Computers & Geosciences 135 (2020) 104403

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

A multi-task multi-class learning method for automatic identification of heavy minerals from river sand

Na Li [a], Huizhen Hao [a,b], Zhiwei Jiang [a], Feng Jiang [c], Ronghua Guo [d], Qing Gu [a,*], Xiumian Hu [d]

[a] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
[b] Department of Communication Engineering, Nanjing Institute of Technology, Nanjing, China
[c] School of Mobile Internet, Taizhou Institute of Sci. & Tech., NUST, Taizhou, China
[d] School of Earth Sciences and Engineering, Nanjing University, Nanjing, China

ARTICLE INFO

MSC:
00-01
99-00

Keywords:
Heavy mineral identification
Multi-task learning
Multi-class classification
Group sparsity
River sand

ABSTRACT

The identification of heavy minerals collected from river sand is an essential work in geology, which has great significance in the sedimentary provenance analysis, tectonic evolution and lithofacies paleogeography. Automatic identification of heavy minerals using classifiers trained by identified minerals has become a topical issue in applying machine learning techniques to geological research. Due to the high cost of manual identification and annotation, the available identified heavy minerals from a single river basin are insufficient for training a classifier with acceptable accuracy. Besides, since there are scores of heavy mineral classes in nature, the common "One vs. One" and "One vs. Rest" strategies used for multi-class classification will cause the problems of class imbalance and time overhead in the case of identifying heavy minerals. In this paper, we propose a Multi-Task Multi-Class learning method (MTMC), which leverages the correlation among heavy minerals collected from multiple river basins, and jointly handles these identification tasks, in order to deal with the problem of insufficient identified heavy minerals from a single river basin while improve the performance of the identification of heavy minerals in all the river basins. MTMC uses softmax as the classifier to handle the multi-class classification of heavy minerals, and formalizes the multi-task learning as a convex optimization problem. The experimental results based on heavy minerals collected from the Yangtse River, Yarlung Zangbo River and PumQu River demonstrate the effectiveness of MTMC, which evidently outperforms current machine learning methods for the identification of heavy minerals.

1. Introduction

The identification of heavy minerals (regarded as a classification task in machine learning) is an essential work in geology. The stable physical and chemical properties of heavy minerals, including those found in river basins, record and retain a large amount of sedimentary provenance information, such as chemical composition, geotectonics, climatic conditions and atmospheric environment. Based on these properties, geologists can infer the terrigenous source areas to which they belong and their parent rock components. Efficient and accurate identification of heavy minerals is of great significance for provenance analysis, tectonic evolution, stratigraphic correlation, and lithofacies paleogeography.

Traditionally, the identification of heavy minerals is done manually using a petrological microscope exploiting the optical properties of heavy minerals (Morton and Hallsworth, 1994), which is either time-consuming or subject to individual experience. The analysis of heavy minerals by inductively coupled plasma-atomic emission spectroscopy (ICP-AES) offers a faster alternative by discriminating key elements associated with specific heavy minerals (Mounteney et al., 2018). Raman spectroscopy is also used as a fundamental complement for heavy mineral analysis (Andò and Garzanti, 2014). For the semi-automatic approach by Raman spectroscopy (Lünsdorf et al., 2019), manual identification is still required for heavy mineral analysis. The electron probe microanalyzer (EPMA) is also an effective technique for the quantitative analysis of heavy minerals (Shimizu et al., 2019), but it is highly costive in both time and resources.

Recently, computer-aided methods using machine learning techniques have been applied to the automatic identification of heavy minerals, which can both reduce the workload of geologists and improve accuracy by incorporating knowledge from identified heavy minerals. Some popular methods take advantage of the optical properties of the minerals for the identification of heavy minerals (Borges and

* Corresponding author.
E-mail address: guq@nju.edu.cn (Q. Gu).

https://doi.org/10.1016/j.cageo.2019.104403
Received 13 July 2019; Received in revised form 28 November 2019; Accepted 12 December 2019
Available online 18 December 2019

de Aguiar, 2019; Yousefi et al., 2016; Aligholi et al., 2017). These methods are close to the process of manual identification, which is strongly recommended by the experts in Geology. However, some optical properties, such as pleochroism and extinction, which are important in manual identification, are still difficult to be represented by heuristically defined image features, which may compromise the prediction performance.

Instrumental analysis combined with machine learning techniques can be used to automatically identify the heavy minerals (Hao et al., 2019; Akkaş et al., 2015; Akkas et al., 2016). A Scanning Electron Microscope (SEM) equipped with an Energy Dispersive X-ray Spectrometer (EDS) can qualitatively and quantitatively analyze the mineral sample, which is a common way for instrumental analysis in the identification of heavy minerals. Ruisanchez et al. (1996) used SEM-EDS to analyze the chemical composition of minerals, and applied neural networks to identify 12 standard mineral classes. Akkaş et al. (2015), Akkas et al. (2016) used the decision tree and neural network to identify natural minerals and zeolite group minerals respectively, based on data obtained from SEM-EDS analysis. Gallagher and Deacon (2002) used the SEM-EDS data to identify 10 mineralogical classes by the multi-layer perceptron and Kohonen SOM network.

At present, the automatic identification of heavy minerals collected from river basins is challenging. On one hand, the number of available identified mineral samples collected from a single river basin is far from sufficient to train a useful classifier (identifier), which leads to poor identification performance. First, heavy minerals are rare in river sand, which makes it hard for collection. Second, the manual annotation of heavy mineral samples requires cumbersome work by professional geologists, which is both costly and less efficient (Akkaş et al., 2015). Third, the available identified samples of a single heavy mineral class are usually few, due to the scores of classes exist in nature. On the other hand, there are scores of heavy mineral classes, while the majority of machine learning methods are designed for binary classification. The common "One vs. One" and "One vs. Rest" strategies used for multi-class classification are not a good choice for the identification of heavy minerals. Specifically, the "One vs. One" strategy leads to a large number of classifiers, while the "One vs. Rest" strategy causes a great class imbalance which increases the risk of over-fitting. The more the heavy mineral classes are, the more serious are the above problems.

In this paper, we study the automatic identification of heavy minerals collected from river sand based on the SEM-EDS analysis. For the insufficiency of identified heavy minerals in a single river basin, we harness the intrinsic relationship of heavy minerals among multiple river basins, and use them simultaneously to obtain good performance for each river basin and each heavy mineral class. Based on this idea, we propose a Multi-Task Multi-Class learning method (MTMC), which regards the identification of heavy minerals in each individual river basin as a single classification task, and trains the classifiers jointly. MTMC captures the common features of heavy minerals shared among multiple river basins via the group-sparse matrix, as well as specific features of individual river basin via the elementwise-sparse matrix. MTMC uses softmax as the classifier for the multi-class classification of heavy minerals, and formalizes the multi-task learning as a convex optimization problem solved by the FISTA algorithm (Beck and Teboulle, 2009). Experiments are conducted based on heavy minerals from river sand collected from the Yangtse River, Yarlung Zangbo River and PumQu River. The results show that MTMC evidently outperforms the current machine learning methods for the identification of heavy minerals.

2. SEM-EDS based feature representation

SEM-EDS is a reliable tool for microchemical analyses of heavy minerals. SEM can provide high-resolution, high-magnification raster data of the target mineral. EDS enables qualitative and quantitative analyses of chemical elements consisted in the mineral. In the approach of SEM-EDS analysis, an incident focused electron beam generated by an electron gun excites the atoms on the surface of a mineral sample, which results in the emission of secondary electrons, backscattered electrons, and characteristic X-rays. Various detectors are applied to detect the particles produced from the mineral sample and their released energies, hence acquire the information that can describe the composition of the mineral sample. Based on the property that the characteristic X-rays emitted by different chemical elements have different energies, EDS counts the collected characteristic X-rays produced by electronic transitions between energy shells. A typical EDS datum is represented as the form of the X-ray spectrum.

In this paper, we acquire SEM-EDS data of the heavy minerals and use them to extract the input features for building the classifiers.

The heavy minerals used in this paper were collected from modern river sands of the Yangtse River, the Yarlung Zangbo River (Yarlung) and the PumQu River respectively in China. All heavy mineral grains were manually picked in random, and mounted in the epoxy resin and polished. Microscopic images were acquired for spot analysis. All mineral samples were carbon coated and the carbon element was removed prior to the analysis.

The heavy mineral samples were analyzed using Carl Zeiss Supra 55 field emission Scanning Electron Microscope equipped with Oxford Aztec X-Max 150 Energy Dispersive X-ray Spectrometer, at State Key Laboratory for Mineral Deposits Research in Nanjing University. The analytical conditions used were 15 kV accelerating voltage and 60 nA beam current, with counting time of 90 s. The working distance was 8.7 mm. More details about the sample preparation and SEM-EDS data acquisition can be found in Hao et al. (2019).

Based on the spectrum produced by the EDS detector, we used standardless quantitative analysis to obtain the approximate weight fractions of oxides, which were adjusted further by matrix correction and normalization. The standard intensity I_{std} in $k = I_{unk}/I_{std}$ is calculated physically. Since the introduction of standard references for each mineral will result in 5–6 times increase in testing time, we used the reserved reference materials that are provided by the testing instrument. Generally, the relative errors of the results from the SEM-EDS analysis will range from 1%–3% for the major element (i.e. the element with a mass concentration greater than 10 wt%), and 3%–5% for the minor element (i.e. the element with a mass concentration between 1 and 10 wt%).

