learning technology has been widely applied in remote sensing images, and the deep learning method for recognition is an effective way to promote the remote sensing automatic extraction method to practicality. From the research methods in the study of open-pit coal mine identification, there are not many studies based on deep learning technology in remote sensing image extraction of open-pit coal mines, and the traditional identification methods are mainly used, but the traditional remote sensing automatic extraction methods have problems such as low accuracy, poor generalization, low efficiency, longer training durations, and automation [4]. We believe, the use of modern cloud technology, which has been relatively unexplored for remote sensing, can

help to store the huge amount of images and train the model quickly; therefore, reducing the overall model latencies and execution times.

To address the above issues, we propose a lightweight CNN model based on dense connection to extract open-pit coal mines from Sentinel-2 satellite images, using north-central Xinzhou, Shanxi, China as the study area. We use the edge cloud infrastructure in such a way that the proposed application is divided into different modules, i.e. training and prediction, and each module runs independently on different nodes. We continuously experimented on the dataset, compared the model evaluation accuracy, and found the optimal size of the input data and the optimal hyperparameters of the model. After detailed analysis, the optimal extraction band for open-cast coal mines was determined. Finally, we trained and tested our dataset on AlexNet [5], VGG-16 [6], GoogLeNet [7], Xception [8], ResNet50 [9], and DenseNet121 [10] models, and compared the accuracy of each model. The major highlights and contributions of our research are illustrated below.

- Based on Convolutional Neural Network and Dense Block, we propose a lightweight densely connected network-AD-Net for the extraction of open-pit coal mining areas from Sentinel-2 re-mote sensing images.
- The AD-Net model consists of two convolutional layers, two pooling layers, a channel attention module and a Dense Block. The two convolutional layers greatly reduce the complexity of the model, and the Dense Block enhances the feature propagation while reducing the parameter computation.
- A deep neural framework is designed for edge cloud with different modules that runs independently and communicate with each other to improve the processing efficiency.
- Furthermore, we create and build a unique remote sensing image service system that connects many datacentres and their associated edge networks, employing the edge-cloud architecture.

The remaining part of this manuscript is structured in the following manner. We offer an overview of the related work and previous research in Section 2. The overview of various models (deep learning), and material along with various methods are described in Section 3. The proposed AD-Net network model is described in Section 4. Evaluation parameters, experimental setup, and empirical results are illustrated in Section 5. Finally, Section 6 concludes this manuscript and offers some future research areas and directions for further investigation.

Related work

Demirel et al. (2011) identified, quantified and analyzed high-resolution multispectral satellite images of the open-pit mining area of the Goynuk open-pit mine in Turkey from 2004 to 2008 using SVM classification method [11]. Yuan et al. (2013) used Landsat-tm images with a resolution of 30m as the data source to extract information about the mine area using object-oriented supervised classification [12]. Cheng (2017) selected a SVM classifier to extract and classify the mine development and utilization information from the high-Gaofen-1 satellite image data of Jiangcang fifth open-pit mine in Qinghai Province, and evaluated and compared the accuracy of the classification results with the traditional pixel-based maximum likelihood method, the result showed that the object-oriented classification method combined with SVM has high accuracy and good quality [13]. Shao (2020) used Sentinel-2A satellite imagery as the data source to extract land use information in mining areas using a combination of supervised classification and normalized index calculation, which can improve the classification accuracy as well as the precision extraction of multiple types of features over a large area [14].

Based on Landsat series remote sensing data, Jia et al. (2020) constructs a complete set of rules for extracting mining exploitation sites based on the land use characteristics of open-pit mines, and adopts an object-oriented classification method to realize land use extraction for open-pit mines [15]. Kang et al. (2020) used histogram comparison to obtain the best segmentation scale of typical features, and applied the method to the Gaofen-1 image and ZY-3 image to realize the identification and classification of typical features in mining areas based on high-resolution images [16]. Run and Li et al. (2020) used Landsat-8 satellite images as the data source to extract the classification of Baotou Baiyun Ebo mining area row by object-oriented classification method and compared it with supervised classification and supervised classification methods, and the results showed that the object-oriented method has higher accuracy [17]. Huo et al. (2020) used the fusion of object-oriented ideas and SVM methods with Gaofen-2 satellite data to extract land use information from mining areas in Yuzhou City, Henan Province, mainly from open-pit mining, and the results showed that the method outperformed the fused K-nearest neighbor and object-oriented methods [18]. Vorovenccii et al. (2021) used the Landsat imagery to map the extent of surface mining and reclamation in mining areas, and used different SVM algorithms to classify satellite images to detect changes in the extent of surface mining and reclamation [19, 20].

Different open-pit coal mines have different shapes and constituent elements, and the open-pit coal mine area is a composite target consisting of several sub-regions such as side gang, mining pit and drainage field, and the regional spatial features are extremely heterogeneous. Traditional image feature extraction mainly relies on manually designed feature extractors with poor generalization ability and robustness, so the accuracy of traditional open-pit coal mine extraction methods is not high in practical applications. In recent years, CNN model has been successfully applied to image classification [21, 22], target localization detection [23, 24], target segmentation [25, 26], and other image fields, and can also be applied to speech recognition [27, 28], text classification [29, 30], and many other fields.

Guan et al. combined a clustering-based band selection method with the residual and capsule networks to create a deep model called ResCapsNet and tested the robustness of ResCapsNet using Gaofen-5 imagery [31]. Qian et al. proposed a multi-stream CNN (3M-CNN) model based on multimodal remote sensing data and multi-scale kernel functions [32]. Chen et al proposed a CNN-based object-oriented framework for mapping open pit mines with Gaofen-2 high-resolution satellite images [33]. Camalan et al. used supervised (E-ReCNN) and semi-supervised (SVM-STV) methods to study binary and multiclass variation within the mining pools in the MDD region [34]. Liu et al. introduced the deep learning full convolutional neural network algorithm into the automatic extraction of open pit mine, which fully mines a large number of high-level abstract features from the underlying features to achieve intelligent, effective, and efficient interpretation of the open pit mine [35].

Hu et al. combined deep learning algorithms with object-oriented ideas and used convolutional neural networks to extract information on the type of development land area occupied by mining areas, mainly open pit quarries, using Gaofen-2 images as the study data [36]. In order to solve the problem of low accuracy of CNN for open pit mining site recognition due to the small amount of training data, Cheng et al. found that the migration learning method of freezing the bottom parameters of the CNN pre-training model network and fine-tuning the top parameters had the best training effect through comparative experiments [37]. Zhang et al. improved a fully convolutional neural network based on dense connections and trained an open-pit mining area extraction model for multi-source remote sensing data to finally achieve fully automatic extraction of open-pit mining areas in Tongling area [38].

From the above study we noted that there are not many studies based on deep learning technology in remote sensing image extraction of open-pit coal mines.

Moreover, the traditional identification methods are mainly used, but the traditional remote sensing automatic extraction methods have problems such as low accuracy, poor generalization, low efficiency, longer training durations, and automation [4]. We believe, the use of modern cloud technology, which has been relatively unexplored for remote sensing, can help to store the huge amount of images and train the model quickly; therefore, reducing the overall model latencies and execution times.

