RESEARCH　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

Image data model optimization method based on cloud computing

Jingyu Liu[1], Jing Wu[2], Linan Sun[3*] and Hailong Zhu[1]

Abstract

In the current age of data explosion, the amount of data has reached incredible proportions. Digital image data constitute most of these data. With the development of science and technology, the demand for networked work and life continues to grow. Cloud computing technology plays an increasingly important role in life and work. This paper studies the optimization methods for cloud computing image data recognition models. The parallelization and task scheduling of the remote-sensing image classification model SCSRC based on spatial correlation regularization and sparse representation are studied in a cloud computing platform. First, cloud detection technology, combined with the dynamic features of the edge overlap region, is implemented in cloud computing mode. For image edge overlap region detection, the SCSRC method is implemented on a single machine, and the time performance of the method is analysed experimentally, which provides a basis for parallelization research under the cloud computing platform. Finally, the speedup and expansion ratio of the SK-SCSRC algorithm are determined by experiment, and MR-SCSRC and SK-SCSRC are compared. The simulation results show that, compared to previous methods, the method of image edge overlap detection is more accurate and the image fusion is better, which improves the image recognition ability in the overlap region and demonstrates the performance improvement of the MR-SCSRC algorithm under scheduling. This method addresses the shortcomings of Hadoop's existing scheduler and can be integrated into remote-sensing cloud computing systems in the future.

Keywords: Cloud computing, Data model, Image processing, Optimization method

Introduction

Proposing a new concept is usually a gradual process, and the concept of cloud computing is no exception. With the development of information technology and the rapid development of image data collection technology, various industries generate a large amount of multimedia data every day, and most of these data come from digital image data. Faced with the explosive growth of digital image data, traditional stand-alone image processing faces many problems, such as low processing speeds and poor concurrency [1]. Therefore, the traditional image processing mode cannot evolve to meet the needs of users, and it is necessary to find a new effective image processing mode [2]. Aminsoofi provided a general

reference to cover a wide range of issues related to data security in cloud computing, from accountability to data sources, identification and risk management [3]. Barsoum, A. F., later proposed that an increasing number of organizations will want to transfer data to third-party cloud computing service providers [4]. Customers can rent a CSP storage infrastructure to store and retrieve virtually unlimited amounts of data and then pay according to GB/month. To improve scalability, availability and durability, some customers may wish to copy their data to multiple servers in the centre [5]. The more copies are to be stored in the CSP, the more fees the customer will be charged, so the customer must be sure that the CSP will retain all copies of the data agreed upon in the service agreement, and all of these copies must correspond to the latest changes posted by the customer. In this paper, we present a map-based ownership scheme for multiple copies of dynamic data that has the

* Correspondence: rgzn_2019@163.com
[3]Heihe University, Heihe 164300, Heilongjiang, China
Full list of author information is available at the end of the article

following characteristics: the CSP does not cheat by providing fewer copies to the client, dynamic data outsourcing is supported, and block-level operations, such as modifying and inserting blocks, are supported.

Cloud computing is an Internet-based computing model with wide participation, in which computing resources (computing, storage, and interaction) are dynamic, scalable, and virtualized services. Cloud computing is a form of distributed computing that is related to breaking down huge data-processing programs into countless small programs through the network "cloud" and then processing and analysing these small programs through a system containing multiple servers. The result is returned to the user [6, 7]. In short, in the early days of cloud computing, simple distributed computing was performed, the distribution problem was solved and the calculation results were combined, so cloud computing is also called grid computing. Thanks to this technology, thousands of data points can be processed in a short time (a few seconds) to provide powerful network services [8]. In cloud computing, the model method for optimizing massive image data can be accurate and reasonable. Cloud computing can extract effective and valuable information and make this information reliable and convincing. By using intelligent optimization algorithms to solve image data model problems with big data and cloud computing, the method of processing image data models in big data and cloud computing can provide new and complex massive data picture theory and assistance tools. The main importance of cloud computing is its consistency; that is, cloud computing is highly scalable, and it is needed to provide users with a whole new experience. The core of cloud computing is the coordination of many computer resources. Users can receive unlimited resources through the network, and resources received at the same time are not limited by geographical or temporal constraints. This simplifies the modification of image processing software code and facilitates various image processing methods [9]. The distributed storage and distributed computing of the Hadoop distributed platform combines the advantages of high reliability and scalability with the parallel processing of large images, allowing users to quickly process large images and assign them to heterogeneous clusters. The task assignment optimization strategy based on the genetic algorithm proves that the optimization method is feasible, reduces the processing time, and significantly improves the image processing efficiency [10].

