


Article

Soft Measurement of Rare Earth Multi-Element Component Content Based on Multi-LightVGG Modeling

Zhen Li ¹, Jun Xiao ², Qihan Zhang ³, Kunming Liu ^{2,*}  and Jinhui Li ^{2,*}

¹ College of Rare Earth, Jiangxi University of Science and Technology, 86 Hongqi Road, Ganzhou 341000, China; leeminchen@163.com

² College of Chemistry and Chemical Engineering, Jiangxi University of Science and Technology, 86 Hongqi Road, Ganzhou 341000, China; 17679326645@163.com

³ School of Information Engineering, Jiangxi University of Science and Technology, 86 Hongqi Road, Ganzhou 341000, China; zhangqihan0722@163.com

* Correspondence: liukunming@jxust.edu.cn (K.L.); jinhuili@jxust.edu.cn (J.L.)

Abstract: The current hardware equipment used to detect the content of each element component in the rare earth extraction process has a complex structure and high maintenance cost. A modeling method for the soft measurement of rare earth multi-element component content is proposed to address this issue. This method uses the Multi-LightVGG multi-tasking learning model and the Multi Gradient Descent Algorithm based on Optimized Upper Bound (MGDA-OUB) to optimize the model for each prediction task and find the Pareto optimal solution. After conducting several experiments, the Multi-LightVGG model loaded with MGDA-OUB has lower MRE, RMSE for Pr, Nd prediction, and MAX(|error|) for Nd prediction than the Multi-LightVGG model without MGDA-OUB by 0.3778%, 0.5208%, 0.0015, 0.0015, and 0.1985%, respectively; and the MRE and RMSE of the Multi-LightVGG model for Pr and Nd prediction under the same optimization conditions are lower than those of Multi-ResNet18 by 0.3297%, 0.5423%, 0.0019, and 0.002, respectively, thus indicating that MGDA-OUB can effectively solve multiple task-specific Pareto solutions to avoid possible conflicts between specific tasks, while the Multi-LightVGG model, compared to the Multi-Resnet18 model, has a backbone network that can effectively capture the abstract representations in the images of the rare earth-extraction mixed solution, which in turn improves the prediction accuracy of the content of each elemental component.

Keywords: soft measurement; rare earth; multi-element component; multi-lightVGG



Citation: Li, Z.; Xiao, J.; Zhang, Q.; Liu, K.; Li, J. Soft Measurement of Rare Earth Multi-Element Component Content Based on Multi-LightVGG Modeling. *Minerals* **2023**, *13*, 1491. <https://doi.org/10.3390/min13121491>

Academic Editor: Manuel Jesús Gázquez

Received: 25 September 2023

Revised: 16 November 2023

Accepted: 21 November 2023

Published: 28 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Rare earths are essential strategic resources that are widely used in the electronics, high-tech, national defense, and military industries. They are specifically utilized in manufacturing new energy vehicles, guided missiles, computers, and other related fields [1–3]. However, rare earth elements share similar physicochemical properties, and their neighboring elements have small separation coefficients [4,5]. As a result, separating rare earth elements with high purity is challenging. The tandem-stage extraction theory [6–8] has been proposed to address this issue, which significantly enhances the purity and yield of separated rare earths. In the rare earth industry, the theory of tandem extraction [9] guides the establishment of a rare earth tandem extraction separation process. This process comprises an n-stage extraction section and an m-stage washing section. The treated rare earth raw material liquid is added from the n-th stage and then separated by stirring with a motor. The easy-to-extract components are gradually distributed to the organic phase of the extraction tanks at each stage, while the difficult-to-extract components are gradually deposited into the aqueous phase. After the multi-stage extraction tank, the difficult-to-extract product with the component content of can be obtained from the 1st extraction tank, and the easy-to-extract product with the component content of can be obtained from the n + m

extraction tank, as shown in Figure 1. A monitoring point can be set up at the extraction stage in the above separation process where the component content of the extraction section and the washing section is sensitive to changes. This monitoring point provides the current component content value of each element in the extraction tank, which can be used for theoretical calculations to obtain the theoretical minimum extraction volume and minimum washing volume required for the separation. This information can then be used to optimize the extraction process to ensure a high purification of the separated product.

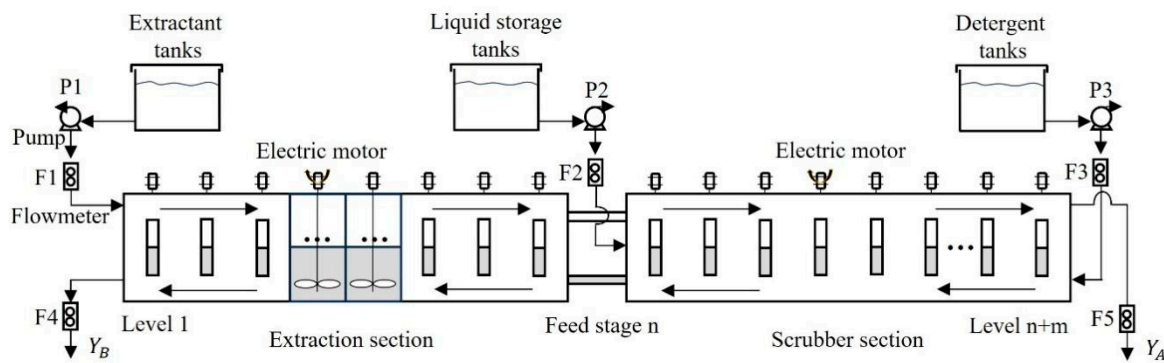


Figure 1. Schematic diagram of rare earth solvent extraction process.