We rank the oxides and use their weight fractions in percent as the input features. Since different classes of heavy minerals have variant chemical compositions, we take the union of oxides of all heavy minerals as the complete feature set, which consists of 55 oxides, e.g. Al_2O_3, CaO, FeO, SiO_2 and MgO. If a specific mineral does not contain certain oxides, the corresponding weight fractions are 0.

The example SEM-EDS features of 13 mineral samples from five heavy mineral classes collected from the three river basins are shown in Table 1, where "-" indicates that the heavy mineral does not contain the corresponding oxides. Only 15 oxides from the total 55 are listed, since all the rest have only "-" elements for the 13 mineral samples. Obviously, such a SEM-EDS feature matrix is sparse. From Table 1, the SEM-EDS features of heavy minerals present the following characteristics.

- Heavy minerals from the same class have common SEM-EDS features shared among the three river basins.

Specifically, due to the stability of the chemical properties of heavy minerals, if they are from the same heavy mineral class, although collected from different river basins, they usually have similar chemical compositions, presented as the shared features.

- Heavy minerals from the individual river basin may have specific features not shared in the mineral class.

Table 1
The examples of SEM-EDS features (partial) of 13 mineral samples from five heavy mineral classes collected from the three river basins. The mineral abbreviations include Mnz (Monazite), Bt (Biotite), Zrn (Zircon), Amp (Amphibole), Tur (Tourmaline).

		SEM-EDS features														
		Al_2O_3	CaO	Ce_2O_3	FeO	IrO_2	K_2O	La_2O_3	MgO	Na_2O	Nd_2O_3	P_2O_5	SiO_2	TiO_2	WO_3	ZrO_2
Mnz	Yarlung	–	–	27.78	–	–	–	13.83	–	–	10.72	29.75	–	–	–	–
	PumQu	–	–	27.54	–	2.02	–	13.71	–	–	10.70	28.71	–	–	–	–
Bt	Yarlung	15.27	–	–	19.59	–	9.65	–	12.21	–	–	–	41.33	3.77	–	–
	PumQu	20.36	–	–	21.09	–	5.31	–	8.21	–	–	–	37.63	–	–	–
Zrn	Yangtse	–	–	–	–	–	–	–	–	–	–	–	30.42	–	1.53	66.69
	Yarlung	–	–	–	–	–	–	–	–	–	–	–	31.08	–	–	67.74
	PumQu	–	–	–	–	–	–	–	–	–	–	–	30.24	–	1.06	66.98
Amp	Yangtse	10.62	11.39	–	15.65	–	–	–	11.98	1.58	–	–	46.24	–	–	–
	Yarlung	6.63	11.19	–	15.05	–	–	–	13.77	0.93	–	–	50.13	–	–	–
	PumQu	12.84	11.31	–	18.35	–	–	–	8.98	1.49	–	–	43.53	–	–	–
Tur	Yangtse	11.84	11.77	–	13.44	–	–	–	12.65	1.5	–	–	47.34	–	–	–
	Yarlung	38.54	0.87	–	9.81	–	–	–	6.11	2.28	–	–	41.40	–	–	–
	PumQu	39.28	0.30	–	11.08	–	–	–	4.41	2.53	–	–	41.51	–	–	–

Besides shared features, some heavy minerals collected from certain river basins may have extra chemical compositions that are specific to the river basin. For example, a sample of Mnz (monazite) from the PumQu River contains IrO_2 not shared by the sample from the Yarlung River, while a sample of Zrn (zircon) from the Yarlung River does not contain WO_3 common in those from the PumQu and Yangtse rivers.

3. Methodology

3.1. Problem definition

As described above, the identification of heavy minerals in each river basin is considered as a separate classification task. Suppose there are T heavy mineral classification tasks $\mathcal{T}_t$, corresponding to T river basins, where $t \in \{1, 2, \ldots, T\}$ means the tth task. The training set for each task $\mathcal{T}_t$ consists of the SEM-EDS feature matrix $\mathbf{X}^t \in \mathbb{R}^{N_t \times (d+1)}$, where each row corresponds to a pre-identified heavy mineral sample, as well as the vector of class labels $\mathbf{y}^t \in \{1, 2, \ldots, C\}^{N_t}$. Here $N_t \geq 1$ is the number of heavy minerals in the tth task and $C \geq 3$ is the number of mineral classes. For ease of calculation, the dimension of each feature vector $\mathbf{x} = (x_0, x_1, \ldots, x_d)$ is $d + 1$, where $x_0 = 1$ corresponds to the intercept term (Sun et al., 2018).

Due to manual identification cost, the available identified heavy minerals collected from the individual river basin are far from sufficient to train a useful classifier to identify unknown heavy minerals. The goal is to apply the multi-task learning, which exploits the relationship among multiple river basins and allows them to borrow knowledge from each other, to jointly train a model

$$\mathbf{f} = \{f^1, f^2, \ldots, f^T\} \tag{1}$$

where f^t is a multi-class classifier for the heavy mineral identification of the tth river basin (i.e. task).

Notations: For any matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$, its $\ell_{2,1}$ norm is $\|\mathbf{M}\|_{2,1} = \sum_{j=1}^{m} \left\|\mathbf{M}_j\right\|_2$, where $\left\|\mathbf{M}_j\right\|_2 = \sqrt{\sum_{k=1}^{n} \left|\mathbf{M}_{j,k}\right|^2}$, i.e. the sum of the ℓ_2 norm of rows in $\mathbf{M}$; its $\ell_{1,1}$ norm is $\|\mathbf{M}\|_{1,1} = \sum_{j=1}^{m} \sum_{k=1}^{n} |\mathbf{M}_{j,k}|$, which equals the sum of absolute values of the elements.

3.2. Multi-task multi-class learning

Framework. Given multiple related tasks, a common regularized framework for learning them simultaneously is formalized as follows:

$$\min_{\mathcal{W}} \quad \sum_{t=1}^{T} \mathcal{L}\left(\mathbf{y}^t, \mathbf{X}^t\right) + \mathcal{R}(\mathcal{W}) \tag{2}$$

where $\mathcal{W}$ denotes the collated parameters of all the classifiers. $\mathcal{L}(\cdot, \cdot)$ is a loss function measuring the empirical loss on the training set of the tth river basin, while $\mathcal{R}(\mathcal{W})$ is the regularization on $\mathcal{W}$. Different choice of $\mathcal{R}(\cdot)$ will lead to different regularization.

Multi-class classification. Generally, a classifier aims to learn a function $f : \mathbb{R}^{(d+1)} \mapsto \mathbb{R}^C$ that maps the heavy minerals to specific classes or vectors of class scores. A common mapping function given a heavy mineral vector $\mathbf{x}$ is defined as:

$$f(\mathbf{x}, \mathbf{W}) = \mathbf{x}\mathbf{W}^\top \tag{3}$$

where $\mathbf{W} = [\mathbf{w}_1; \mathbf{w}_2; \ldots; \mathbf{w}_C] \in \mathbb{R}^{C \times (d+1)}$ is the parameter matrix, in which each row corresponds to a mineral class.

To deal with the multi-class classification of heavy minerals, we use the softmax classifier, which interprets the class scores resulted from the function f as the probabilities for class labels. The probability assigned to the cth class given the heavy mineral vector $\mathbf{x}$ and the parameter vector of the cth class $\mathbf{w}_c$ (i.e. the cth row of the parameter matrix $\mathbf{W}$) is defined by the formula:

$$\text{softmax}(\mathbf{x}, \mathbf{w}_c) = \frac{\exp\left(\mathbf{x}\mathbf{w}_c^\top\right)}{\sum_{j=1}^{C} \exp\left(\mathbf{x}\mathbf{w}_j^\top\right)} \tag{4}$$

where $\exp(\cdot)$ is the exponential function. The softmax classifier directly outputs the probability distribution over C candidate classes, which can avoid the problems of class imbalance and time overhead introduced by the common "One vs. One" and "One vs. Rest" strategies for multi-class classification.

Based on the softmax classifier, the cross-entropy function is used to measure the empirical loss. Hence, the loss function $\mathcal{L}(\cdot, \cdot)$ in Formula (2) over all mineral classes and all training heavy minerals is formalized as:

$$\mathcal{L}\left(\mathbf{y}^t, \mathbf{X}^t\right) = -\sum_{i=1}^{N_t} \sum_{c=1}^{C} \mathbb{I}\left(y_i^t = c\right) \log\left(\text{softmax}(\mathbf{x}_i^t, \mathbf{w}_c^t)\right) \tag{5}$$

where $\mathbf{x}_i^t$ and y_i^t, corresponding to the ith row of the feature matrix $\mathbf{X}^t$ and the ith element of the label vector $\mathbf{y}^t$, are the feature and ground truth label of the ith heavy mineral sample respectively, while $\mathbf{w}_c^t$ is the parameter vector of the cth mineral class for the tth task $\mathcal{T}_t$. Separately, $\mathbb{I}(\cdot)$ and $\log(\cdot)$ are the indicator function and logarithmic function.