Currently, cloud services are not only a form of distributed computing but also include hybrid computing and computing technologies such as distributed computing, service computing, load balancing, parallel computing, network storage, hot backup, and virtualization. In fact, before the advent of the Internet, people realized that it was theoretically possible to create computer-controlled resources such as electricity in large public "power plants" and transmit them to various places through a network. When computer scientists had just begun thinking about how to talk to a computer, Internet interconnection experts predicted: "The future of computing can become public, just as the phone system has become public." When creating the World Wide Web, a huge, publicly accessible online data warehouse was used instead. All private online data storage and offline data storage seem very closely related, and many application service providers are expected to transfer software to enterprises over the Internet. However, this does not happen very quickly. The biggest obstacles appear in the network. The complex software that can replace software on a hard disk drive needs the ability to transfer large amounts of data quickly. This was impossible in the traditional low-speed dialling era. The initialization of the software would quickly overload the phone line or damage the modem, and the PC would shut down, but even in this case, the early Internet still provided a prototype of cloud computing [11]. The emergence of fibre-optic cables and the fibre-optic Internet eliminated the bottleneck of data transmission. The importance of network space ultimately exceeded the importance of computer memory. The role of the fibre-optic Internet in computer applications is similar to the role of AC systems in electricity, which made the device location no longer important to users. Since data can be transmitted over the Internet at the speed of light, it is possible to provide a remote user with the full computing power of a computer. The server computer for the software used by the operator can be located in the data centre near the user or in a data centre somewhere across the country [12]. In addition, the fibre-optic Internet also acts as a rotary converter, allowing a user to connect various devices to the network to generate electricity so that incompatible computers can work together as a system. The optical Internet has spurred the formation of central computing plants by providing a common medium for transmitting data directly or indirectly. In addition to the powerful capabilities of the fibre-optic Internet, the expansion of virtualization technology has provided the necessary impetus for the development of public computing. Virtualization refers to the use of hardware to model software, and because there are many components of a computer system (from the microprocessor and storage device to network devices such as routers and firewalls), the improvement of virtualization capabilities and the explosive growth of computer chip performance are inseparable. In addition, all equalizers work in digital mode, so software can be used to replace or virtualize hardware. Building a variety of hardware and software tools into business

applications may sound like science fiction, but it is quickly becoming a reality. With the development of cloud computing services, anyone can easily use rich software services on the Internet, use unlimited online storage and access and share data through a variety of devices, such as mobile phones and TVs. Personal computers may become antiques in the near future that remind people of an era in which everyone was an amateur technician.

The parallelization and task scheduling of the remote-sensing image classification model SCSRC based on spatial correlation regularization and sparse representation are studied on a cloud computing platform. First, cloud detection technology combined with the dynamic features of the edge overlap region is implemented in cloud-computing mode. Image edge overlap region detection is performed. The SCSRC method is implemented on a single machine, and the time performance of the method is analysed experimentally, which provides a basis for parallelization research under the cloud computing platform. Finally, the speedup and expansion ratio of the SK-SCSRC algorithm are given by experiment, and MR-SCSRC and SK-SCSRC are compared. The simulation results show that, compared to other methods, the method of image edge overlap detection in MR-SCSRC is more accurate and the image fusion is better, which shows the improvement of the image recognition ability in the overlap region and verifies the performance improvement of the MR-SCSRC algorithm under scheduling. It addresses the shortcomings of Hadoop's existing scheduler and can be integrated into remote sensing cloud computing systems in the future.

Proposed method
Cloud computing image data model
(1) Cloud computing.

The "cloud" of cloud computing stems from the practice of drawing Internet maps, which are described as clouds. The so-called "public cloud" model for remotely executing workloads over the Internet in a data centre of a commercial provider is a model for adding, using, and delivering Internet services that typically involves providing dynamic and scalable information on the server. In terms of content, the Internet is usually a virtualized resource [13, 14]. Cloud computing is an information technology (IT) model that provides instant access to shared, customizable system resource pools and more advanced services that are typically available over the Internet with minimal administrative work. Similar to business computing, cloud computing is based on shared resources to achieve consistency and economies of scale. In addition, cloud computing can be more accurately described as virtualization and centralized data centre resource management [15].

(2) Cloud computing classification.

Depending on the type of service, cloud computing is generally divided into three categories: infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). It can be said that the private cloud is a small public IaaS cloud that allows software to be deployed and run in the customer data centre. Like a public cloud, internal clients can provide their own virtual resources to create, test and run applications and can charge according to the cost of resource consumption; at the same time, a cloud computing service has the following features: on-demand self-service, access to any network device anytime and anywhere, a resource pool shared by several people, rapid redeployment, and services that can be monitored and measured. Rapid deployment, based on virtualization technology, resources or access to services, reduces the burden on the user's terminal and reduces the user's reliance on IT resources [16, 17].