Materials and methods

Materials

Study area

The study area of this paper is Xuejiawa Township and several adjacent areas. The study area is located in north-central Xinzhou City, Shanxi Province, with a latitude and longitude range of $112^{\circ}20'45''\sim 112^{\circ}41'30''E$ and $38^{\circ}56'48''\sim 39^{\circ}08'46''N$ (Figure 1). The Xuejiawa Township is a high mountainous and cold dry area with large temperature differences, with an average annual temperature of $6.2^{\circ}C$ and an average annual precipitation of 470-770 mm [39]. The unique geological conditions of the study area make it exceptionally rich in mineral resources. The main mineral species are coal and alumina, and there are also iron, manganese, limestone and other minerals. The main components of the coal field are mostly Carboniferous and Jurassic coal, and the coal reserves are rich in coal with good low ash, low wrestling, and high calorific value, and shallow burial, easy to mine, and mostly open-pit mining [40].

Dataset description

The research data for this paper were obtained from the ESA's Copernicus Data Centre, which provides free access to Sentinel-2 data. The Sentinel-2 consists of two satellites (A and B), launched in June 2015 and March 2017, respectively. Both satellites carry the Multi-Spectral Imager (MSI) with 13 spectral bands that provide images of vegetation, soil and water cover, inland waterways and coastal areas, and can also be used for emergency relief services. Among the optical data, the Sentinel-2 data is the only one containing three bands in the red-edge range, which is very effective for monitoring vegetation health information. The spatial resolution is increased to 10 and 20 meters in the visible and near-infrared bands. One satellite has a revisit period of 10 days, and two complementary satellites have a revisit period of 5 days. We used 10-m resolution Sentinel-2 LA data in bands 2, 3, 4, and 8. the L2A-level data contain mainly radiometric-calibrated and atmospherically corrected atmospheric bottom reflectance data. We downloaded 2018 Sentinel-2 images of the study area with less than 1% image cloud coverage.

Dataset processing

Figure 2 outlines the steps to prepare the dataset. We download Sentinel-2 satellite images of the study area at the ESA Copernicus Data Centre, selecting images with less cloud coverage. The downloaded satellite images have 13 bands, and three datasets were produced using the red, green, blue and near-infrared bands to determine

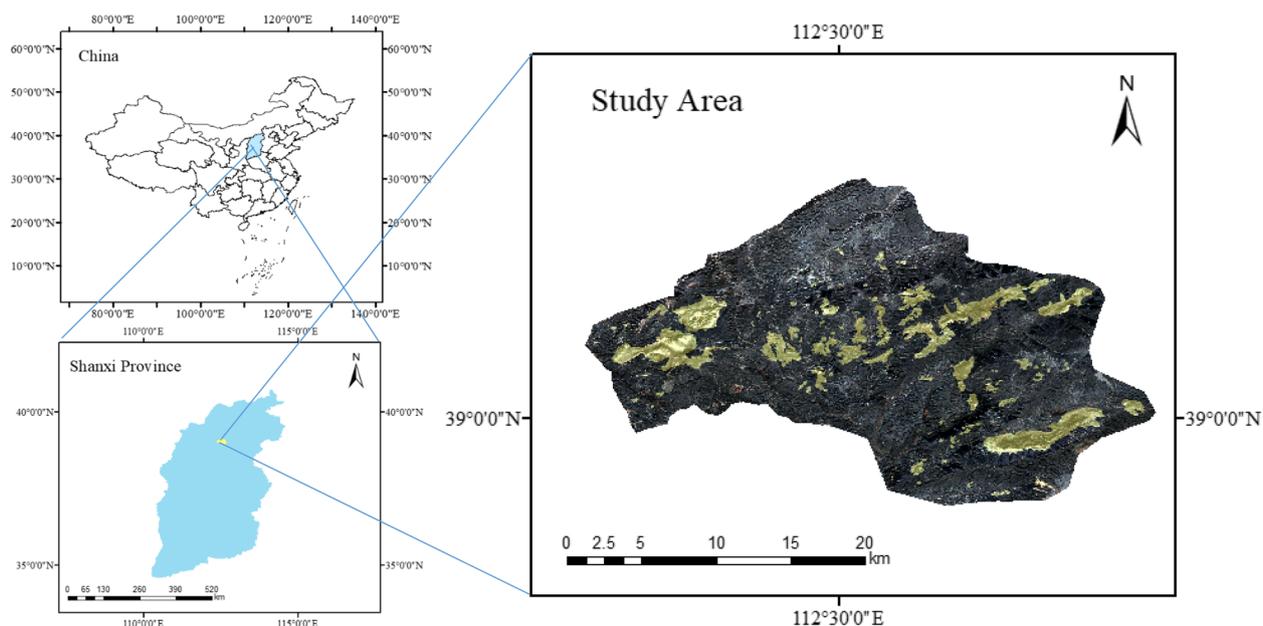


Fig. 1 A brief description of the study area

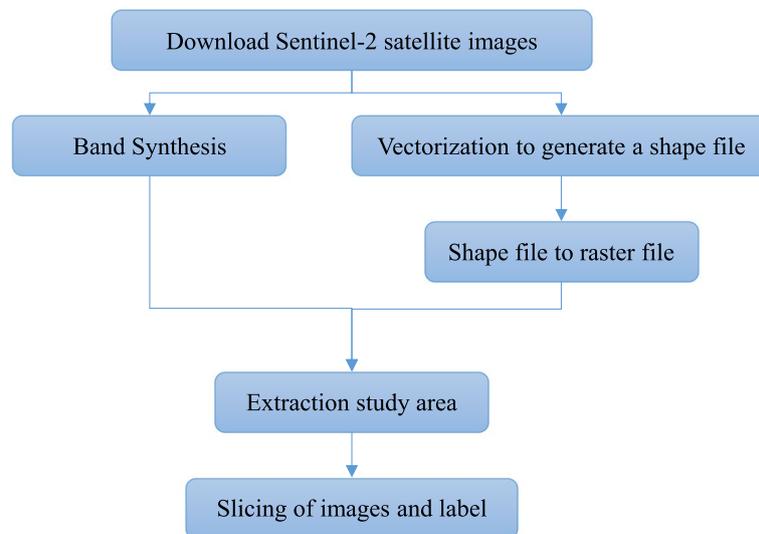


Fig. 2 Data preprocessing and various steps

which band combination was more suitable for the extraction of open-pit coal mines. The first dataset was formed by synthesizing the red, green and blue bands, the second by synthesizing the red, green and NIR bands, and the third using the NIR, red, green and blue bands. Then, the mining area was vectorized using Google Earth high-resolution remote sensing image as a reference and converted to a raster file in the same format as the satellite image, and the format used in this paper is TIFF. The Sentinel-2 images and shp files in the dataset have been projected according to the same coordinate system, ensuring that the mining areas correspond to the labels. The satellite image and label image are cropped according to the shp file of the administrative area, so that the area covered by the label image is obtained to be the same as the cropped satellite image. Finally, the cropped satellite image and label image slices are input to the model.

Methods and techniques

The CNN Model

Convolutional Neural Network is a deep feed-forward artificial neural network, which has been widely used in the field of image recognition due to its local perception and weight sharing [41]. The CNN mainly consists of an input layer, convolutional layer, activation layer, pooling layer, and a fully connected layer.

In a CNN network model that processes images, the input layer is generally a four-dimensional matrix. The first dimension is the number of input images, and the other three-dimensional matrices represent one image. Note that the three-dimensional matrices represent the

number of color channels, the length, and width of the image, respectively. This relationship is mathematically described in Eq. 1.

$$\text{inputsize}(N, C_{in}, H, W) \quad (1)$$

The convolutional layer is the core layer for building the CNN model. The convolutional layer consists of a set of filters, which can be considered as a two-dimensional numerical matrix, and usually convolution helps us to extract image features. The two-dimensional convolution formula is as follows in Eq. 2:

$$s(i, j) = (X * W)(i, j) = \sum_{m=0}^{m_r-1} \sum_{n=0}^{m_c-1} x(i+m, j+n)w(m, n) \quad (2)$$

where X represents the input image, W is the core filter, the number of rows and columns of X are m_r and m_c , respectively, and the number of rows and columns of W are n_r and n_c , respectively. And i, j satisfy the condition $0 < i < m_r + n_r - 1, 0 \leq j < m_c + n_c - 1$.