Multi-task learning. In order to jointly learn the classification tasks across multiple the river basins, we collate the parameter matrices of the classifiers for all the river basins together and $\mathcal{W}$ in Formula (2) is set to be a 3-dimensional matrix, i.e. $\mathcal{W} \in \mathbb{R}^{C \times (d+1) \times T}$.

As mentioned in Section 2, the automatic identification of heavy minerals depends on both shared features among all the river basins and specific features for the individual river basin. We decompose the parameter matrix $\mathcal{W}$ as two components $\mathcal{P}$ and $\mathcal{Q}$ accordingly, i.e. $\mathcal{W} = \mathcal{P} + \mathcal{Q}$, where $\mathcal{P}$ captures the shared features and $\mathcal{Q}$ indicates the specific features. We apply different regularization terms on the parameter matrices $\mathbf{P}_{c::}$ and $\mathbf{Q}_{c::} \in \mathbb{R}^{(d+1) \times T}$, which are the cth horizontal slices of

Algorithm 1: The FISTA Algorithm for MTMC.

Input: Feature set $\{\mathbf{X}^t\}_{t=1}^T$; Label set $\{\mathbf{y}^t\}_{t=1}^T$; Number of iterations M;
Output: $\mathcal{W} = \mathcal{P} + \mathcal{Q}$;
1: Initialize the matrices $\mathcal{P}^0$, $\mathcal{Q}^0$ and step size L^0;
2: Let $\theta^1 = 1$ and $\tilde{\mathcal{P}}^1 = \mathcal{P}^0$ and $\tilde{\mathcal{Q}}^1 = \mathcal{Q}^0$;
3: **for** k = 1 to M **do**
4: Update θ^{k+1}:

$$\theta^{k+1} = \frac{1+\sqrt{1+4(\theta^k)^2}}{2};$$

5: Search the proper step size $\bar{L}$ linearly, such that

$$f\left(\text{prox}_{\bar{L}}(\tilde{\mathcal{P}}^k), \text{prox}_{\bar{L}}(\tilde{\mathcal{Q}}^k)\right) \le Q_{\bar{L}}\left(\tilde{\mathcal{P}}^k, \tilde{\mathcal{Q}}^k\right);$$

6: **for** each class **do**
7: Set $L^k = \bar{L}$, and compute $\mathbf{P}_c^k$ and $\mathbf{Q}_c^k$:

$$\mathbf{P}_c^k = \text{prox}_{L^k}(\tilde{\mathbf{P}}_c^k),$$

$$\mathbf{Q}_c^k = \text{prox}_{L^k}(\tilde{\mathbf{Q}}_c^k);$$

8: Update $\tilde{\mathbf{P}}_c^{k+1}$ and $\tilde{\mathbf{Q}}_c^{k+1}$:

$$\tilde{\mathbf{P}}_c^{k+1} = \mathbf{P}_c^k + \frac{\theta^k - 1}{\theta^{k+1}}\left(\mathbf{P}_c^k - \mathbf{P}_c^{k-1}\right),$$

$$\tilde{\mathbf{Q}}_c^{k+1} = \mathbf{Q}_c^k + \frac{\theta^k - 1}{\theta^{k+1}}\left(\mathbf{Q}_c^k - \mathbf{Q}_c^{k-1}\right);$$

9: **end for**
10: **end for**

$\mathcal{P}$ and $\mathcal{Q}$ respectively, to exploit these two types of features. Formally, we define $\mathcal{R}(\cdot)$ in Formula (2) as:

$$\begin{aligned} &\mathcal{R}(\mathcal{W}) = \alpha \sum_{c=1}^{C} \|\mathbf{P}_{c::}\|_{2,1} + \beta \sum_{c=1}^{C} \|\mathbf{Q}_{c::}\|_{1,1} \\ &\text{s.t.} \quad \mathcal{W} = \mathcal{P} + \mathcal{Q} \end{aligned} \tag{6}$$

where α and β are regularization parameters to control the trade-off between the two terms. Specifically, the first term restricts each matrix $\mathbf{P}_{c::}$ to be row-sparse via the $\ell_{2,1}$ norm. That is, only a small part of rows in $\mathbf{P}_{c::}$ consist of all non-zero elements, which is equivalent to selecting shared features (i.e. the non-zero rows) of heavy minerals across the river basins. The second term enforces parameter matrix $\mathbf{Q}_{c::}$ to be elementwise-sparse via the $\ell_{1,1}$ norm. That is, there exist only a small part of non-zero elements in $\mathbf{Q}_{c::}$, which correspond to the specific features useful for the individual river basin.

Based on above, the goal of MTMC is to minimize the following objective function:

$$\begin{aligned} \min_{\mathcal{W},\mathcal{P},\mathcal{Q}} \quad & -\sum_{t=1}^{T}\sum_{i=1}^{N_t}\sum_{c=1}^{C} \mathbb{I}\left(y_i^t = c\right) \log\left(\text{softmax}(\mathbf{x}_i^t, \mathbf{w}_c^t)\right) \\ & + \alpha \sum_{c=1}^{C} \|\mathbf{P}_{c::}\|_{2,1} + \beta \sum_{c=1}^{C} \|\mathbf{Q}_{c::}\|_{1,1} \\ \text{s.t.} \quad & \mathcal{W} = \mathcal{P} + \mathcal{Q} \end{aligned} \tag{7}$$

From the perspective of the heavy mineral classes, we use multi-task learning to jointly train the model for all the river basins, taking $\mathbf{W}_{c::}$ as the parameter matrix for the cth mineral class; while from the perspective of the river basins (i.e. tasks), we derive a multi-class classifier for each of them, taking the parameter matrix $\mathbf{W}_{::t}$, which is the tth frontal slice of $\mathcal{W}$. Fig. 1 illustrates the details of MTMC.

3.3. Optimization

Since the loss function and the regularization terms are both convex, Formula (7) is evidently convex. Hence, we can use the FISTA (Beck and Teboulle, 2009) algorithm to find the optimal solution. The FISTA algorithm is used to minimize a linear combination of two convex functions, i.e. $\min_{\boldsymbol{\Phi}} f(\boldsymbol{\Phi}) + g(\boldsymbol{\Phi})$, where $f(\boldsymbol{\Phi})$ is both convex and smooth, while $g(\boldsymbol{\Phi})$ is convex but can be non-smooth.

For notational convenience, we abbreviate $\mathbf{P}_{c::}$ and $\mathbf{Q}_{c::}$ as $\mathbf{P}_c$ and $\mathbf{Q}_c$. In order to apply the FISTA algorithm, we set

$$\begin{aligned} f\left(\mathbf{P}_c, \mathbf{Q}_c\right) &= -\sum_{t=1}^{T}\sum_{i=1}^{N_t}\sum_{c=1}^{C} \mathbb{I}\left(y_i^t = c\right) \log\left(\text{softmax}(\mathbf{x}_i^t, \mathbf{w}_c^t)\right) \\ g\left(\mathbf{P}_c, \mathbf{Q}_c\right) &= \alpha \sum_{c=1}^{C} \|\mathbf{P}_c\|_{2,1} + \beta \sum_{c=1}^{C} \|\mathbf{Q}_c\|_{1,1} \end{aligned} \tag{8}$$

and define the first order Taylor expansion of $f\left(\mathbf{P}_c, \mathbf{Q}_c\right)$ at $(\tilde{\mathbf{P}}_c, \tilde{\mathbf{Q}}_c)$, using the squared Euclidean distance between $\left(\mathbf{P}_c, \mathbf{Q}_c\right)$ and $(\tilde{\mathbf{P}}_c, \tilde{\mathbf{Q}}_c)$, which is formulated as:

$$\begin{aligned} Q_L\left(\mathbf{P}_c, \mathbf{Q}_c\right) \simeq\ & f\left(\tilde{\mathbf{P}}_c, \tilde{\mathbf{Q}}_c\right) + \left\langle \nabla_{\tilde{\mathbf{P}}_c} f\left(\tilde{\mathbf{P}}_c, \tilde{\mathbf{Q}}_c\right), \mathbf{P}_c - \tilde{\mathbf{P}}_c \right\rangle \\ & + \frac{L}{2}\left\|\mathbf{P}_c - \tilde{\mathbf{P}}_c\right\|_F^2 + \left\langle \nabla_{\tilde{\mathbf{Q}}_c} f\left(\tilde{\mathbf{P}}_c, \tilde{\mathbf{Q}}_c\right), \mathbf{Q}_c - \tilde{\mathbf{Q}}_c \right\rangle \\ & + \frac{L}{2}\left\|\mathbf{Q}_c - \tilde{\mathbf{Q}}_c\right\|_F^2 \end{aligned} \tag{9}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product of two matrices. $\nabla_{\tilde{\mathbf{P}}_c} f(\tilde{\mathbf{P}}_c, \tilde{\mathbf{Q}}_c)$ and $\nabla_{\tilde{\mathbf{Q}}_c} f(\tilde{\mathbf{P}}_c, \tilde{\mathbf{Q}}_c)$ are the partial derivatives of $f(\tilde{\mathbf{P}}_c, \tilde{\mathbf{Q}}_c)$, with respect to $\mathbf{P}_c$ and $\mathbf{Q}_c$ at $(\tilde{\mathbf{P}}_c, \tilde{\mathbf{Q}}_c)$. We can obtain the solution by minimizing the following proximal operator iteratively:

$$\begin{aligned} \min_{\mathbf{P}_c, \mathbf{Q}_c} \quad & \frac{L}{2}\left\|\mathbf{P}_c - \left(\tilde{\mathbf{P}}_c - \frac{1}{L}\nabla_{\tilde{\mathbf{P}}_c} f\left(\tilde{\mathbf{P}}_c, \tilde{\mathbf{Q}}_c\right)\right)\right\|_F^2 \\ & + \frac{L}{2}\left\|\mathbf{Q}_c - \left(\tilde{\mathbf{Q}}_c - \frac{1}{L}\nabla_{\tilde{\mathbf{Q}}_c} f\left(\tilde{\mathbf{P}}_c, \tilde{\mathbf{Q}}_c\right)\right)\right\|_F^2 \\ & + g\left(\mathbf{P}_c, \mathbf{Q}_c\right) \end{aligned} \tag{10}$$