(3)The status quo of cloud computing

The Elastic Compute Service (ECS) is an IaaS-class cloud computing service provided by the Alibaba Cloud. It has excellent performance, stability and reliability, and flexible expansion capabilities. Cloud Server ECS eliminates the need for pre-production of IT equipment, allowing the server to be used as easily and efficiently as public resources such as water, electricity and gas to provide off-the-shelf computing resources [18]. The Alibaba Cloud ECS continues to provide innovative servers to meet a variety of business needs and help users grow their businesses. Compared to conventional IDC rooms and server providers, the Alibaba Cloud uses stricter IDC standards, server access standards and operating and maintenance standards to ensure that cloud servers are used in a way that promotes the high availability and data reliability of cloud computing. Based on this, each area provided by the Alibaba Cloud has several Availability Zones. When higher availability is required, an Alibaba Cloud Multiple Access Zone can be used to create primary and backup services or real-time services. For two- and three-centre solutions in the financial sector, services with better availability can be created in multiple regions and multiple access areas. For disaster recovery, backup and other services, the Alibaba Cloud has a very mature solution. The Alibaba Cloud provides users with three types of support: there are products and services that increase availability, including cloud servers, load balancing, multi-backup database services and DTS data migration services; industry partners and environmental partners can help users create more stable architectures; and to ensure continuity of service and extensive training services, users can obtain high

availability across the entire communication channel from the enterprise to the underlying infrastructure server.

(4)The positive role of cloud computing

In today's highly competitive industry market, cloud computing an excellent opportunity for not only innovation but also faster and more cost-effective business operations than ever before. The cloud is a very effective way to provide IT services. Because users can create new virtual servers in the cloud with unprecedented speed and consistency and can automatically allocate resources such as computing power and storage to IT services, cloud computing can make new services work faster than traditional architectures. In addition, since cloud computing operations are paid for based on actual usage, operating costs are reduced compared to the costs of traditional storage; capital costs are especially reduced, which greatly reduces the risk of introducing new services into the organization, even if the services provided are unsuccessful. On the other hand, if the service usage is high, the cost will be reasonable, and the end user will incur the expected cost; for external services, revenue will be significantly increased. In addition, the benefits of cloud computing are ideal for a variety of services, and cloud computing can create new values of many different types. It was almost impossible to implement the concept of cloud computing a few years ago. Today, for many organizations, using cloud computing has become standard operating procedure. In cloud computing, virtual servers are created to meet business needs, which means that they are used daily for internal and external services. A new virtual server can replicate any or all production servers, and this can be done in the cloud in a short time. Enterprises can now use the cloud computing platform to replicate some or all of their data centre functionality in at least two different ways: they can create third-party cloud policies, create new virtual servers when necessary, and then use the appropriate software stack and data; and they can perform recovery services. To respond more quickly to crashes, they can create virtual servers in advance and have them initiate a hot backup site at any time. The site can perform all the necessary preliminary configuration of the software, and before providing full recovery services, the latest data can simply be transferred. In either case, the impact on a business is huge. For example, business continuity is significantly improved because the organization no longer relies on a single-service delivery infrastructure and no longer "puts all its eggs in one basket" if the service fails; this supports long-term thinking. During the assessment and correction process, the organization may temporarily switch to cloud backup infrastructure. This feature can greatly reduce the negative impact on the business; otherwise, it may continue until

the problem in the data centre is resolved. Cloud backup provides a recovery tool that can be easily processed and prepared regardless of the situation. Especially for small and medium-sized companies with small IT budgets, cloud computing is far superior to any other choice, and it eliminates the cost and management complexity associated with traditional backup solutions that require redundant hardware and software infrastructure. Cloud computing simplifies processes and significantly reduces capital costs [19, 20].

Image data model optimization method

(1)Image data model

A data model is an abstraction of data functions; data is a symbolic representation of a thing, and a model is an abstraction of the real world. It abstracts the static characteristics of the system—its dynamic behaviour and limitations—and provides an abstract basis for representing information and the operation of the database system. The data model comprises three parts: the data structure, data manipulation, and data constraints. China has created a series of data models to assess the potential of mineral resources. Three basic types of data models have been used: hierarchical data models, grid data models and relational data models. The layered model was the earliest development: its basic structure is a tree structure, and a typical representative is the IMS model. Since the connections among data in most practical tasks do not have a pure tree structure, the hierarchical data model was gradually eliminated; the grid data model expresses the relationships among data with a network structure. It developed earlier and has certain advantages. It is still used more frequently. A typical representative is the DBTG model. The relational model was developed later; it includes a set of entities that represent the relationships among data through a two-dimensional table that satisfies specific conditions. It has a solid mathematical foundation and theoretical foundation, and it is flexible and easy to use. It also has a wide range of adaptation possibilities.

(2) Data model content classification.

The content described in the data model consists of three parts: a data structure, data manipulation, and data restrictions. The data structure mainly describes the type, content, nature and relationship of data, and it is a set of target types. Target types are components of the database and can usually be classified into two categories: relationships between data types and data types themselves, such as record types, data elements, relationships, and domains. In the database task group (DBTG) grid model, some of the relationships are also used in the relational model. The data structure is the basis of the data model, and the operations and limitations of the data are mainly based on the data structure.