The activation layer is a nonlinear mapping of the convolutional layer output. The commonly used and well-known activation functions are Sigmoid, Tanh, and ReLU. The principle of the Sigmoid activation function is to map a real number to the interval (0, 1). The Sigmoid activation function is formulated as follows in Eq. 3.

$$\text{sigmoid}(x) = \frac{1}{1 + e^x} \in (0, 1) \quad (3)$$

The Tanh is also a very common activation function. Compared to the sigmoid, its output has a mean value of zero, making it converge faster than the sigmoid and,

therefore, reducing the total number of iterations. The Tanh activation function is mathematically expressed as given in Eq. 4.

$$\tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \in (-1, 1) \tag{4}$$

The ReLU activation function is the most commonly used activation function in deep learning. It can effectively solve the problem of gradient disappearance. The principle of ReLU activation function is to keep the positive numbers and change the negative part to 0 [42]. The formula of ReLU activation function is as follows in Eq. 5.

$$\text{relu}(x) = \max(x, 0) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \in [0, +\infty) \tag{5}$$

Usually, a pooling layer is inserted periodically between successive convolutional layers. It is used to gradually reduce the spatial size of the data body, so that the number of parameters in the network can be reduced, making the computational resources consumed less, and can also effectively control overfitting, the common pooling layer methods are Max pooling and average pooling.

Maximum pooling is taking the maximum value for a local, which is mathematically defined by the following Eq. 6.

$$Y_{kR_{ij}} = \max_{(p,q) \in R_{ij}} x_{k(p,q)} \tag{6}$$

where k denotes the kth feature map, R_{ij} denotes the rectangular region, i and j denote the number of ranks in the rectangular region, and $x_{k(p,q)}$ denotes the element located at (p,q) in the rectangular region R_{ij} .

The averaging pooling is the process of averaging over local values and is defined by the following Eq. 7.

$$Y_{kR_{ij}} = \frac{1}{i \times j} \sum_{(p,q) \in R_{ij}} x_{k(p,q)} \tag{7}$$

where k denotes the kth feature map, R_{ij} denotes the rectangular region, $i \times j$ is the number of elements in the rectangular region, and $x_{k(p,q)}$ denotes the elements located at (p,q) in the rectangular region R_{ij} .

After multiple rounds of convolutional and pooling layers, the final classification result is generally given by the fully connected layer at the end of the CNN. While operations such as convolution, pooling and activation function layers map the original data to the hidden feature space, the fully connected layer serves to map the learned "distributed feature representation" to the sample labeling space.

Dense block and its architecture

DenseNet was proposed by Huang et al. The model borrows the idea of ResNet, both contain jump connection

structure, the difference is that in ResNet the layers and layers are summation of elements, in Densenet the layers and layers are concatenated in dimension, so its network structure is completely new. The most important contribution of DenseNet is the tightly connected CNN, which uses a large number of Dense Blocks (Figure 3) in the network. Any two layers in this neural network are directly connected, meaning that the input of each layer in the network is the concatenation of the output of all the previous layers, and the features learned in this layer are directly passed to all the later layers as input.

Suppose the input is an image X_0 , which passes through an L-layer neural network, where the nonlinear transformation of layer i is denoted as $H_i(*)$, and $H_i(*)$ can be the accumulation of various function operations such as BN, ReLU, Pooling or Conv. The feature output of the ith layer is denoted as X_i , which is mathematically expressed as shown in the following Eq. 8.

$$X_i = H_i([X_0, X_1, \dots, X_{i-1}]) \tag{8}$$

where $[]$ stands for concatenation, which is the overlay of all output feature maps from X_0 to X_{i-1} layers by Channel.

Channel attention module

The channel attention structure is shown in Figure 4, and the formula is shown in (2). Input feature F of size $H \times W \times T$, where H and W represent the length and width of the input feature, respectively, and T represents the number of channels. First, global average pooling and global max pooling are performed on the input feature layer, respectively, to compress the input features into a size of $1 \times 1 \times T$. Secondly, through the two-layer full connection of reducing the number of channels

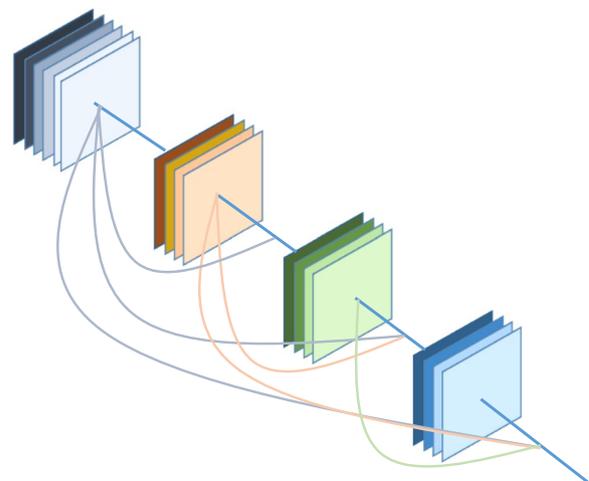


Fig. 3 The dense Block and its structure

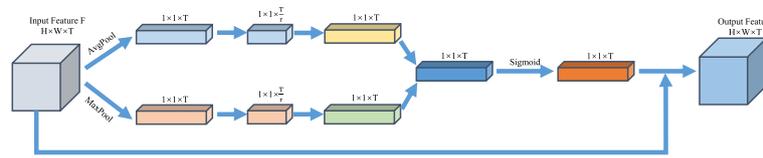


Fig. 4 Channel attention and its model architecture

and restoring the number of channels, the results of the two full connections are added element-wise to obtain $1 \times 1 \times T$ features. Perform a sigmoid operation on the obtained features, so that the weight of each feature point is between 0-1, and obtain the weight coefficient of each channel. Finally, the obtained weight coefficients are multiplied by the corresponding input features to realize the weighted distribution of features in different channels. The entire process is expressed mathematically using Eq. 9.

$$M_C(F) = Sigmoid(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (9)$$

The proposed AD-Net model

Based on the existing knowledge, we constructed AD-Net (Figure 5), a lightweight densely connected network, to realize the extraction of open-pit coal mines. First, we use a two-layer convolutional network to extract features, and second, we replace the convolutional layers with a dense block, in which six layers of convolution are designed, and the input of each layer is a concatenation of the outputs of all previous layers, and the output of each

layer is a determined number of channels. Such a design number alleviates the problem of gradient disappearance, enhances the propagation of features, greatly reduces the number of parameters, has a regularization effect, and reduces overfitting even on a smaller training set. Finally, after the feature extraction network, a channel attention module is introduced, which makes the convolutional neural network pay more attention to the information of each channel.