Rare earth-extraction production sites currently use the offline assay method to determine the component content value of the extraction tank. However, this method has shortcomings and cannot provide real-time feedback on the effectiveness of extraction and separation. Existing detection methods, such as X-ray fluorescence analysis and spectrophotometry, are not practical due to their complex hardware device structures and high cost of use and maintenance. Therefore, it is essential to develop a non-contact rare earth element component content soft measurement method that is simple to operate and has low monitoring costs.

Current soft-measurement methods can predict the elemental component content of the extraction tank individually [10–17], but they do not consider possible correlations between the contents of multiple elemental components. With the rise of deep learning methods in machine learning, many researchers have used deep learning methods for image classification in various research fields and achieved excellent results [18–20]. Multi-task learning methods [21–23] in deep learning methods, on the other hand, allow for the joint training of multiple tasks, aiming to improve the generalization ability by exploiting specific features contained in the training signals of the associated tasks. Compared to single-task learning, multi-task learning has several advantages:

- (1) Multiple tasks can share a shared layer network, significantly reducing memory footprint.
- (2) Avoiding double-computation of features in the shared-layer network improves the speed of model training.
- (3) If related tasks share complementary information, it has the potential to improve the overall performance of the model.

There are different methods used for multi-task learning based on multi-task optimization strategies. Gradient normalization, proposed by Chen et al. [24], balances the gradient and rate of multi-task network training and encourages the network to learn all tasks at an equal rate. However, the method may not work well when the magnitude of different tasks varies. Uncertainty weighting, proposed by Kendall et al. [25], uses the noise parameter σ to balance the weights among specific tasks by acting on the loss function. In contrast, Désidéri et al. [26] viewed multi-task learning as a multi-objective optimization problem. They proposed the Multiple Gradient Descent Algorithm (MGDA) to reduce the value of the loss for any particular task without increasing the value of other losses. This is achieved by finding the Pareto smoothing point among different tasks. Sener et al. [27] built on the research of the multiple gradient descent algorithm by proposing the

multi-objective loss upper bound. This method utilizes the upper-bound-based Multiple Gradient Descent Algorithm (Multiple Gradient Descent Algorithm—Upper Bound) to find the Pareto optimal solution among all tasks. It greatly reduces the training iteration time and achieves similar or even better test performance than the original method through a single inverse computation process without the need to explicitly specify the task-specific gradient. Both the multi-task-based model structure and the multi-task-based optimization strategy can be used together to improve the model's overall performance.

The work of Zhou et al. [28] and Zhao et al. [29], involved the use of multiple gradient descent algorithms in their respective research. Zhou et al. proposed an end-to-end license plate recognition method while Zhao et al. achieved good prediction results in identifying working conditions in the froth flotation process. Meanwhile, Zhang et al. established a multi-task learning-based rare earth multi-element component content and concentration prediction model. In their study, they explored the existence of a commonality between the component content and concentration of rare earth elements, or between the component contents. They proposed a multiple gradient descent algorithm based on the optimization of the upper bound for optimizing the above model, and the experimental results demonstrated that there is a good commonality between the component contents of the rare earth elements. However, the previous studies only compared different multi-task optimization methods and did not innovate the backbone network. Thus, we innovatively propose the Multi-LightVGG model, based on the previous work, for training and predicting the component content values of Pr/Nd. We also used MGDA-OUB to optimize the Multi-LightVGG model. Our repeated experiments showed that MGDA-OUB can effectively improve the accuracy of the rare earth multi-element component content prediction model for Pr/Nd component content. We found that Multi-LightVGG has higher prediction accuracy than Multi-Resnet18, which provides a new idea for the soft measurement method of rare earth element component content.

In the Section 2, this paper introduces the basic architecture of the Multi-LightVGG network and explains the optimization algorithm MGDA-OUB. The Section 3 provides a detailed description of the process of creating an image dataset of Pr/Nd mixed solution. The prediction results of the Multi-LightVGG model for the component contents of Pr and Nd are then compared with those of other multi-tasking models, and the prediction error evaluation indexes are presented in the form of charts and graphs. Finally, the conclusions are drawn in the Section 4.

2. Rare Earth Multi-Element Component Content Prediction Model

In this chapter, the basic architecture of the Multi-LightVGG model and the definition and algorithmic flow of MGDA-OUB are specifically introduced.