Different data structures have different operations and restrictions. Data processing in the data model mainly involves the type and operation mode of the operations and the corresponding data structure. The structure is a set of operators, including multiple operations and inference rules that work with a database of valid instances of the target type. The data constraints in the data model primarily describe the syntax, word meanings, and constraints and dependencies among data structures. Rules for dynamically changing data to ensure correct, efficient, and compatible data are also included. There is a set of integrity rules that determine the state of the database and state changes that correspond to the data model. The constraints can be divided into constraints on data values and constraints on data connections according to various principles: static and dynamic constraints, entity constraints, and reference constraints between entities.

Image data model optimization method for cloud computing

An example algorithm is given below.

Definition one: uncertain reasoning in the cloud.

In practical applications, a set of single-conditional rules can be formally expressed as:

$$If A, then B, \cdot i = 1, 2, 3, \cdots, m \tag{1}$$

Similarly, multiple rules and multiple conditions can be formally expressed as:

$$If \ A_1 A_2, \cdots, A_n, then \ B_1 \tag{2}$$

Definition two: principles of cloud computing

1) Calculation logic

If F is a calculation and $\times 1, \times 2, ..., xn$ are the n calculated parameter variables, then F is called an n-ary calculation, and S is the result of the calculation. This can be written as:

$$S = F(\chi_1, \chi_2, \cdots, \chi_N) \tag{3}$$

If $a_1, a_2, ..., a_n$ are the values of the n parameter variables $x_1, x_2, ..., x_n$, then:

$$S = F(a_1, a_2, \cdots, a_n) \tag{4}$$

If $A_1, A_2, ..., A_n$ are n conditions and B is the conclusion, then a rule called an n rule, denoted as R, can be expressed as:

$$If \ A_1 \ A_2, \cdots, A_n, then B \tag{5}$$

2) Computational logic transformation

Given a calculation F and an eq. $S=F(a_1,a_2,...,a_n)$, an n rule R can be generated, which is expressed as:

$$If \ A_1 A_2, \cdots, A_n, then \ B_1, B_2, \cdots, B_n \tag{6}$$

Definition three: computing clouding process.
1) For the n calculation F, if its domain is:

$$\Omega = (U_1 \times U_2 \times \cdots \times U_n) \tag{7}$$

then m parameter values can be extracted as Ω sample parameter values, and the calculation of F can be performed. The result is Si, where:

$$S_i = F(a_{1i}, a_{2i}, \cdots, a_{ni}) \ i = 1, 2, 3, \cdots, m \tag{8}$$

By the conversion theorem:

$$If \ a_{1i}, a_{2i}, \cdots, a_{ni}, then \ S_i \ i = 1, 2, 3, \cdots, m \tag{9}$$

3) Clouding

Once a numerical variable has been clouded by a clouding process P, it is assumed that process A constructs a set A of qualitative concepts corresponding to U, which contains K qualitative concepts, namely:

$$A = \{A_1, A_2, \cdots, A_K\} \tag{10}$$

Experiments

Experimental data set

This experiment used four datasets, where Data 1 and Data 2 are the Indian Pines dataset and the University of Pavia dataset, respectively; the Mosaic software feature of ENVI software was used to mosaic the Indian Pines dataset up and down to 29,000. Data of size 145, 58000 × 145, were used as Data 3 and Data 4. The size of each data set is shown in Table 1.

The experiments were tested under the Hadoop and Spark cloud platforms, and three types of experiments were performed:

1) an acceleration ratio test was performed based on the parallel SK_SCSRC algorithm of the Spark platform;
2) a parallel SK_SCSRC algorithm expansion ratio test was performed based on the Spark platform;

Table 1 Data sets used in the experiment

Data set	Data size
Data 1	128*128 (8 MB)
Data 2	640*320 (40 MB)
Data 3	30000*155 (2.03 GB)
Data 4	61000*155 (2.5 GB)

3) the execution efficiency of the MR_SCSRC algorithm under the Hadoop platform and the SK_SCSRC algorithm under the Spark platform were compared.

Experimental environment

To verify the correctness of the classification method of SCSRC, the code of the SCSRC algorithm is written in Java. The hardware environment of the experiment is the Intel Xeon E7–4807 with 8.0 G memory and a 500 GB hard disk. The parameters used in the experiment, $\lambda_1 = 10^{-3}$, $\lambda = 2$, $\mu = 1$, obtained classification results for the algorithm that did not change significantly over many experiments to avoid reducing the performance of the algorithm. The SCSRC algorithm includes many matrix operations, and a matrix library was used to implement the SCSRC algorithm.

The common matrix calculation libraries in Java are JAMA, UJMP and JPLAS.

1) JAMA is a basic linear algebra Java package that provides basic matrix operation functions.
2) UJMP encapsulates common Java matrix libraries into different modules, and different matrix libraries can be chosen when performing matrix operations. Moreover, it also provides a visualization function for the matrix.
3) JPLAS is a Java linear algebra library based on BLAS and LAPACK. It can call native libraries through JNI, so it has a faster operating speed.