Since Formula (10) is decomposable, we can cast it into two separate proximal operators. So the solutions of $\mathbf{P}_c$ and $\mathbf{Q}_c$ at the kth iteration are formulated respectively as follows:

$$\mathbf{P}_c^k = \text{prox}_{\frac{\alpha}{L}g}\left(\tilde{\mathbf{P}}_c^{k-1} - \frac{1}{L}\nabla_{\tilde{\mathbf{P}}_c} f\left(\tilde{\mathbf{P}}_c, \tilde{\mathbf{Q}}_c\right)\right) \tag{11}$$

$$\mathbf{Q}_c^k = \text{prox}_{\frac{\beta}{L}g}\left(\tilde{\mathbf{Q}}_c^{k-1} - \frac{1}{L}\nabla_{\tilde{\mathbf{Q}}_c} f\left(\tilde{\mathbf{P}}_c, \tilde{\mathbf{Q}}_c\right)\right) \tag{12}$$

where L is the step size and $\text{prox}_{\frac{*}{L}g}(\cdot)$ is the proximal operator. For the $\ell_{2,1}$ norm, $\text{prox}_{\frac{*}{L}g}(\cdot)$ is defined as:

$$\left(\text{prox}_{\frac{\alpha}{L}g}(\mathbf{P})\right)_c = \mathbb{I}\left(\|\mathbf{P}_c\|_2 \geqslant \frac{\alpha}{L}\right)\left(\left(1 - \frac{\alpha}{L\|\mathbf{P}_c\|_2}\right)\mathbf{P}_c\right) \tag{13}$$

While for the $\ell_{1,1}$ norm, it is defined as:

$$\left(\text{prox}_{\frac{\beta}{L}g}(\mathbf{Q})\right)_c = \mathbb{I}\left(\left|\mathbf{Q}_{cij}\right| > \frac{\beta}{L}\right)\left(\text{sgn}\left(\mathbf{Q}_{cij}\right)\left(\left|\mathbf{Q}_{cij}\right| - \frac{\beta}{L}\right)\right) \tag{14}$$

where $\mathbf{Q}_{cij}$ returns the (i,j)th element of $\mathbf{Q}_c$, $\text{sgn}(\cdot)$ is the sign function, and $|\cdot|$ returns the absolute value. Algorithm 1 lists the pseudo code of the FISTA algorithm.

4. Experimental settings

4.1. Dataset

In this paper, totally 1882 manually identified heavy minerals are acquired, which include 908 heavy mineral samples from the Yangtse River, 456 samples from the Yarlung River and 518 samples from the PumQu River. SEM-EDS features are taken from the heavy mineral samples. All the heavy minerals belong to 11 mineral classes, i.e. monazite (Mnz), amphibole (Amp), apatite (Ap), biotite (Bt), epidote (Ep), garnet (Grt), pyroxene (Px), rutile (Rt), sphene (Spn), tourmaline (Tur) and zircon (Zrn). Table 2 lists the numbers of heavy minerals of the 11 mineral classes collected from the three river basins. As shown in Table 2, the collected heavy mineral samples are relatively few in each of the 11 mineral classes from each river basin.

4.2. Performance measurements

We adopt the precision, recall and F1-score weighted by the proportion of the mineral classes to evaluate the performance of different

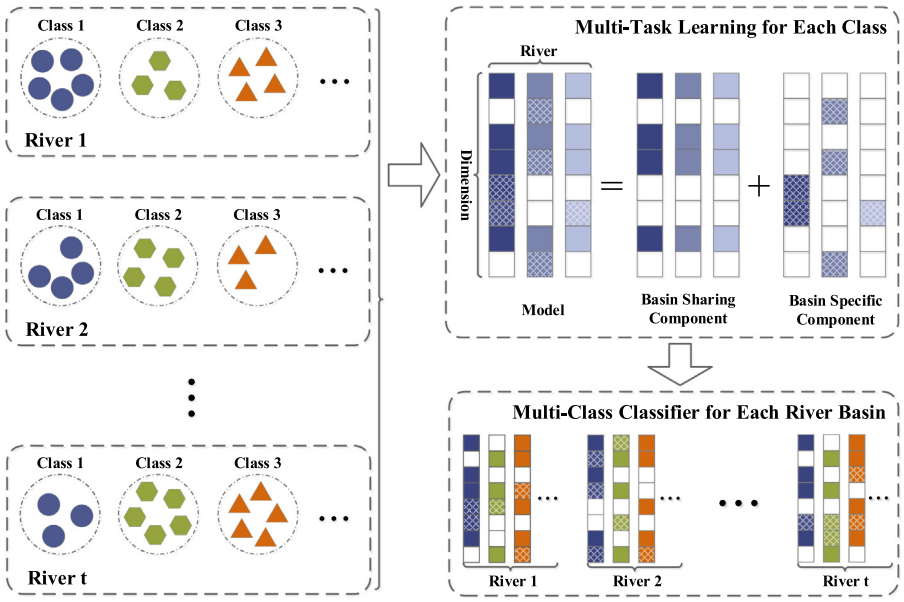


Fig. 1. Illustration of MTMC, where the heavy mineral classes are marked by different colors (i.e. blue, green, orange, etc.). MTMC jointly trains the multi-class classifiers for the river basins via the multi-task learning. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Statistics of the heavy minerals collected from the three river basins.

Mineral Name	Abbr.	Yangtse	Yarlung	PumQu
Monazite	Mnz	0	50	52
Amphibole	Amp	238	48	39
Apatite	Ap	77	29	47
Biotite	Bt	0	43	38
Epidote	Ep	138	33	55
Garnet	Grt	149	56	47
Pyroxene	Px	33	36	48
Rutile	Rt	0	38	57
Sphene	Spn	80	50	40
Tourmaline	Tur	45	43	45
Zircon	Zrn	148	30	50
Total number		908	456	518

methods. These measurements are suitable for multi-class classification and account for the class imbalance. The weighted precision (Pr_c), recall (Rec_c) and F1-score ($F1_c$) for a single mineral class are respectively computed as follows:

$$Pr_c = w_c \times \frac{TP_c}{TP_c + FP_c} \tag{15}$$

$$Rec_c = w_c \times \frac{TP_c}{TP_c + FN_c} \tag{16}$$

$$F1_c = w_c \times \frac{2 \times Pr_c \times Rec_c}{Pr_c + Rec_c} \tag{17}$$

where $w_c = \frac{N_c}{\sum_{c=1}^{C} N_c}$ is the proportion of minerals from the cth mineral class and N_c is the number of heavy minerals belonging to the cth class. TP_c, FP_c and FN_c denote the number of true positive (i.e. heavy minerals belonging to the cth class correctly identified as the cth class), false positive (i.e. heavy minerals belonging to other classes incorrectly identified as the cth class) and false negative (i.e. heavy minerals belonging to the cth class incorrectly identified as other classes) respectively. The overall measurements (Pr, Rec and $F1$) are the weighted average over the sub-measurements for all the mineral classes. For example, precision Pr is computed by $Pr = \frac{1}{C}\sum_{c=1}^{C} Pr_c$. All the measurements range from 0 to 1, and better methods lead to greater values.

4.3. Experimental design

To evaluate the performance of MTMC, we design two groups of experiments. Firstly, we compare MTMC with the single task learning methods, which trains the classifier for each river basin individually, and measure the performance. The single task learning methods for comparison include Support Vector Machine (SVM) (Cortes and Vapnik, 1995), Linear Discriminant Analysis (LDA) (Fisher, 1936), Naïve Bayes (NB) (Nigam et al., 2000), Nearest Neighbor (1NN) (Cover et al., 1967), Decision Tree (DT) (Quinlan, 1986), AdaBoost (Freund and Schapire, 1997), Random Under Sampling (RUSBoost) (Seiffert et al., 2009), Bootstrap Aggregation (Bagging) (Breiman, 1996), Linear Programming Boost (LPBoost) (Demiriz et al., 2002) and TotalBoost (Warmuth et al., 2006). These are classic, commonly used methods for single task learning. We train the classifiers for the single task learning methods under two settings: one is to use the basin's own training set, the other is to use the combination of the training sets from all the river basins. We evaluate the performance on the test set from each of the three river basins respectively. Note that if some heavy mineral classes are not found in the river's specific dataset (e.g. the Yangtse River), they will be ignored during both the training and evaluation, i.e. the classifier corresponding to this river only identifies the mineral classes it contains. By comparing with the above methods, we can validate the performance of MTMC with high confidence.