Loss function and evaluation metrics

The loss function is usually used to measure the goodness of the model prediction, and the model can be solved and evaluated by minimizing the loss function. The Cross Entropy Loss Function is a commonly used loss function, in the case of binary classification, the model finally needs to predict the results of only two cases, for each category our prediction to get the probability of p and $1-p$, binary classification of CE Loss formula is as follows in Eq. 10.

$$L = \frac{1}{N} \sum_i - [y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)] \quad (10)$$

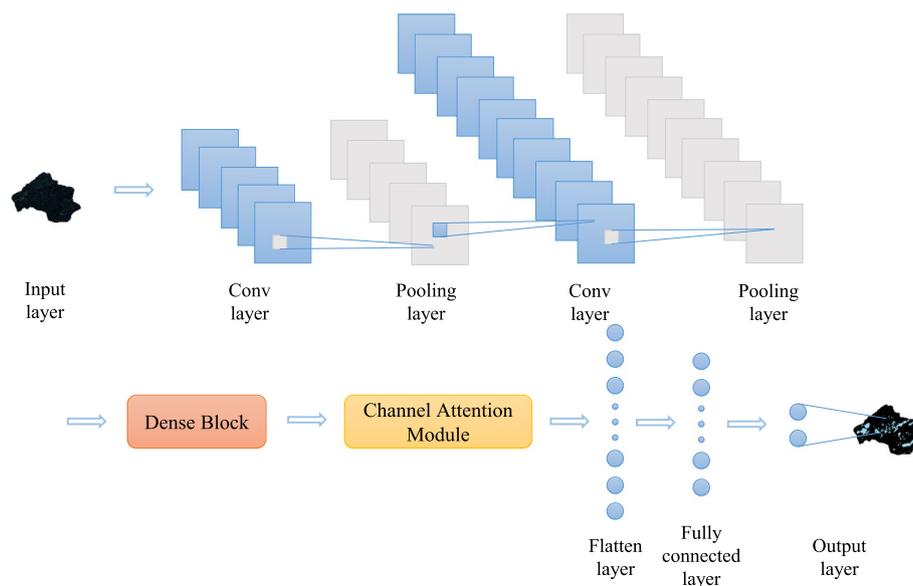


Fig. 5 The proposed AD-Net deep learning architecture

where N denotes the number of samples, y_i denotes the label of sample i , positive class is 1, negative class is 0, and p_i denotes the probability that sample i is predicted to be a positive class.

Evaluation metrics are used to evaluate the accuracy of the semantic segmentation model. The commonly used evaluation metrics are Precision, Recall, F1-Score, OA and Kappa. In all formulas, "TP" is read as true positive, showing correctly identified open-pit coal mine pixels, "FP" is read as false positive, showing incorrectly identified open-pit coal mine pixels, "TN" is read as true negative, showing correctly identified surface coal mine pixels rejected, and "FN" is read as false negative, which means incorrect surface coal mine pixels are rejected.

Precision

The number of correct predictions as a percentage of the total number of positive predictions is known as precision. This metric can be computed using the rate of true positive (TP) and false positive (FP) as given in Eq. 11.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (11)$$

Recall

The number of positive classes predicted to be correct for all data marked as positive. The Recall metric is computed using the ratio of true positive and false negative as given in Eq. 12.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (12)$$

F1-Score

This metric is a statistical measure of the accuracy of a binary classification model, defined as the summed average of precision and recall. The F1-Score can be computed using the formula as illustrated in Eq. 13.

$$\text{F1 - score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (13)$$

OA (Overall Accuracy)

The ratio of the number of correctly classified samples to the number of all samples is denoted as the overall accuracy. The overall accuracy metric is computed using the formula given in Eq. 14.

$$\text{OA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (14)$$

Kappa (Kappa Coefficient)

The Kappa Coefficient is an indicator which is used for consistency testing, for classification problems is to test whether the model prediction results and the actual classification results are consistent. The Kappa Coefficient is computed using the formula stated in Eq. 15.

$$K = \frac{P_0 - P_e}{1 - P_e} \quad (15)$$

where P_0 is the sum of the number of correctly classified samples in each category divided by the total number of samples, which is, in fact, the OA metric as described above.

Assuming that the number of real samples in each category is a_1, a_2, \dots, a_c , and the number of predicted samples in each category is b_1, b_2, \dots, b_c , and the total number of samples is n , then we have: $P_e = \frac{a_1 \times b_1 + a_2 \times b_2 + \dots + a_c \times b_c}{n \times n}$

Edge-cloud for distributed CNN model

The edge cloud platform is used to implement various phases of the smart application as shown in Figure 6. The edge platform is responsible to preprocess the collected data and then send it for storage and training purposes to the remote cloud. Figure 7 discuss how the two important modules of the smart application i.e. channel attention module and the space attention module are implemented over the edge-cloud infrastructure so that the data can be process locally and only essential data is used for training purposes. This can help in reducing the training time and the application latency. The prediction happens at the edge nodes while the model training occurs at the remote cloud in a distributed AI fashion. In fact, the entire application is designed in two different modules i.e. model training and model prediction. Both modules communicate with each other while the training happens at the remote cloud due to the huge amount of images and their characteristics. Similarly, the prediction module is running on the edge node and use the trained model to get outcomes. Despite this, the edge node can also implement some sort of data aggregation and pre-processing method that enables to remove redundant data and sends only important data to the remote cloud.

Results and discussion

All experiments in this study were implemented using the TensorFlow (version 2.8.0) framework on a GeForce RTX 3090 server with 24 GB of memory and Windows 10 as the operating system. For each dataset, the non-mining pixels were larger than the mining pixels, and to balance the positive and negative samples, an equal number of negative samples were selected based on the positive

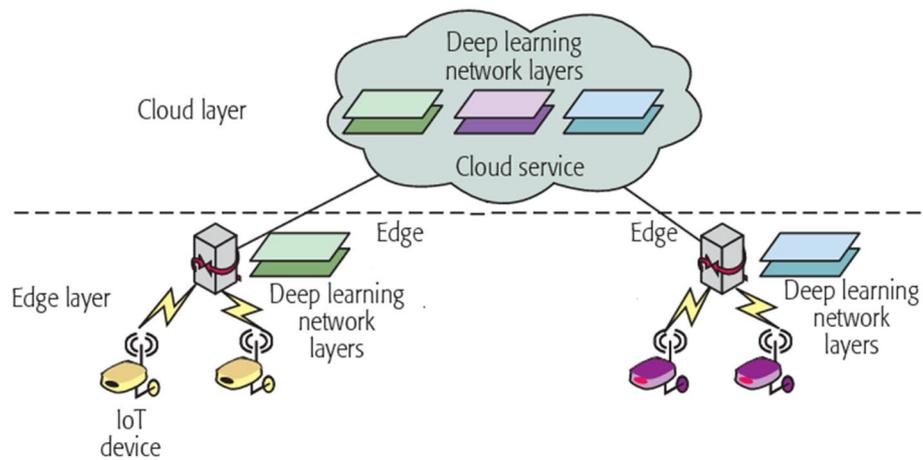


Fig. 6 The proposed edge-cloud deep learning architecture

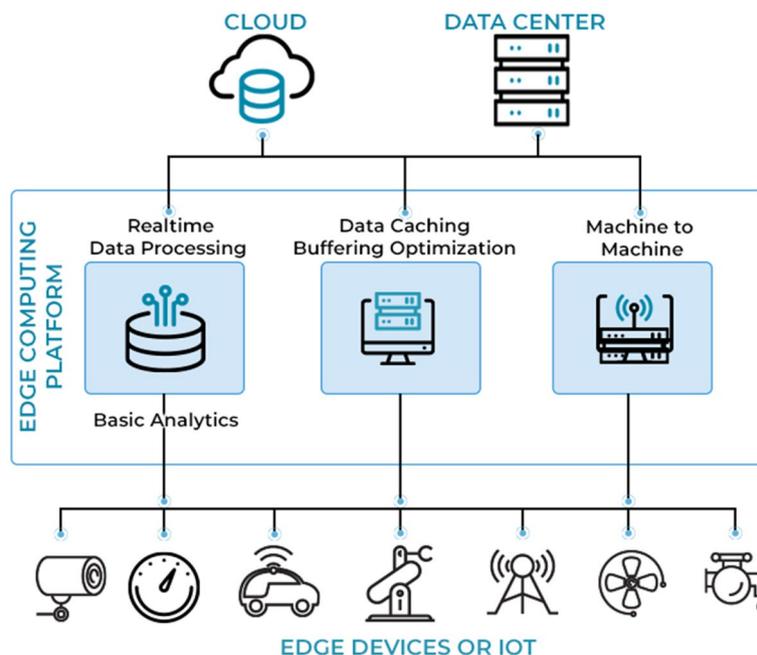


Fig. 7 An edge-cloud platform for distributed AI

sample pixels, and the training The ratio of the training set to the validation set is 6:4. To achieve the best accuracy of the model, various experiments are conducted on the model, which are divided into two main parts:

- Part I: The optimal slice size of the input data and the optimal optimizer of the model are determined.
- Part II: The optimal model is compared with the advanced CNN and RNN models, and Recall, Precision, F1, OA, and kappa are used as evaluation metrics.