Implementation of the algorithm

The algorithm used is a recent regular subspace classification algorithm based on MapReduce, a parallel programming framework, and it includes three parts:

(1) Split: In the Hadoop architecture, the input test data set is divided into multiple serialized splits. This operation is performed by calling getSplit. The larger the input file, the more splits the input is divided into. Parallel processing of splits can reduce the processing time of massive data files. The RecordReader function parses each split into key-value pairs. A split shard can be parsed into multiple key-value pairs and passed as input data to the Map function for processing. A split shard corresponds to a map process, and it also corresponds to a shard. The default block size is 64 M. To balance the machine load of the entire platform and improve the utilization of the cluster, when the amount of input data to be processed is small, the data should be divided into as many shards as possible. Through the formula below, the fragment size is determined according to goalSize and minSize, both of which are determined by the configuration file they are the number of map tasks started and the minimum fragment size, respectively. Let D be the size of the input data set and S be the number of shards.

(2) Map: The input of the Map function is in the form of <key, value>. The key represents the pixel offset, and the value represents a piece of test data. The final result is <key', value'>, where key' is consistent with key, and value' is represented as (Euclidean distance, class number). In the Map phase, the NRS classifier is used to calculate the Euclidean distance between each test sample data point and each type of reference data; the sample offset, the compared category, and the corresponding Euclidean distance are recorded so that the test data can be processed during the Reduce phase. The final classification operation, the Map stage, includes a variety of matrix operations, including matrix inversion, matrix multiplication, and diagonalization of matrices. For these operations, corresponding functions are designed. Because this is a large matrix operation, it is time consuming.

(3) Reduce. In the Reduce phase, the input key-value pairs are the output of the Map phase, and they have the form <data offset, Euclidean distance, class number>. According to the data offset and the length of the test data, the serial numbers of the test data can be obtained. For key-value pairs with the same data offset, the Euclidean distances of the values are compared to find the minimum Euclidean distance. According to the category number corresponding to the smallest Euclidean distance, the final class label is added to the test data, and the classification result is finally stored on the distributed file system. The output key is of the form (test sample serial number, class number), and the output value is set to empty. Then, other operations are performed on the hyperspectral image data based on the classification results.

Discussion
Performance comparison of image edge detection based on cloud computing

To test the application performance of this method in the detection of image edge overlap areas, a simulation experiment is carried out, and the matching coefficient of image frame feature points is 1.25. This method can effectively detect image edge overlap areas based on cloud computing; the information enhancement performance of the detection output is good, and the accuracy of image detection by different methods is tested. The comparison results are shown in Fig. 1, and the analysis is performed. The method in this paper has a high accuracy in detecting the overlapping area of image edges based on cloud computing.

Fig. 1 Test performance comparison
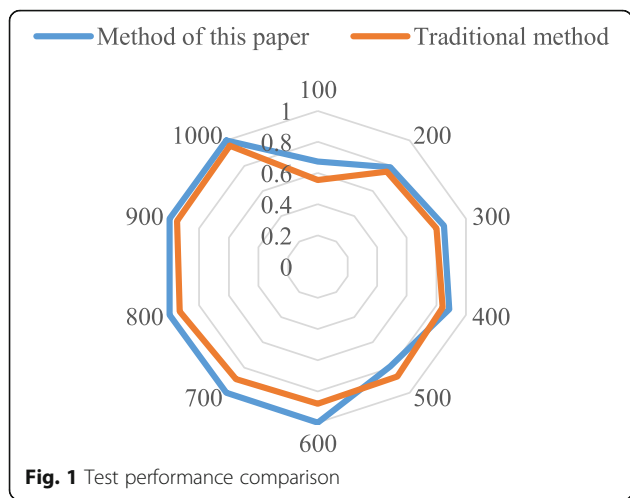
Parallelization of the SCSRC algorithm on the cloud platform

(1) Acceleration ratio test of the SK UU SCSRC algorithm.

Speedup refers to the performance improvement after the serial program is parallelized. It is an important index to measure the performance of parallel computing, and it is shown in formula (11):

$$sp = \frac{T_1}{T_P}. \tag{11}$$

T1 is the time consumed by the calculation of the serial program (the degree of parallelization is 1), and TP is the time consumed by the calculation after the parallelization of the program (the degree of parallelization is p). First, taking data set Datal as input, we run the SCSRC serial program based on CPU (the number of partitions is 1) and the SK_SCSRC parallel program based on Spark under the same execution environment. Spark divides the input data into multiple partitions (the partitions here are similar to the partitions in map reduction but have different meanings to the partitions in the reduction phase of map reduction). It starts a thread task for each partition to process the partition data, so the number of partitions is the parallelization of the spark parallel process (to ensure that there are enough resources so that all tasks can be

carried out at the same time). Table 2 shows the operation results of the SCSRC program and the SK_SCSRC program. The results are the average values after each program is repeatedly executed 10 times. Figure 2 shows the speedup ratio after parallel processing of Datal.