Secondly, we compare MTMC with other multi-task learning methods. The compared methods include: (1) LASSO (Tibshirani, 1996), which applied the $\ell_{1,1}$ norm on the parameter matrix $\mathcal{W}$; (2) MTFS (Obozinski et al., 0000), which applied the $\ell_{2,1}$ norm on $\mathcal{W}$; (3) DMTL (Jalali et al., 2010), which applied the $\ell_{\infty,1}$ and $\ell_{1,1}$ norms on the decomposed parameter matrices $\mathcal{P}$ and $\mathcal{Q}$ respectively; (4) rMTFL (Gong et al., 2012), which applied the $\ell_{2,1}$ norm on the rows of $\mathcal{P}$ and columns of $\mathcal{Q}$ respectively; (5) STML (Chen et al., 2012), which defined a convex constraint on $\mathcal{P}$ and applied the $\ell_{2,1}$ norm on $\mathcal{Q}$; (6) RMTL (Chen et al., 2011), which applied the trace norm on $\mathcal{P}$ and the $\ell_{2,1}$ norm on the columns of $\mathcal{Q}$. Note that the available multi-task learning methods are designed for binary classification. We use the " One vs. Rest" strategy to adapt them for the multi-class classification suitable for heavy minerals.

For evaluation, we partition the original dataset into a training set for training the classifiers, and a test set for evaluating the performance for the identification of heavy minerals. Specifically, we use 10-fold

Table 3
The performance measurements (mean ± std) of different methods using single task learning for the identification of heavy minerals, two rows correspond to two sets of training data: self vs. combined. Measurements of MTMC are listed for comparison. The method abbreviations include SVM (Support Vector Machine), LDA (Linear Discriminant Analysis), NB (Naïve Bayes), 1NN (Nearest Neighbor), DT (Decision Tree), RUSBoost (Random Under Sampling), Bagging (Bootstrap Aggregation), and LPBoost (Linear Programming Boost).

Methods	Yangtse			Yarlung			PumQu		
	Pr	Rec	F1	Pr	Rec	F1	Pr	Rec	F1
SVM	0.94 ± 0.01 0.88 ± 0.01	0.93 ± 0.01 0.86 ± 0.01	0.93 ± 0.01 0.85 ± 0.01	0.94 ± 0.01 0.88 ± 0.02	0.94 ± 0.01 0.85 ± 0.02	0.93 ± 0.01 0.84 ± 0.02	0.97 ± 0.00 0.83 ± 0.03	0.97 ± 0.00 0.82 ± 0.03	0.97 ± 0.00 0.79 ± 0.04
LDA	0.94 ± 0.01 0.87 ± 0.01	0.93 ± 0.01 0.84 ± 0.01	0.93 ± 0.01 0.84 ± 0.01	0.95 ± 0.01 0.89 ± 0.01	0.95 ± 0.01 0.82 ± 0.01	0.94 ± 0.01 0.83 ± 0.01	0.98 ± 0.02 0.96 ± 0.01	0.97 ± 0.02 0.95 ± 0.01	0.98 ± 0.02 0.95 ± 0.01
NB	0.67 ± 0.06 0.57 ± 0.02	0.58 ± 0.06 0.47 ± 0.04	0.57 ± 0.06 0.47 ± 0.03	0.65 ± 0.04 0.48 ± 0.05	0.65 ± 0.04 0.48 ± 0.04	0.61 ± 0.04 0.43 ± 0.04	0.75 ± 0.04 0.58 ± 0.04	0.80 ± 0.03 0.63 ± 0.02	0.75 ± 0.04 0.57 ± 0.03
1NN	0.93 ± 0.01 0.89 ± 0.01	0.93 ± 0.01 0.84 ± 0.01	0.92 ± 0.01 0.85 ± 0.01	0.95 ± 0.01 0.73 ± 0.03	0.94 ± 0.01 0.67 ± 0.02	0.94 ± 0.01 0.65 ± 0.02	0.98 ± 0.00 0.85 ± 0.01	0.97 ± 0.00 0.75 ± 0.02	0.97 ± 0.01 0.77 ± 0.03
DT	0.90 ± 0.02 0.85 ± 0.03	0.89 ± 0.02 0.78 ± 0.03	0.89 ± 0.03 0.80 ± 0.03	0.70 ± 0.05 0.79 ± 0.03	0.73 ± 0.05 0.77 ± 0.02	0.69 ± 0.05 0.75 ± 0.03	0.91 ± 0.01 0.84 ± 0.02	0.88 ± 0.02 0.86 ± 0.01	0.88 ± 0.02 0.84 ± 0.02
AdaBoost	0.87 ± 0.04 0.89 ± 0.02	0.87 ± 0.04 0.88 ± 0.01	0.86 ± 0.04 0.88 ± 0.01	0.84 ± 0.03 0.88 ± 0.01	0.84 ± 0.03 0.86 ± 0.01	0.82 ± 0.03 0.86 ± 0.01	0.67 ± 0.14 0.90 ± 0.02	0.68 ± 0.12 0.90 ± 0.03	0.65 ± 0.13 0.88 ± 0.03
RUSBoost	0.93 ± 0.01 0.81 ± 0.05	0.91 ± 0.01 0.72 ± 0.04	0.91 ± 0.01 0.75 ± 0.05	0.87 ± 0.03 0.81 ± 0.02	0.87 ± 0.02 0.81 ± 0.02	0.86 ± 0.03 0.79 ± 0.02	0.11 ± 0.09 0.77 ± 0.02	0.17 ± 0.08 0.82 ± 0.02	0.11 ± 0.09 0.78 ± 0.02
Bagging	0.94 ± 0.01 0.91 ± 0.01	0.94 ± 0.01 0.86 ± 0.01	0.93 ± 0.01 0.87 ± 0.01	0.96 ± 0.01 0.92 ± 0.01	0.95 ± 0.01 0.89 ± 0.01	0.95 ± 0.01 0.89 ± 0.01	0.97 ± 0.01 0.96 ± 0.01	0.97 ± 0.01 0.96 ± 0.01	0.97 ± 0.01 0.95 ± 0.01
LPBoost	0.92 ± 0.03 0.86 ± 0.02	0.90 ± 0.03 0.81 ± 0.02	0.90 ± 0.03 0.81 ± 0.02	0.89 ± 0.04 0.82 ± 0.02	0.88 ± 0.03 0.81 ± 0.02	0.87 ± 0.03 0.79 ± 0.02	0.78 ± 0.10 0.82 ± 0.02	0.79 ± 0.10 0.84 ± 0.01	0.77 ± 0.10 0.82 ± 0.01
TotalBoost	0.90 ± 0.01 0.86 ± 0.02	0.89 ± 0.02 0.81 ± 0.02	0.88 ± 0.02 0.81 ± 0.02	0.87 ± 0.04 0.79 ± 0.04	0.86 ± 0.02 0.78 ± 0.03	0.85 ± 0.03 0.76 ± 0.04	0.69 ± 0.17 0.79 ± 0.02	0.69 ± 0.15 0.82 ± 0.02	0.67 ± 0.16 0.79 ± 0.02
MTMC	0.96 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.98 ± 0.01	0.98 ± 0.01	0.98 ± 0.01	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00

cross validation, which randomly splits the dataset into 10 equal size subsets, in which a single subset is retained as the test set while the remaining 9 subsets are used as the training set. This process is repeated 10 times, with each of the 10 subsets used exactly once as the test set. The results from the 10 folds are averaged to produce an overall measurement of performance.

For the hyperparameters α and β in MTMC, we choose the optimal values from the candidate range $[0.001, 1]$ using the validation data reserved from the training set. For the compared methods, the required parameters are set as the optimal values suggested by the authors. During the experiments, we evaluate the performance measurements on each river basin separately, and make the comparisons accordingly.

5. Results and discussion

5.1. Comparison with single task learning

Table 3 reports the performance of the 10 most commonly used single task learning methods, each of which has two rows of measurements, where the first row corresponds to the results from the basin's own training set and the second row corresponds to the results from the combined training set (as described in Section 4.3). Fig. 2 shows the F1 measurements of these methods accordingly.

The results in Table 3 and Fig. 2 show that MTMC indeed outperforms all the single task learning methods. Some of the single task learning methods, such as SVM, LDA, 1NN and Bagging can achieve passable performance, while other methods such as NB, perform poorly on certain river basins. This explains that the shortage of identified heavy minerals does compromise the performance of the trained classifiers. By simply combining the training data of all the river basins, the performance of most methods becomes even worse. This demonstrates that simply combining the training sets cannot compensate for the shortage of training data, and may compromise the performance by introducing noise. On the other hand, MTMC obtains 2%–3% performance improvement compared to the best single task learning method, which demonstrates that MTMC can take advantage of the shared knowledge among heavy minerals collected from different river basins by using multi-task learning, i.e. jointly learning the multiple classification tasks.

5.2. Comparison with multi-task learning

To further demonstrate the effectiveness of MTMC, we compare it with the current multi-task learning methods. Table 4 reports the performance measurements of 6 most used multi-task learning methods, along with MTMC. Fig. 3 shows the F1 measurements accordingly.