Slice experiments

The slice size of the input network has an impact on the training results, because the image features vary depending on the image scale. The larger the image is, the more texture and contextual information it has, and the more important features it can capture. In addition, some discriminative features are better obtained when the image becomes larger. However, when the size becomes large enough, the classification performance may no longer improve or even decrease, and the computational effort

Table 1 Evaluation of the proposed model with the different slice sizes on dataset 1

Slice size	Recall	Precision	F1-Score	OA	Kappa
9 × 9	0.975	0.897	0.934	0.931	0.861
17 × 17	0.934	0.958	0.946	0.947	0.908
33 × 33	0.974	0.914	0.943	0.941	0.881

Table 2 Evaluation of the proposed model with the different slice sizes on dataset 2

Slice size	Recall	Precision	F1-Score	OA	Kappa
9 × 9	0.959	0.927	0.942	0.941	0.882
17 × 17	0.967	0.937	0.952	0.951	0.902
33 × 33	0.931	0.945	0.938	0.934	0.889

Table 3 Evaluation of the proposed model with the different slice sizes on dataset 3

Slice size	Recall	Precision	F1-Score	OA	Kappa
9 × 9	0.937	0.947	0.942	0.942	0.885
17 × 17	0.943	0.961	0.952	0.952	0.905
33 × 33	0.886	0.964	0.923	0.927	0.898

becomes much larger accordingly. The size of Sentinel-2 images and shp files in this experimental study area is 3001×2205, which were cut into slices of 9×9, 17×17, and 33×33 sizes, respectively. Tables 1, 2, and 3 show the evaluation results of AD-Net based on three datasets, including Recall, Precision, F1-Score, OA, and Kappa.

In this experiment, the input sizes were sliced in intervals of two times minus one for the three datasets starting from 9 × 9. In all three datasets, 17 × 17 provided the highest Kappa relative to the other slice sizes. overall, the Kappa first increased and then decreased as the slice size increased, and the Precision and F1-score of 9 × 9, 17 × 17, and 33 × 33 were all around 0.9, and the OA of all 12 experiments was above 0.9, indicating that they all extracted well open pit coal mine. However, from the kappa coefficient, only 17×17 exceeded 0.9, and its Recall, Precision, F1 and OA were basically higher than the other slice size evaluation results, so the slice size of 17×17 was selected for the follow-up study and analysis.

Experimental parameters

After calculating the loss value of the model using the loss function, the loss value needs to be used to optimize the model parameters. The internal parameters of the model affect the model training, as well as, the training

Table 4 Evaluation of the proposed model with the different optimizer on dataset 1

Slice size	Recall	Precision	F1-Score	OA	Kappa
SGD	0.949	0.953	0.951	0.951	0.902
AdaGrad	0.968	0.907	0.936	0.934	0.867
RMSprop	0.904	0.963	0.933	0.935	0.871
Adma	0.970	0.945	0.957	0.957	0.913

Table 5 Evaluation of the proposed model with the different optimizer on dataset 2

Slice size	Recall	Precision	F1-Score	OA	Kappa
SGD	0.970	0.948	0.959	0.958	0.916
AdaGrad	0.958	0.936	0.947	0.946	0.892
RMSprop	0.944	0.956	0.950	0.950	0.900
Adma	0.986	0.938	0.961	0.960	0.920

Table 6 Evaluation of the proposed model with the different optimizer on dataset 3

Slice size	Recall	Precision	F1-Score	OA	Kappa
SGD	0.987	0.935	0.960	0.959	0.918
AdaGrad	0.927	0.953	0.940	0.941	0.882
RMSprop	0.974	0.938	0.955	0.955	0.909
Adma	0.953	0.964	0.959	0.959	0.918

results, so we need to use various optimization strategies and algorithms to update and calculate the network parameters that affect the model training and model output to approximate or reach the optimal values. The SGD, AdaGrad, RMSprop, and Adma were tested on three different datasets. The following tables 4, 5, and 6 show the evaluation metrics of different optimization functions based on the three datasets, including the Recall, Precision, F1-Score, OA and the Kappa key performance indicators (KPIs).

Comparing Tables 4, 5, and 6, the model with the Adam optimizer outperforms all the optimizers in the same table, and it performs better on all three data sets. In Dataset 1, comparing OA and Kappa values from four experiments, AdaGrad performed the worst among the four optimizers, with Adma exceeding AdaGrad by 2.5% and 5.3%, RMSprop by 2.4% and 4.8%, and SGD by 0.06% and 1.2%. In dataset 2, Adma’s OA and Kappa values exceeded AdaGrad by 1.5% and 3.1%, exceeded RMSprop by 1.0% and 2.2%, and exceeded SGD by 0.2% and 0.4%. From the experiments of dataset 1 and dataset 2, it can be seen that SGD performs second only to Adma, and the

difference is not large. In dataset 3, the OA and Kappa values of both reach the same, but Adma is faster, so the Adma optimizer is chosen for subsequent research and analysis.

Comparison of different CNN models

The optimal slice and the optimal optimizer were selected based on the slice experiment and the parameter experiment, in addition to comparing the OA and Kappa values of the three datasets with the same slice size and the same optimizer respectively, it is easy to see that the experimental results on dataset 3 are better than the other two datasets. This implies that the combination of NIR, red, green and blue bands is more suitable for the extraction of open pit coal mines in 10 m spatial resolution satellite images. The results are shown in Figure 8.

To further validate the model accuracy, we used other classical convolutional neural networks to compare with the proposed model. All these models were trained on dataset 3 and their performance was recorded.

Table 7 shows the Recall, Precision, F1-Score, OA and Kappa of the proposed model with the classical classification model on each dataset 3. Figure 8 visually shows that the proposed model outperforms all other comparative algorithms, and it can be seen that AD-Net has significantly better accuracy than AlexNet, VGG-16, GoogLeNet, Xception and ResNet50 with a significant improvement in accuracy, and although the improvement in accuracy over DenseNet121 is not significant, the processing speed is much faster than DenseNet121 because of the relatively simple model structure.