(2) Speed ratio performance test for different data sets.

There is a certain gap between the actual acceleration ratio and the ideal acceleration ratio, because Spark starts a thread for each partition to process the data of the partition. In this experiment, the amount of data in each partition is small, and the network communication cost incurred by collecting data from each partition and the system cost of creating threads account for a certain proportion of the calculation time of the partition, so there is a certain gap between the actual acceleration ratio and the ideal acceleration ratio. The above experiments are repeated with datasets Data 2, Data 3 and Data 4 to test the acceleration performance. The results are shown in Fig. 3.

As the amount of data increases, the acceleration effect becomes increasingly obvious. The larger the amount of data, the closer the actual acceleration ratio is to the ideal acceleration ratio, and the larger the number of partitions is, the better the acceleration effect. However, when the number of partitions increases to a certain value, the acceleration ratio no longer increases, and this is reflected as a smooth curve on the graph. If the number of partitions continues to increase, the acceleration ratio will decline because after this point, the overhead of network transmission will gradually increase.

### Performance analysis of image data model optimization based on cloud computing

The MR_SCSRC algorithm is executed under the FIFO scheduler and StaticTaskScheduler scheduler, and Data 2 is set as the algorithm input to obtain the average time of each iteration of the algorithm, as shown in Fig. 4. The performance of the MR_SCSRC algorithm is improved under the scheduling strategy of StaticGA-TaskScheduler, which is 40% faster than FIFO.

Conclusions

(1) The concept of cloud computing has developed rapidly since its birth and has changed tremendously. In today's highly competitive industry market, cloud

Table 2 Data set processing results with different numbers of partitions

Number of partitions	OA	AA	Kappa	Total time	Speedup ratio
1	0.9756	0.9490	0.9737	6461 s·	1.1
3	0.9756	0.9490	0.9737	4544.13 s	1.45
6	0.9756	0.9490	0.9737	2562.54 s	2.67
9	0.9756	0.9490	0.9737	1715.30 s	3.56
12	0.9756	0.9490	0.9737	1329.26 s	4.35

Fig. 2 Data 1 acceleration ratio performance test
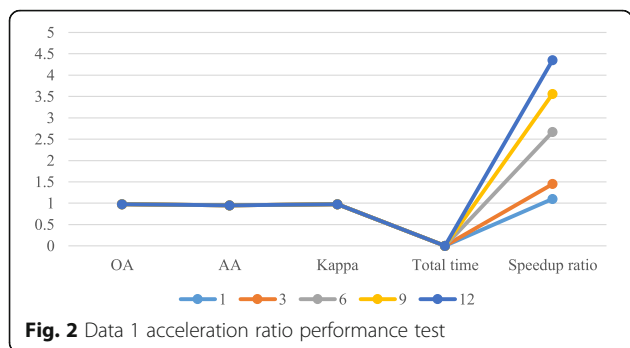
computing is not only an excellent opportunity for innovation but also an excellent opportunity for faster, more cost-effective business operations than ever before. Thus, cloud computing can make new services work faster than traditional architectures can. Since the cost of cloud computing operations is based on actual usage, operating costs are reduced, especially capital costs, which greatly facilitates the introduction of cloud computing in organizations. The financial risk of adopting new services, even if the services provided are unsuccessful, is much lower. In addition, the benefits of cloud computing are well suited to several services, and cloud computing can create many new types of value. Compared with traditional web application models, cloud computing has the advantages of high flexibility, scalability and popularity, as well as virtualization technology, dynamic scalability and high reliability. It loosens temporal and spatial constraints and brings effective computing power. Adding cloud computing capabilities based on a source server speeds up calculations, ultimately enabling the dynamic expansion of the virtualization layer to extend an application's goals, and cloud computing can quickly provide computing power and resources based on user needs. It is compatible with many applications, and it can not only be integrated with low-profile machines and hardware products from different manufacturers but also enable higher-performance calculations with peripheral devices.