From Table 4 and Fig. 3, we can see that MTMC consistently outperforms other multi-task learning methods. LASSO and MTFS actually correspond to the case where α or β is set to be 0 in MTMC, respectively. Compared with these two, MTMC obtains significantly improved performance across all the river basins. This is because LASSO only considers the specific features for the individual river basin and ignores the shared features among different river basins, while MTFS is just the opposite, which focuses on the shared features and ignores the specific ones. This demonstrates that both the shared features and specific features of the river basins are required comprehensively to improve the identification performance of heavy minerals. The other four compared methods all consider both shared and specific features, similar to MTMC. DMTL achieves the best performance among the compared multi-task learning methods, although performs worse than MTMC. Specifically, MTMC obtains the identification accuracy of 100% on the PumQu River, which is 2% greater than DMTL. This demonstrates the superiority of multi-class classification (i.e. softmax) used in MTMC, against the "One vs. Rest" strategy used in the compared methods. Class imbalance introduced by the "One vs. Rest" strategy dose compromise the performance of trained classifiers in the case of identification of heavy minerals.

5.3. Discussion

To further study the identification accuracy of MTMC among the mineral classes, we randomly choose 60% heavy mineral samples as the training set, 20% as the validation set, and the rest 20% as the test set from each river basin respectively, and conduct the experiments using MTMC. Fig. 4 presents the resulted confusion matrices on the three river basins, where the horizontal axis corresponds to the predicted mineral classes and the vertical axis corresponds to the ground truth. According to Figs. 4(b) and 4(c), MTMC obtains excellent performance

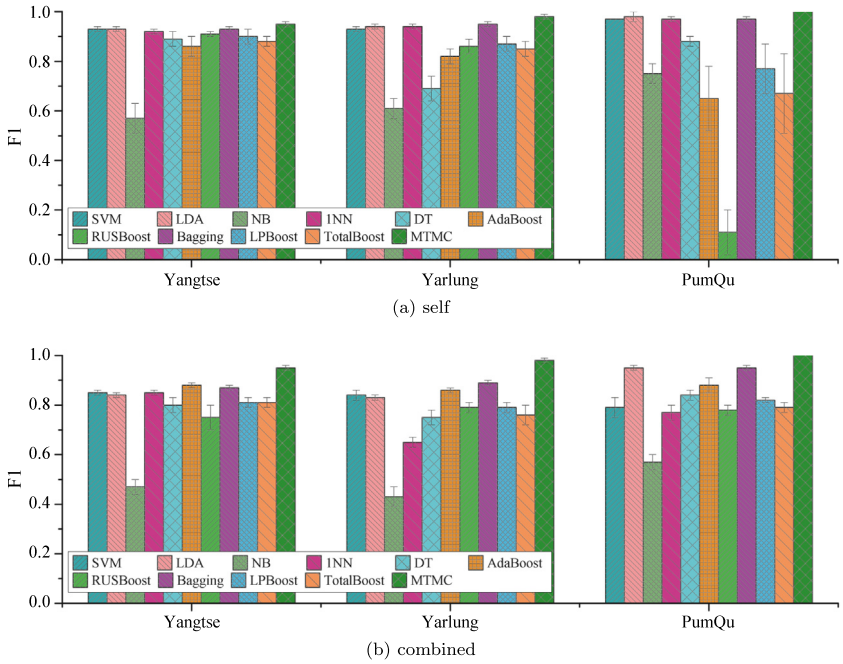


Fig. 2. The F1 measurements of different single task learning methods using self vs. combined training sets.

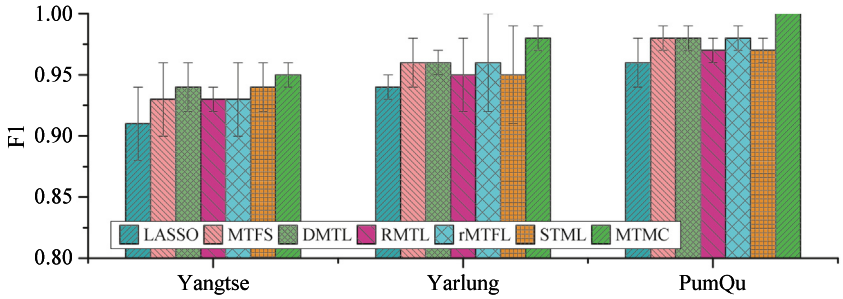


Fig. 3. The F1 measurements of different multiple task learning methods.

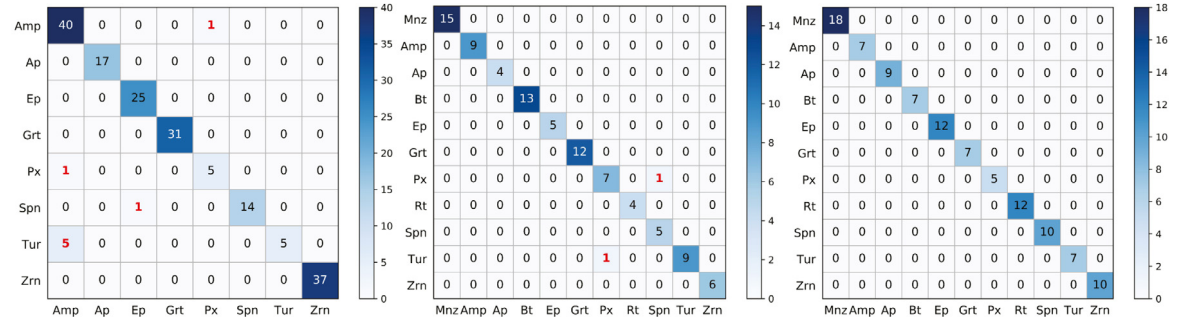
Table 4
The performance measurements (mean ± std) of different methods using multi-task learning for the identification of heavy minerals. The method abbreviations include LASSO ($\ell_{1,1}$ norm on $\mathcal{W}$), MTFS ($\ell_{2,1}$ norm on $\mathcal{W}$), DMTL ($\ell_{\infty,1}$ norm on $\mathcal{P}$ and $\ell_{1,1}$ norm on $\mathcal{Q}$), RMTL (Trace norm on $\mathcal{P}$ and $\ell_{2,1}$ norm on the columns of $\mathcal{Q}$), rMTFL ($\ell_{2,1}$ norm on the rows of $\mathcal{P}$ and columns of $\mathcal{Q}$), STML (Convex constraint on $\mathcal{P}$ and $\ell_{2,1}$ norm on $\mathcal{Q}$).

Methods	Yangtse			Yarlung			PumQu		
	Pr	Rec	F1	Pr	Rec	F1	Pr	Rec	F1
LASSO	0.90 ± 0.04	0.93 ± 0.02	0.91 ± 0.03	0.94 ± 0.02	0.94 ± 0.01	0.94 ± 0.01	0.96 ± 0.02	0.96 ± 0.02	0.96 ± 0.02
MTFS	0.93 ± 0.03	0.94 ± 0.02	0.93 ± 0.03	0.97 ± 0.01	0.96 ± 0.02	0.96 ± 0.02	0.98 ± 0.01	0.98 ± 0.01	0.98 ± 0.01
DMTL	0.94 ± 0.01	0.94 ± 0.01	0.94 ± 0.02	0.97 ± 0.01	0.96 ± 0.01	0.96 ± 0.01	0.98 ± 0.01	0.98 ± 0.01	0.98 ± 0.01
RMTL	0.95 ± 0.01	0.94 ± 0.01	0.93 ± 0.01	0.96 ± 0.02	0.95 ± 0.02	0.95 ± 0.03	0.98 ± 0.01	0.97 ± 0.01	0.97 ± 0.01
rMTFL	0.94 ± 0.03	0.94 ± 0.02	0.93 ± 0.03	0.96 ± 0.02	0.96 ± 0.03	0.96 ± 0.04	0.98 ± 0.01	0.98 ± 0.01	0.98 ± 0.01
STML	0.94 ± 0.02	0.94 ± 0.01	0.94 ± 0.02	0.95 ± 0.03	0.95 ± 0.03	0.95 ± 0.04	0.98 ± 0.01	0.97 ± 0.02	0.97 ± 0.01
MTMC	0.96 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.98 ± 0.01	0.98 ± 0.01	0.98 ± 0.01	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00

on both the Yarlung and PumQu rivers, especially on the PumQu River, where all the heavy minerals are identified correctly. In contrast, the accuracy on the Yangtse River (Fig. 4(a)) is slightly inferior. As can be seen from Fig. 4, the vast misidentification between Tur (tourmaline) and Amp (amphibole) on the Yangtse River does not appear on the Yarlung and PumQu rivers. This may suggest noises exist within the dataset of the Yangtse River.