The predicted results are continuous probability values and not a binary map of 0 or 1. By using GIS

Table 7 Evaluation of the models and the proposed model on dataset 3

Net	Recall	Precision	F1-Score	OA	Kappa
VGG-16	0.470	0.991	0.638	0.733	0.503
AlexNet	0.997	0.817	0.898	0.887	0.766
GoogLeNet	0.973	0.904	0.034	0.912	0.824
ResNet50	0.927	0.959	0.942	0.943	0.887
Xception	0.983	0.913	0.947	0.945	0.889
DenseNet121	0.975	0.926	0.950	0.949	0.896
AD-Net	0.953	0.964	0.959	0.959	0.918

software and setting a suitable threshold, the probability values can be binarized. The threshold values were set to 0.7, 0.8 and 0.9 for comparison, and the experiments proved that 0.8 gave better results. In general, most of the surface coal mines can be predicted well and the boundaries are more accurate with strong continuity. However, there are some errors in the prediction results. As shown in Figure 9, certain features with similar characteristics to the open-pit mines are also split out, or certain open-pit mines with insignificant characteristics are missed.

The goodness of the extraction results depends largely on the quantity and quality of the dataset. To improve the detection accuracy, some image pre-processing methods such as de-blurring can be added and the number of such training samples can be increased to improve the detection accuracy. In the future, we would like to refine the task of extracting open pit coal mines to extract more information from the images, such as mining sites, discharge sites and side gangs.

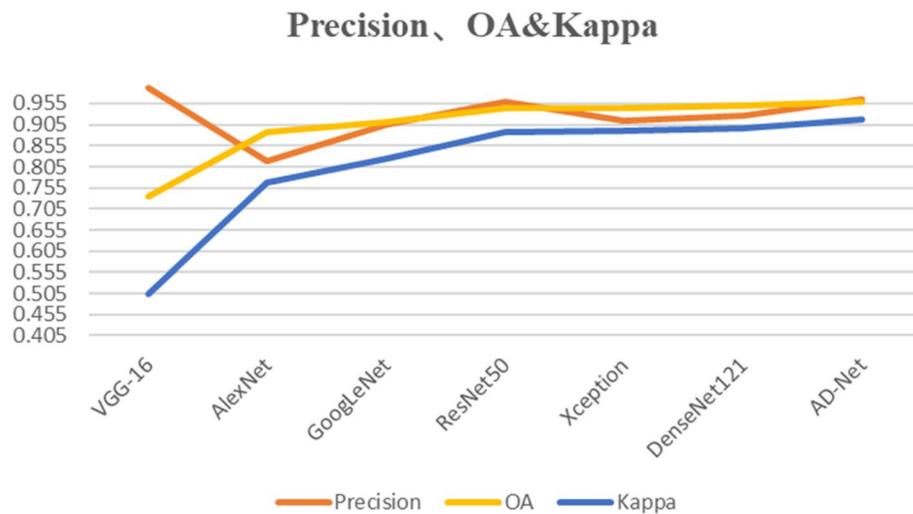


Fig. 8 Comparing the Precision, OA, and Kappa metrics of the models and the proposed model on dataset 3

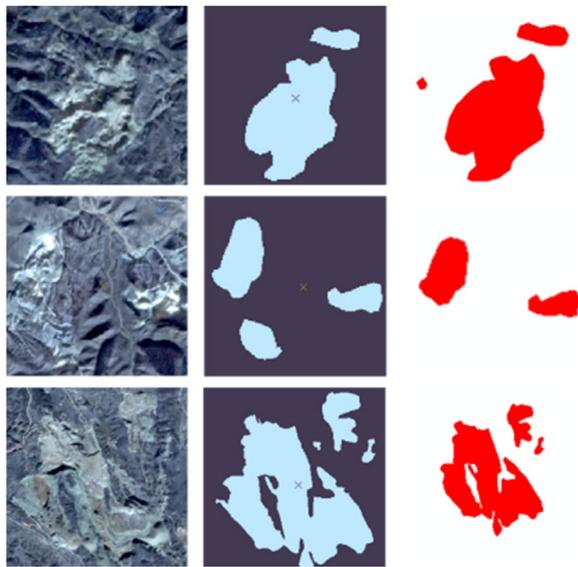


Fig. 9 The first column is the original image, the second column is the ground true data, and the third column is the proposed model extraction result

Table 8 Characteristics of machines used for simulated edge-cloud setup

CPU model	Speed (GHz)	Cores	ECUs	Memory (GB)	Storage (TB)
Servers					
E5645	2.400	12	18.8	32	8
Virtual Machines					
t1.tiny	800 MIPS	1	1	0.6	650 MB
t2.micro	1200 MIPS	1	1	0.613	1 GB

Results of AD-Net model over the edge-cloud setup

The two small edge centers (edge nodes) and one big datacenter (remote cloud) depicted in the above Figure 6 were assumed to be used for running and executing various modules of the proposed AD-Net system. The iFogSim simulator was used to mimic the infrastructure [43]. Each edge center has a single computer, while the datacenter has ten identically-designed machines with the same specifications as listed in

Table 8 below. Note that the ECU is related to the cloud notion (elastic compute unit) and is computed through multiplying the CPU speed (GHz) with the total number of cores. A network with a 1GB/s connection capacity connects the datacenter to the edges [44]. Additionally, we assume that all servers are virtualized and that the same servers are hosting a variety of virtual machines, as shown in Table 8. The AD-Net-Cloud application modules then execute within the virtual computers. Furthermore, it is anticipated that all services would utilize the computing resources they are given in a regularly distributed fashion. We also assume that each AD-Net module use the location or resource that was first assigned to it. Different metrics such as model training durations, model prediction times, RMSE, and the MAPE indicators were used to measure the accuracy and performance of the proposed AD-Net model against the closest rivals (i.e. CNN and RNN) over the edge-cloud environment.

The results, in terms of model training and prediction times, which we obtained on the above setup are shown in Table 9 and Figure 10. We observed a significant decrease (approximately 23.6% to 28.1%) in the model prediction when using the edge node. Similarly, we also noted an approximate improvement of 12.8% to 16.4% in the model training time. This should be noted that the prediction time is proportional to the amount of images that are stored on the remote cloud. Therefore, this may variate significantly due to images, sizes, numbers, and even the amount computational resources of the cloud infrastructure.

The results for various models against the proposed AD-Net scheme, in terms of accuracy (RMSE and MAPE), are shown in Figure 11. The mean absolute percentage error is referred to as the MAPE factor. The MAPE metric is the typical multiplicative relationship amongst each calculated average of mean value and the predicted value of result that was observed. The root mean squared error, or RMSE, is another name for statistical measure i.e. standard deviation. We can observe that the overall impact of the accuracy when using the cloud over the edge cloud is similar. Albeit, for the AD-Net model there is a significant increase in the MAPE value for the edge-cloud; however, we

Table 9 Experimental results for model training and prediction using cloud and edge-cloud setup

	CLOUD		EDGE-CLOUD	
	Training time (s)	Prediction time (s)	Training time (s)	Prediction time (s)
CNN	10,543.6	4598.34	8451.6	1297.12
RNN	10,751.5	4384.4	8341.66	1188.3
U-net	10,734.95	4381.64	8344.01	1131.53