Fig. 3 Speedup performance test on different datasets

Fig. 4 Running time under two schedulers
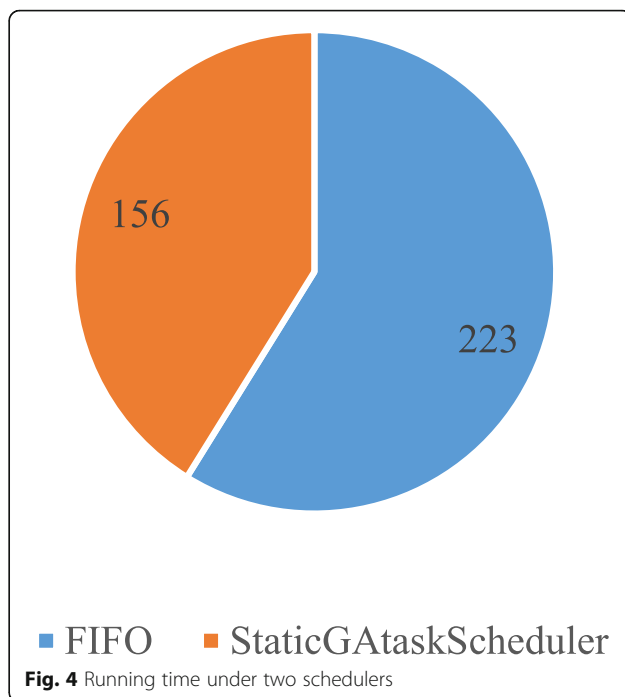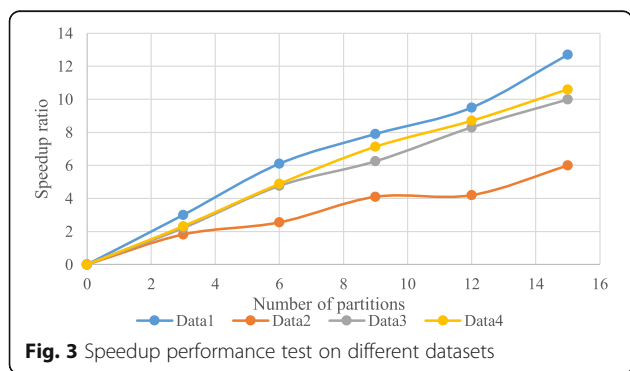
(2) There are many optimization methods for the image data model. Each algorithm has its own application domain and application conditions as well as its advantages and disadvantages. The characteristics of various algorithms, such as genetic algorithms, evolutionary algorithms, particle swarm optimization algorithms, and single-purpose or multifunctional intelligent optimization algorithms such as ant colony algorithms, can be used to optimize the image data model algorithm. These intelligent optimization algorithms are applied to the data model algorithm to improve the efficiency of the algorithm, expand its scope, and eliminate the flaws and drawbacks of the algorithm. When traditional data algorithms process large amounts of data, they cannot perform tasks efficiently and quickly. The data model optimization method of cloud computing improves the processing power of big data. The data are improved, and slight noise and variations are added to the training data. On the one hand, the training data can be augmented to improve the generalization ability of the model. On the other hand, the noise data can be increased, thereby increasing the reliability of the model. Using the cloud computing big-data model optimization method, we can create three-dimensional traffic and completely solve traffic problems in urban development. Cloud Medical is used in medical information platforms, remote treatment diagnosis and consultation. The image data model optimization method uses methods such as image flipping, image cutting, and image whitening to increase the amount of data and improve fit of the model. Model enhancement techniques prevent re-

equipment and improve the generalization capabilities of the model. The optimization method changes the learning speed in the learning process so that the model can converge better, the ability of model fitting is strengthened, the network level and remaining network deepen, and finally the accuracy of classification is improved. The image data model optimization method can also be used to predict the mineralization of mineral resources and to assess coal, uranium, and chemical mineral potential, thereby eliminating the cost and management complexity required in traditional backup solutions. With redundant hardware and software infrastructure, cloud computing simplifies processes and significantly reduces capital costs.

(3) Research is carried out on the parallelization of the hyperspectral remote-sensing image classification algorithm under the cloud computing platform, and the global research status is summarized. The cloud computing platforms Hadoop and Spark are studied as well as the HDFS component. Map reduction is carried out on the programming model and Spark RDD programming model. The generation background of HDFS is analysed. Map reduction and Spark RDD technology are examined, focusing on the implementation mechanism of map reduction tasks and comparing maps. The differences between the reduced programming model and the Spark RDD programming model are studied.

Abbreviations
PC: Personal Computer; AC: Alternating Current; GSP: Completely-self-Protected; DBTG: Database Task Group; BLAS: Basic Linear Algebra Subprograms

About the authors
Jingyu Liu was born in Heilongjiang province in 1982.She got her doctor's degree from Harbin engineering university. Now I work in Harbin normal university. Research interests include artificial intelligence and pattern recognition, image processing, cloud computing, information security, etc.
E-mail: liujingyu926@163.com
Jing Wu was born in Zhengzhou,Henan,P.R.China,in 1979. She received the Master degree from Huazhong University of science and teachnology, P.R. China. Now, she works in ZhengZhou Preschool education college, her research interests include Artificial Intelligence, Data, Image Processing, Communication and Information Security.
E-mail: Szmt_2019@163.com
Linan Sun was born in Jiamusi, Heilongjiang, P.R. China, in 1983. She received the Master degree from Harbin Normal University, P.R. China. Now, she works in Heihe University, Her research interests include pattern recognition and artificial intelligence, cloud computing, image processing, big data analysis, statistical analysis.
E-mail: rgzn_2019@163.com
Hailong Zhu was born in heilongjiang province in 1972.He received his doctor's degree from Harbin engineering university. Now work in Harbin normal university. Research interests include artificial intelligence and pattern recognition, image processing, cloud computing, information security, etc.
E-mail: 15136502144@163.om

Authors' contributions
All authors take part in the discussion of the work described in this paper. The author(s) read and approved the final manuscript.