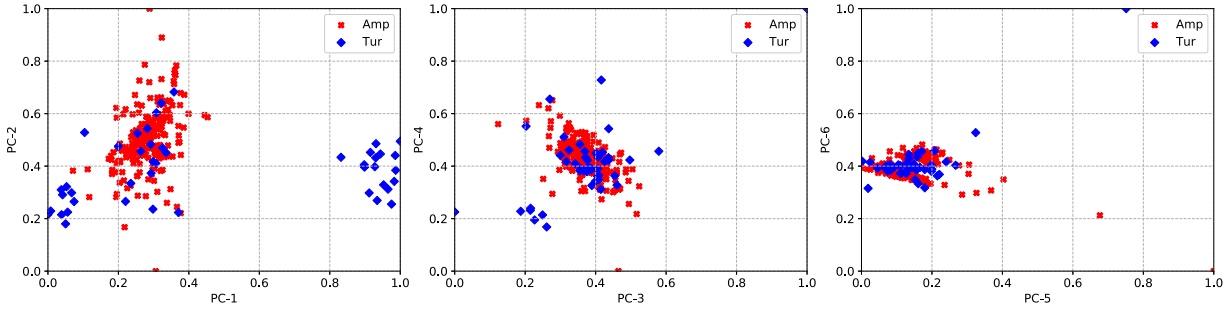
Another reason for the misidentification on the Yangtse River is that some classes of heavy minerals have similar or even identical chemical compositions and very close contents. For example, Tur and Amp have similar Al_2O_3, CaO, FeO, MgO, Na_2O and SiO_2 contents. For visualization, we perform Principal Component Analysis (PCA) (Jolliffe, 2011) on the SEM-EDS data of Tur and Amp from the Yangtse River. Table 5 lists the coefficients of the top six principal components. The scatter plots between pairs of the principal components are shown in Fig. 5. From Fig. 5, we can see that the data distribution of Tur and Amp on the Yangtse River is quite mixed up with each other, and none of the principal components can distinguish Amp and Tur effectively. In addition, we build the scatter plots between pairs of the main chemical compositions of Amp and Tur in Fig. 6, which indicates that the chemical compositions may lead to the misidentification of the two classes.

One possible solution is to take in the microscopic images of heavy minerals (e.g. in Fig. 7, where the visual difference between the Tur



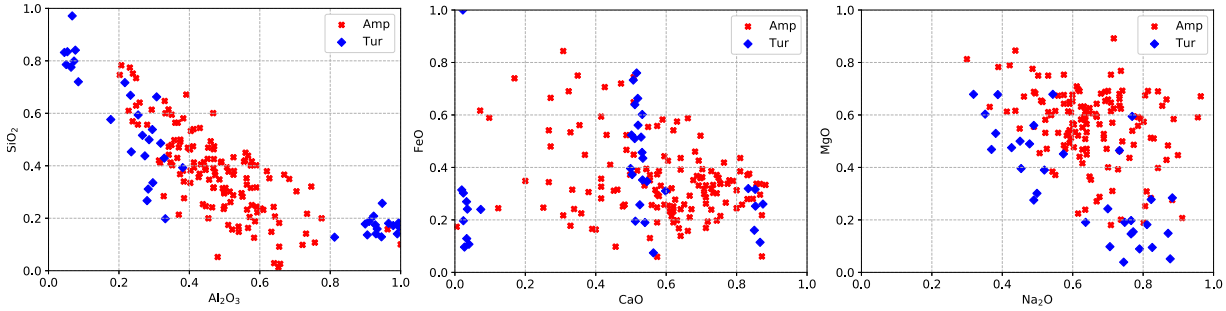
(a) Yangtse (b) Yarlung (c) PumQu

Fig. 4. The confusion matrices resulted from MTMC on the three river basins. The mineral abbreviations include Mnz (Monazite), Amp (Amphibole), Ap (Apatite), Bt (Biotite), Ep (Epidote), Grt (Garnet), Px (Pyroxene), Rt (Rutile), Spn (Sphene), Tur (Tourmaline), and Zrn (Zircon).



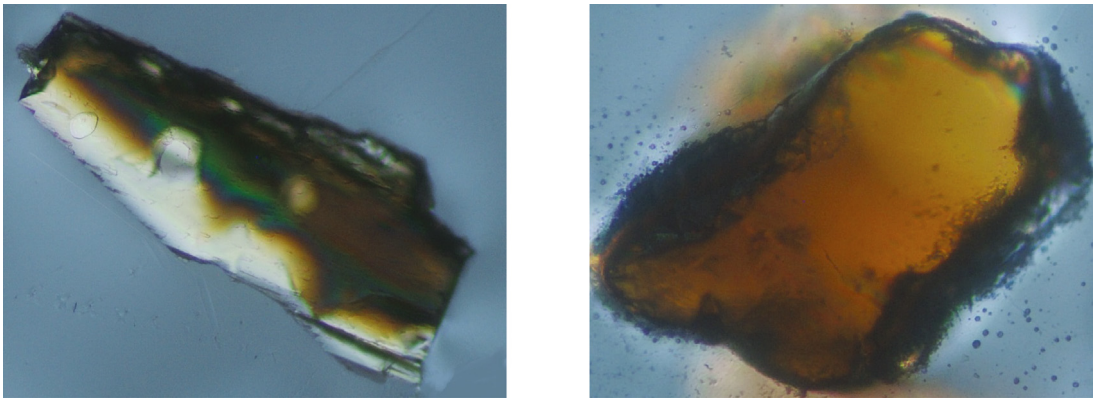
(a) PC-1 vs. PC-2 (b) PC-3 vs. PC-4 (c) PC-5 vs. PC-6

Fig. 5. The distribution of SEM-EDS data of Amp (Amphibole) and Tur (Tourmaline) from the Yangtse River, the axes correspond to pairs of principal components resulted from PCA respectively.



(a) Al_2O_3 vs. SiO_2 (b) CaO vs. FeO (c) Na_2O vs. MgO

Fig. 6. The distribution of SEM-EDS data of Amp (Amphibole) and Tur (Tourmaline) from the Yangtse River, the axes correspond to pairs of the main chemical compositions respectively.



(a) Amp (b) Tur

Fig. 7. The two samples of Amp (Amphibole) and Tur (Tourmaline) collected from the Yangtse River.

Table 5
The coefficients of the top six principle components (PC) resulted from PCA. Since the weight fractions of other oxides in Amp and Tur are usually 0, which will not affect the calculations of the principal components, their coefficients are not listed here.

PC	Al_2O_3	CaO	FeO	MgO	Na_2O	SiO_2
1	0.88	−0.34	−0.11	−0.21	−0.05	−0.22
2	−0.09	0.01	0.81	−0.43	0.03	−0.37
3	−0.04	−0.69	0.36	0.14	−0.04	0.58
4	−0.14	−0.39	0.03	0.64	0.10	−0.61
5	−0.28	−0.33	−0.29	−0.45	−0.02	−0.06
6	0.07	0.11	0.04	0.07	−0.17	−0.05

and Amp samples is quite evident) and extract the optical properties (Aligholi et al., 2015; Izadi et al., 2017; Li et al., 2017) as auxiliary features, which will be part of our future works. Another possible solution is to utilize the content of the boron (B) element to distinguish them. However, due to the chemical shift, the quantitative analysis result for the B element is, unfortunately, less accurate using our current techniques. The peak position of the B element varies from mineral to mineral during our testing. Since the peak is usually broader and overlaps with the adjacent peaks, the B element and its accurate content cannot be detected effectively. We are not able to achieve it adequately under existing laboratory conditions, but we plan to improve this in our future work.

6. Conclusion

In this paper, we propose a Multi-Task Multi-Class learning method MTMC for the identification of heavy minerals, to handle the problem of the insufficiency of identified heavy minerals from the individual river basin. MTMC makes use of all the identified heavy minerals from multiple river basins to jointly train the classifiers for the individual river basin, in order to improve the performance for each river basin. To avoid the problems of class imbalance and time overhead introduced by the common "One vs. One" and "One vs. Rest" strategies, MTMC uses the softmax classifier for the multi-class classification of heavy minerals. Based on the heavy minerals collected from the three river basins in China, we conduct the experiments to evaluate the performance of MTMC. The results show that MTMC consistently outperforms the available single task and multi-task learning methods for the identification of heavy minerals.

Using MTMC, most heavy mineral classes are identified correctly, while some mineral classes, such as Tur and Amp, are misidentified on certain river basins. In our future work, we will enhance our method by incorporating the optical properties of heavy minerals using the microscopic images. We will also try to detect the boron element more accurately, to distinguish Amp and Tur. In addition, we plan to apply deep learning techniques, which can automatically emulate end to end optical properties, to improve the accuracy of the identification of heavy minerals.

Computer code availability

- Name of code: MTMC
- Developer: Na Li
- E-mail: li.na@smail.nju.edu.cn
- Year first available: 2019
- Software required: MATLAB 2017a
- Program language: MATLAB
- Program size: 294kb
- Details on how to access the source code: the code is available at https://github.com/ln-nju/MTMC.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Table B.6
The F1 measurements of different single task learning methods for the identification of heavy minerals. The method abbreviations include SVM (Support Vector Machine), LDA (Linear Discriminant Analysis), NB (Naïve Bayes), 1NN (Nearest Neighbor), DT (Decision Tree), RUSBoost (Random Under Sampling), Bagging (Bootstrap Aggregation), and LPBoost (Linear Programming Boost).