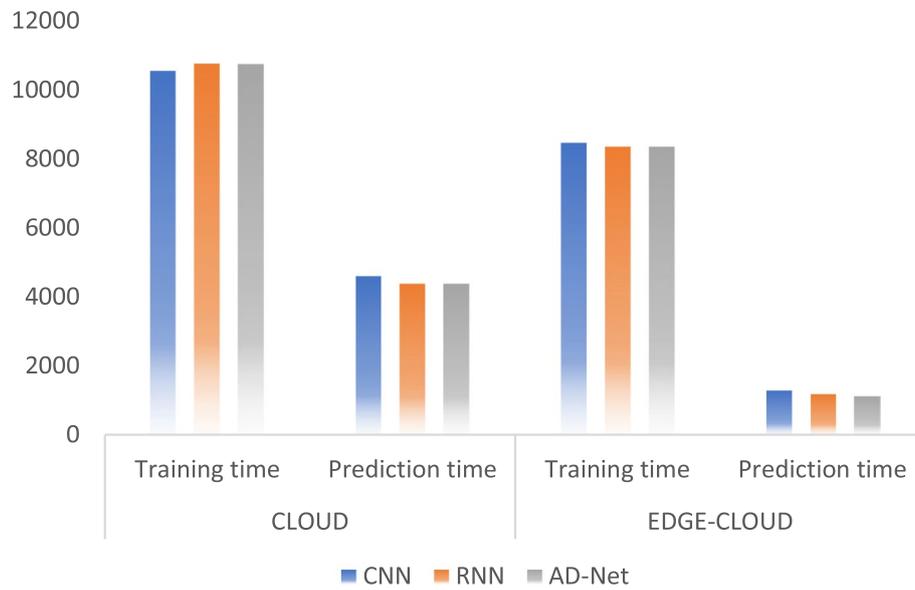


Fig. 10 Model training and prediction times using cloud and edge-cloud platform across various deep learning methods

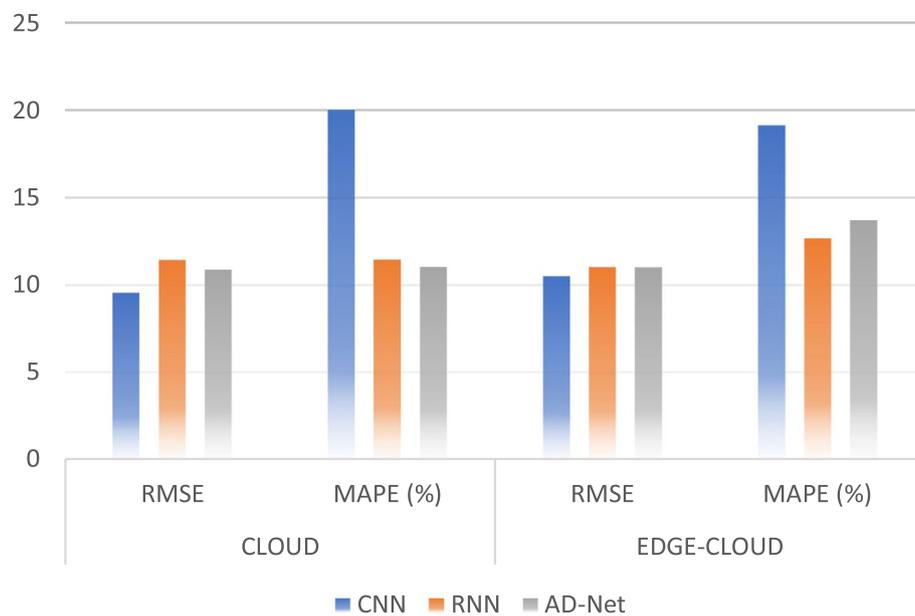


Fig. 11 Accuracy of various deep learning methods using cloud and edge-cloud platforms

also observed a decrease for the CNN model. Also, the RMSE values for all methods are similar on both setup; however, the MAPE values are quite different and show significant improvements. In general, the model training is more accurate on the cloud resources while the model prediction is latency-efficient of the edge infrastructure. Compared to CNN and RNN, the proposed AD-Net model is more accurate both in training and prediction modules.

Conclusions and future work

In this paper, three datasets are created for the task of extracting open-pit coal mines in Sentinel-2 images, using north-central Xinzhou City, Shanxi Province as the study area. A lightweight dense connection model, AD-Net, based on CNN and DenseNet, is proposed. For this model, the optimal slice size and optimizer were determined by extensive experiments. To better extract open-air coal mines from Sentinel-2 data, the

optimal combination of Sentinel-2 satellite bands is investigated, and NIR, red, green and blue band synthesis is found to be superior to red, green and blue band synthesis and NIR, red and green band synthesis. The proposed model is compared with other CNN models, including AlexNet, VGG-16, GoogLeNet, Xception, ResNet50, and DenseNet121. The experimental results show that the Recall, Precision, and F1-score of AD-Net are better than the other compared models, and the proposed model can extract open-air coal mines quickly and accurately.

To improve the detection accuracy, some image pre-processing methods such as de-blurring can be added and the number of such training samples can be increased to improve the detection accuracy. In the future, we would like to refine the task of extracting open pit coal mines to extract more information from the images, such as mining sites, discharge sites and side gangs. Similarly, we will also investigate the impacts of various network activation functions over the achieved results and findings. We will investigate some data pre-processing methods that can reduce the redundancy of collected data so that only important data is stored and used for the training purpose. This will significantly reduce the model training and prediction durations. Furthermore, it will reduce the network traffic that will subsequently improve the network performance.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Project No. 42171424).

Authors' contributions

Conceptualization, methodology, and validation were carried out by Y.L.; original draft preparation by Y.L.; review and editing were carried out by J.Z. All authors have read and agreed to the published version of the manuscript.

Availability of data and materials

The data was freely downloaded from the given website (<https://scihub.copernicus.eu/dhus/#/home>), all accessed on 1 August 2022. Not applicable.

Declarations

Competing interests

The authors declare no competing interests.