Authors' information
Jingyu Liu was born in Heilongjiang province in 1982.She got her doctor's degree from Harbin engineering university. Now I work in Harbin normal university. Research interests include artificial intelligence and pattern recognition, image processing, cloud computing, information security, etc.
Jing Wu was born in Zhengzhou, Henan, P.R. China, in 1979. She received the Master degree from Huazhong University of science and teachnology, P.R. China. Now, she works in ZhengZhou Preschool education college, her research interests include Artificial Intelligence, Data, Image Processing, Communication and Information Security.
Linan Sun was born in Jiamusi, Heilongjiang, P.R. China, in 1983. She received the Master degree from Harbin Normal University, P.R. China. Now, she works in Heihe University, Her research interests include pattern recognition and artificial intelligence, cloud computing, image processing, big data analysis, statistical analysis.
Hailong Zhu was born in heilongjiang province in 1972.He received his doctor's degree from Harbin engineering university. Now work in Harbin normal university. Research interests include artificial intelligence and pattern recognition, image processing, cloud computing, information security, etc.

Availability of data and materials
All the data and materials in this article are real and available.

Competing interests
These no potential competing interests in our paper. And all authors have seen the manuscript and approved to submit to your journal. We confirm that the content of the manuscript has not been published or submitted for publication elsewhere.

Author details
[1]Harbin Normal University, Harbin 150000, Heilongjiang, China. [2]ZhengZhou Preschool education college, Zhengzhou 450000, Henan, China. [3]Heihe University, Heihe 164300, Heilongjiang, China.

References
1. Williams E, Moore J, WI LS (2017) Image Data Resource: a bioimage data integration and publication platform. Nat Methods 14(8):775
2. Fan N, Wang Y, Lv Y (2017) Improved chirp scaling algorithm for processing squinted mode synthetic aperture sonar data. Cybern Inf Technol 16(6):111–122
3. Aminsoofi A, Irfan Khan M, Fazaleamin FA (2017) A review on data security in cloud computing. Int J Comput Appl 96(2):95–96
4. Barsoum AF, Hasan MA (2015) Provable multicopy dynamic data possession in cloud computing systems. IEEE Trans Inf Forensics Secur 10(3):485–497
5. Rost P, Mannweiler C, Michalopoulos DS (2017) Network slicing to enable scalability and flexibility in 5G Mobile networks. IEEE Commun Mag 55(5): 72–79
6. Wang Q, Zhao Y, Lin F (2017) Correlation between the finite element calculation and experimental mode of a mechanical elastic wheel. J Harbin Eng Univ 38(1):86–93
7. Giesl J, Aschermann C, Brockschmidt M (2017) Analyzing program termination and complexity automatically with AProVE. J Autom Reason 58(1):3–31
8. Flocchini P, Prencipe G, Santoro N (2017) Distributed computing by mobile robots: uniform circle formation [J]. Distrib Comput 8878(6):1–45
9. Drake MS, Thornock JR, Twedt BJ (2017) The internet as an information intermediary [J]. Rev Acc Stud 22(2):543–576

10. Tavakkolimoghaddam R, Safari J, Sassani F (2017) Reliability optimization of series-parallel systems with a choice of redundancy strategies using a genetic algorithm. Reliab Eng Syst Saf 93(4):550–556
11. Mijumbi R, Serrat J, Gorricho JL (2017) Network function virtualization: state-of-the-art and research challenges. IEEE Commun Surv Tutorials 18(1):236–262
12. Han F, Reily B, Hoff W (2017) Space-time representation of people based on 3D skeletal data: a review. Comput Vision Image Underst 158(C):85–105
13. Cao S, Biesiada M, Jackson J (2017) Measuring the speed of light with ultra-compact radio quasars. J Cosmol Astropart Phys 2017(02):012–012
14. Yan H, Lu J, Zhou X (2017) Prototype-based discriminative feature learning for kinship verification. IEEE Trans Cybern 45(11):2535–2545
15. Jubb HC, Higueruelo AP, Ochoa-Montaño B (2017) Arpeggio: a web server for calculating and Visualising interatomic interactions in protein structures. J Mol Biol 429(3):365–371
16. Darton RA (2018) Modeling multigroup populations. J R Stat Soc 154(3):493–494
17. Duggan C, Gunn J (2018) Medium-term course of disaster victims: a naturalistic follow-up. Br J Psychiatry 167(2):228–232
18. Spiering VL, Bouwstra S, Spiering RMEJ (2017) On-chip decoupling zone for package-stress reduction. Sensors Actuators A Phys 39(2):982–985
19. Burton E, Goldsmith J, Mattei N (2018) How to teach computer ethics through science fiction. Commun ACM 61(8):54–64
20. Martin W, Sarro F, Yue J (2017) A survey of app store analysis for software engineering. IEEE Trans Softw Eng 43(9):817–847

Publisher's Note