	Yangtse		Yarlung		PumQu	
	Yarlung	PumQu	Yangtse	PumQu	Yangtse	Yarlung
SVM	0.72	0.84	0.74	0.78	0.85	0.80
LDA	0.82	0.78	0.92	0.92	0.85	0.87
NB	0.30	0.47	0.45	0.58	0.35	0.64
1NN	0.77	0.81	0.80	0.74	0.87	0.82
DT	0.74	0.72	0.85	0.76	0.78	0.75
AdaBoost	0.87	0.71	0.89	0.72	0.83	0.78
RUSBoost	0.78	0.62	0.86	0.66	0.60	0.78
Bagging	0.78	0.76	0.92	0.83	0.90	0.83
LPBoost	0.86	0.74	0.91	0.86	0.74	0.78
TotalBoost	0.80	0.72	0.88	0.83	0.81	0.75
MTMC	0.95 ± 0.01		0.98 ± 0.01		1.00 ± 0.00	

CRediT authorship contribution statement

Na Li: Conceptualization, Methodology, Software, Writing - original draft. **Huizhen Hao:** Writing - review & editing, Funding acquisition. **Zhiwei Jiang:** Writing - review & editing, Funding acquisition. **Feng Jiang:** Writing - review & editing. **Ronghua Guo:** Writing - review & editing. **Qing Gu:** Writing - review & editing, Funding acquisition. **Xiumian Hu:** Resources, Data curation, Funding acquisition.

Acknowledgments

We are grateful for helpful advice and constructive suggestions from the editor and the five anonymous reviewers. This study was supported by the Collaborative Innovation Center of Novel Software Technology and Industrialization, the Second Tibetan Plateau Scientific Expedition and Research Program (STEP), Ministry of Science and Technology, China (2019QZKK020604), and National Natural Science Foundation of China Project (61972192, 61906085, 41972111).

Appendix A. Oxide list

The 55 oxides we considered in this paper include: Ag_2O, Al_2O_3, As_2O_3, Au_2O_3, BaO, Br_2O_5, CaO, Ce_2O_3, CoO, Cr_2O_3, Cs_2O, Dy_2O_3, Er_2O_3, Eu_2O_3, FeO, Gd_2O_3, HfO_2, HgO, In_2O_3, IrO_2, K_2O, La_2O_3, MgO, MnO, Na_2O, NaO, Nb_2O_5, Nd_2O_3, NiO, OsO_2, P_2O_5, Pm_2O_3, Pr_2O_3, PtO_2, Rb_2O, Ru_2O_3, Sc_2O_3, SiO_2, Sm_2O_3, SnO_2, SO_3, SrO, Ta_2O_5, TeO_2, ThO_2, TiO, TiO_2, Tl_2O, UO_3, V_2O_5, WO_3, Y_2O_3, Yb_2O_3, ZrO_2, ZnO.

Appendix B. Additional experiment

We conduct an additional experiment to evaluate the soundness of our method. We take each pair of river basins (tasks), train a single task learning classifier on the heavy mineral samples from one river basin, and evaluate its prediction performance on the other. Note that, the classifier trained on the Yangtse River only identifies 8 heavy mineral classes. We compare their prediction performance with MTMC.

Table B.6 lists the results, where the first row gives the target rivers and the second row gives the river of training. Compared with the performance measurements by the river's self training set(in Table 3), most obtain a worse prediction performance using heavy mineral samples from other river basins for training. This demonstrates that simply using additional data from other river basins as supplements does not help improve the prediction performance when the self sampled heavy minerals are insufficient. It is worth noting that MTMC consistently outperforms all the single task learning methods under the above settings, which proves the fitness of MTMC in the identification of heavy minerals collected from multiple river basins.

References

Akkaş, E., Akin, L., Çubukçu, H.E., Artuner, H., 2015. Application of decision tree algorithm for classification and identification of natural minerals using sem–eds. Comput. Geosci. 80, 38–48.

Akkas, E., Evren Çubukçu, H., Akin, L., Erkut, V., Yurdakul, Y., Karayigit, A.I., 2016. Identification of some zeolite group minerals by application of artificial neural network and decision tree algorithm based on sem-eds data. In: EGU General Assembly Conference Abstracts, vol. 18.

Aligholi, S., Khajavi, R., Razmara, M., 2015. Automated mineral identification algorithm using optical properties of crystals. Comput. Geosci. 85, 175–183.

Aligholi, S., Lashkaripour, G.R., Khajavi, R., Razmara, M., 2017. Automatic mineral identification using color tracking. Pattern Recognit. 65, 164–174.

Andò, S., Garzanti, E., 2014. Raman spectroscopy in heavy-mineral studies. Geol. Soc. Lond. Spec. Publ. 386 (1), 395–412.

Beck, A., Teboulle, M., 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imag. Sci. 2 (1), 183–202.

Borges, H.P., de Aguiar, M.S., 2019. Mineral classification using machine learning and images of microscopic rock thin section. In: Mexican International Conference on Artificial Intelligence. Springer, pp. 63–76.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24 (2), 123–140.

Chen, J., Liu, J., Ye, J., 2012. Learning incoherent sparse and low-rank patterns from multiple tasks. ACM Trans. Knowl. Discov. Data 5 (4), 22.

Chen, J., Zhou, J., Ye, J., 2011. Integrating low-rank and group-sparse structures for robust multi-task learning. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 42–50.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297.

Cover, T.M., Hart, P.E., et al., 1967. Nearest neighbor pattern classification. IEEE Trans. Inform. Theory 13 (1), 21–27.

Demiriz, A., Bennett, K.P., Shawe-Taylor, J., 2002. Linear programming boosting via column generation. Mach. Learn. 46 (1–3), 225–254.

Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Ann. Eugenics 7 (2), 179–188.

Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55 (1), 119–139.

Gallagher, M., Deacon, P., 2002. Neural networks and the classification of mineralogical samples using x-ray spectra. In: Proceedings of the 9th International Conference on Neural Information Processing, 2002 ICONIP'02, 5. IEEE, pp. 2683–2687.

Gong, P., Ye, J., Zhang, C., 2012. Robust multi-task feature learning. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 895–903.

Hao, H., Guo, R., Gu, Q., Hu, X., 2019. Machine learning application to automatically classify heavy minerals in river sand by using sem/eds data. Minerals Eng. 143, 105899.

Izadi, H., Sadri, J., Bayati, M., 2017. An intelligent system for mineral identification in thin sections based on a cascade approach. Comput. Geosci. 99, 37–49.

Jalali, A., Sanghavi, S., Ruan, C., Ravikumar, P.K., 2010. A dirty model for multi-task learning. In: Advances in Neural Information Processing Systems. pp. 964–972.

Jolliffe, I., 2011. Principal Component Analysis. Springer.

Li, N., Hao, H., Gu, Q., Wang, D., Hu, X., 2017. A transfer learning method for automatic identification of sandstone microscopic images. Comput. Geosci. 103, 111–121.

Lünsdorf, N.K., Kalies, J., Ahlers, P., Dunkl, I., von Eynatten, H., 2019. Semi-automated heavy-mineral analysis by raman spectroscopy. Minerals 9 (7), 385.

Morton, A.C., Hallsworth, C., 1994. Identifying provenance-specific features of detrital heavy mineral assemblages in sandstones. Sediment. Geol. 90 (3–4), 241–256.

Mounteney, I., Burton, A., Farrant, A., Watts, M., Kemp, S., Cook, J., 2018. Heavy mineral analysis by icp-aes a tool to aid sediment provenancing. J. Geochem. Explor. 184, 1–10.

Nigam, K., McCallum, A.K., Thrun, S., Mitchell, T., 2000. Text classification from labeled and unlabeled documents using em. Mach. Learn. 39 (2–3), 103–134.

Obozinski, G., Taskar, B., Jordan, M., Multi-task feature selection. Tech. Rep 2. Statistics Department, UC Berkeley.

Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 1 (1), 81–106.

Ruisanchez, I., Potokar, P., Zupan, J., Smolej, V., 1996. Classification of energy dispersion x-ray spectra of mineralogical samples by artificial neural networks. J. Chem. Inform. Comput. Sci. 36 (2), 214–220.

Seiffert, C., Khoshgoftaar, T.M., Van Hulse, J., Napolitano, A., 2009. Rusboost: A hybrid approach to alleviating class imbalance. IEEE Trans. Syst. Man Cybern. A 40 (1), 185–197.

Shimizu, M., Sano, N., Ueki, T., Komatsu, T., Yasue, K.-i., Niwa, M., 2019. Provenance identification based on epma analyses of heavy minerals: Case study of the toki sand and gravel formation, central japan. Island Arc 28 (2), e12295.

Sun, J., Zhang, Y., Mao, H., Cong, S., Wu, X., Wang, P., 2018. Research of moldy tea identification based on rf-rfe-softmax model and hyperspectra. Optik 153, 156–163.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1), 267–288.

Warmuth, M.K., Liao, J., Rätsch, G., 2006. Totally corrective boosting algorithms that maximize the margin. In: Proceedings of the 23rd International Conference on Machine Learning. ACM, pp. 1001–1008.

Yousefi, B., Sojasi, S., Castanedo, C.I., Beaudoin, G., Huot, F., Maldague, X.P., Chamberland, M., Lalonde, E., 2016. Mineral identification in hyperspectral imaging using sparse-pca. In: Thermosense: Thermal Infrared Applications XXXVIII, 9861. International Society for Optics and Photonics, 986118.