Received: 24 October 2022 Accepted: 13 February 2023

Published online: 07 March 2023

References

- Xu YJ, Wang L (2011) The importance of mineral resources. *Publ Comms Sci Technol* 19(2):28–29
- Wu W.D. Research on the Harm of Illegal Mining and Its Countermeasures. *Journal of the Party School of the Taiyuan's Committee of the C.P.C.* 2006, *S1*, 19–21. DOI: CNKI: SUN: TYSW.0.2006-S1-008
- Zhao JL, Chen H (2019) Application of high resolution remote sensing image for dynamic monitoring of illegal mining in coal mines. *Sat App* 07:18–23
- Dong WX, Liang HT, Liu GZ, Hu Q, Yu X (2022) Review of deep convolution applied to target detection algorithms. *J Front Comput Sci Technol* 16(05):1025–1042. <https://doi.org/10.3778/j.issn.1673-9418.2111063>
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. (2012) ImageNet classification with deep convolutional neural networks. *advances in neural information processing systems* . 25(2). <https://doi.org/10.1145/3065386>
- Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* 2014, [arXiv:1409.1556](https://arxiv.org/abs/1409.1556).
- Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA 21:1–9. <https://doi.org/10.1109/CVPR.2015.7298594>
- Chollet, F.; Xception: (2017) Deep learning with depthwise separable convolutions. *arXiv* , [arXiv:1610.02357](https://arxiv.org/abs/1610.02357).
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. *IEEE Comput Soc Conf Comput Vis Pattern Recognit* 2016:770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. *IEEE Conf Comput Vis Pattern Recognit* 2017:2261–2269. <https://doi.org/10.1109/CVPR.2017.243>
- Demirel N, Emil MK, Duzgun HS (2011) Surface coal mine area monitoring using multi-temporal high-resolution satellite imagery. *Int J Coal Geol* 86(1):3–11. <https://doi.org/10.1016/j.coal.2010.11.010>
- Yuan DB, Liu CL, Wang GB (2013) The application and the research of object-oriented method for extraction of mining area information. *Remote Sens Inform* 28(02):110–115
- Cheng, L. (2017) Application of object - oriented combined SVM information extraction of open - pit mine. *Qinghai University, Xining, Qinghai, CNKI: CDMD:2.1017.828168*
- Shao AR, Li XJ, Zhou JJ (2020) Extraction method of mining land use information based on sentinel-2A image. *J Shandong Agric Univ (Natural Science Edition)* 51(03):441–446. <https://doi.org/10.3969/j.issn.1000-2324.2020.03.010>
- Jia, M, (2020) Information extraction and dynamic monitoring of open-pit mining area based on remote sensing technology. *North China University of Science and Technology, Tangshan, Hebei.* <https://doi.org/10.27108/d.cnki.ghelu.2020.000697>
- Kang JF, Zhou L, Zhao DJ, Wen XJ (2020) High resolution remote sensing image object recognition method based on histogram feature knowledge base. *Chinese Rare Earths* 41(02):32–40. <https://doi.org/10.16533/J.CNKI.15-1099/TF.202002004>
- Huang PY, Li A (2020) Contrastive analysis of extraction methods for ground object information in mining area based on remote sensing image. *J Tangshan Univ* 33(03):42–46. <https://doi.org/10.16160/j.cnki.tsxyxb.2020.03.009>
- Huo GJ, Hu NX, Chen T, Zhen N (2021) Mining land use information extraction based on combining support vector machine and object-oriented method. *J Polytechnic Univ (Natural Science)* 40(02):70–75. <https://doi.org/10.16186/j.cnki.1673-9787.2020070002>
- Vorovencii I (2021) Changes detected in the extent of surface mining and reclamation using multi temporal Landsat imagery: a case study of Jiu Valley, Romania. *Environ Monit Assess* 193(1):30. <https://doi.org/10.1007/s10661-020-08834-w>
- Hu, Kai, et al. (2022) "LCDNet: Light-weighted cloud detection network for high-resolution remote sensing images." *IIEEE J Sel Top Appl Earth Obs Remote Sens* 15 4809–4823.
- Tas S, Sari O, Dalveren Y, Pazar S, Kara A, Derawi M (2022) Deep learning-based vehicle classification for low quality images. *Sensors* 22:4740. <https://doi.org/10.3390/s22134740>
- Khoeun R, Chopuk P, Chinnsarn K (2022) Emotion recognition for partial faces using a feature vector technique. *Sensors* 22:4633. <https://doi.org/10.3390/s22124633>
- Itu R, Danescu R (2022) Part-based obstacle detection using a multiple output neural network. *Sensors* 22:4312. <https://doi.org/10.3390/s22124312>
- Charouh Z, Ezzouhri A, Ghogho M, Guennoun Z (2022) A resource-efficient CNN-based method for moving vehicle detection. *Sensors* 22:1193. <https://doi.org/10.3390/s22031193>
- Perrolas G, Niknejad M, Ribeiro R, Bernardino A (2022) Scalable fire and smoke segmentation from aerial images using convolutional neural networks and quad-tree search. *Sensors* 22:1701. <https://doi.org/10.3390/s22051701>

26. Hwang B, Kim J, Lee S, Kim E, Kim J, Jung Y, Hwang H (2022) Automatic detection and segmentation of thrombi in abdominal aortic aneurysms using a mask region-based convolutional neural network with optimized loss functions. *Sensors* 22:3643. <https://doi.org/10.3390/s22103643>
27. Mihalache S, Burileanu D (2022) Using voice activity detection and deep neural networks with hybrid speech feature extraction for deceptive speech detection. *Sensors* 22:1228. <https://doi.org/10.3390/s22031228>
28. Trinh Van, L.; Dao Thi Le, T.; Le Xuan, T.; Castelli, E. (2022) Emotional speech recognition using deep neural networks. *Sensors* , 22, 1414. <https://doi.org/10.3390/s22041414>
29. Yu H, Bae J, Choi J, Kim H (2021) LUX: smart mirror with sentiment analysis for mental comfort. *Sensors* 21:3092. <https://doi.org/10.3390/s21093092>
30. Nagaoka Y, Miyazaki T, Sugaya Y, Omachi S (2021) Text detection using multi-stage region proposal network sensitive to text scale. *Sensors* 21:1232. <https://doi.org/10.3390/s21041232>
31. Guan R, Li Z, Li T, Li X, Yang J, Chen W (2022) Classification of heterogeneous mining areas based on rescapsnet and gaofen-5 imagery. *Remote Sens* 14:3216. <https://doi.org/10.3390/rs14133216>
32. Qian M, Sun S, Li X (2021) Multimodal data and multiscale kernel-based multistream CNN for fine classification of a complex surface-mined area. *Remote Sens* 13:5052. <https://doi.org/10.3390/rs13245052>
33. Chen T, Hu N, Niu R, Zhen N, Plaza A (2020) Object-oriented open-pit mine mapping using gaofen-2 satellite image and convolutional neural network, for the Yuzhou City, China. *Remote Sens* 12:3895
34. Camalan S, Cui K, Pauca VP, Alqahtani S, Silman M, Chan R, Plemmons RJ, Dethier EN, Fernandez LE, Lutz DA (2022) Change detection of amazonian alluvial gold mining using deep learning and sentinel-2 imagery. *Remote Sens* 14:1746. <https://doi.org/10.3390/rs14071746>
35. Liu FF, Han HT, Zhang M, Ma LW (2021) Research on automatic extraction method of open-pit mine based on deep learning and remote sensing images. *Chn Energ Environ Prot* 43(06):82–85. <https://doi.org/10.19389/j.cnki.1003-0506.2021.06.014>
36. Hu NX, Chen T, Zhen N, Niu RQ (2021) Object-oriented open pit extraction based on convolutional neural network. *Remote Sens Technol App* 36(02):265–274. <https://doi.org/10.11873/j.issn.1004-0323.2021.2.0265>
37. Cheng GX, Niu RQ, Zhang KX, Zhao LR (2018) Opencast mining area recognition in high-resolution remote sensing images using convolutional neural networks. *Earth Sci* 43(S2):256–262. <https://doi.org/10.3799/dqkx.2018.987>
38. Zhang FJ, Wu YL, Yao XD, Liang ZY (2020) Opencast mining area intelligent extraction method for multi-source remote sensing image based on improved densenet. *Remote Sens Technol Appl* 35(3):673–684. <https://doi.org/10.11873/j.issn.1004-0323.2020.3.0673>
39. Yuan, D.N.; Jia, N.F. (2015) Proceedings of cross-strait symposium on soil and water conservation 2015, China, Shanxi, ; pp. 576–580.
40. Jiang J, Li J, Guo D, Chai M (2016) Application of remote sensing technology for monitoring the environmental footprint of Ningwu Xuejiawa mine. *Huabei Nat Resour* 04:17–19. <https://doi.org/10.3969/j.issn.1672-7487.2016.04.005>
41. Zhang GG, Wu J, Yi Y, Wang ZQ, Sun HX (2019) Traffic sign recognition based on ensemble convolutional neural network. *J Chongqing Univ Posts Telecommu (Natural Science Edition)* 31(04):571–577. <https://doi.org/10.3979/j.issn.1673-825X.2019.04.019>
42. Sun, H.R. (2020) Research on image compression algorithm based on deep learning. Shanghai Jiao Tong University, Shanghai, <https://doi.org/10.27307/d.cnki.gsjtu.2020.002698>
43. H. Gupta, A. Vahid Dastjerdi, S. K. Ghosh, and R. Buyya, "ifogsim: A toolkit for modeling and simulation of resource management techniques in the internet of things, edge and fog computing environments," *Software: Practice and Experience*, vol. 47, no. 9, pp. 1275–1296, 2017.
44. Taleb T, Ksentini A, Frangoudis PA (2016) Follow-me cloud: When cloud services follow mobile users. *IEEE Trans Cloud Comput* 7(2):369–382

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