2.1. Multi-LightVGG Modeling

Due to the good commonality between the component contents of different rare earth elements, they can promote each other in multi-task learning model training to improve the overall prediction accuracy of the model, and the lightweight VGG model [30] is outstanding in processing the images of mixed rare earth extraction solutions. Therefore, this paper establishes a multi-task learning model based on lightweight VGG, named Multi-LightVGG, in which the backbone network is used to extract the abstract representation of the image to the shared-layer network, and branches out multiple specific network layers for outputting the component content values of multiple specific elements to be measured. The specific architecture of the model is shown in Figure 2.

The backbone network of the Multi-LightVGG model is LightVGG, and its specific network parameters are shown in Table 1, which reduces the input image of the rare earth mixed extraction solution with a size of $224 \times 224 \times 3$ to a $7 \times 7 \times 512$ digital matrix and outputs it to its respective task-specific layers, where each task-specific layer establishes a three-layer fully connected operation in which the number of fully connected hidden nodes

in the first and second layers is 1024, and the number of fully connected hidden nodes in the third layer is 121.

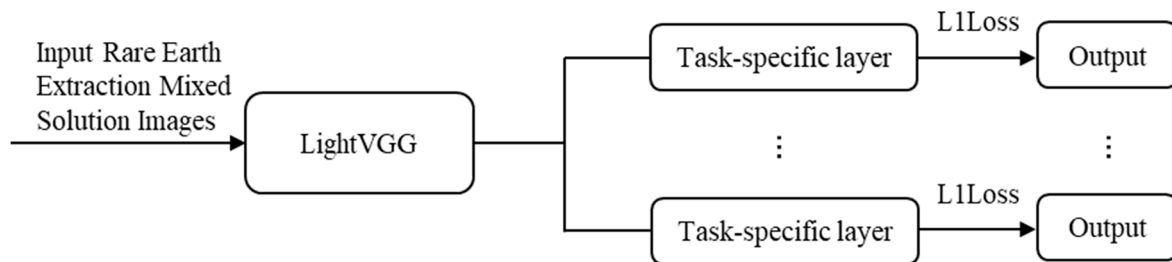


Figure 2. Basic architecture of Multi-LightVGG model.

Table 1. Multi-LightVGG backbone network parameters (minus end fully connected layers).

Layers	Layer Type	Channel Number	Filter Size	Slide
1	Conv	64	3 × 3	1 × 1
2	Pool	64	2 × 2	2 × 1
3	Conv	64	3 × 3	1 × 1
4	Pool	64	2 × 2	2 × 1
5	Conv	64	3 × 3	1 × 1
6	Conv	64	3 × 3	1 × 1
7	Pool	64	2 × 2	2 × 1
8	Conv	64	3 × 3	1 × 1
9	Conv	64	3 × 3	1 × 1
10	Pool	64	2 × 2	2 × 1
11	Conv	512	3 × 3	1 × 1
12	Conv	512	3 × 3	1 × 1
13	Pool	512	2 × 2	2 × 1

In particular, in order to obtain a single accurate value of rare earth element component content, this paper connects the output of the fully connected layer in the forward propagation process in the task-specific layer network to the Softmax function [31], and uses the Softmax function to establish a regression model between the outputs of the probabilities of each category of the task-specific categories and the true values of each category through the linear regression loss function, and then back-propagates the value of the loss function to optimize the network parameters. The loss functions commonly used in linear regression models are L1Loss and L2Loss to compute the mean absolute error *MAE* (Equation (1)) and the mean square error *MSE* (Equation (2)), respectively.

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \tag{1}$$

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \tag{2}$$

In the above equations, where y_i is the model predicted value, \hat{y}_i is the true label value of the task corresponding to the predicted data point, and n is the number of predicted data points. Compared to L1Loss, the L2Loss loss function squares the error and exacerbates the neglect of small errors. Therefore, we adopted L1Loss as the loss function for each model.

2.2. Multi-Objective Optimization Algorithm MGDA-OUB

During the process of training the Multi-LightVGG model for predicting the component content of multiple rare earth element fractions, conflicts may occur between multiple specific tasks, resulting in a situation where the parameters of the shared-layer network are biased towards one specific task. MGDA-OUB is used to balance the competition

between tasks by finding the Pareto [32] solution between multiple specific tasks, so that the shared layer network parameters are not biased towards a specific task, thus improving the accuracy of prediction for each specific task.

2.2.1. Multi-Objective Optimization Definition

Given that the input space \mathcal{X} in the rare earth multi-element component content prediction problem is the collected rare earth mixed extraction solution images, $\{\mathcal{Y}^t\}_{t \in [T]}$ is the space of the real values of the corresponding component contents in a set of rare earth mixed extraction solution images, and the data points in the whole rare earth mixed extraction solution image dataset are $\{x_i, y_i^1, \dots, y_i^T\}_{i \in [N]}$, where T is the number of rare earth element component content tasks to be predicted, N is the number of rare earth mixed extraction solution image data points, and y_i^t is the true value label of the i -th data point corresponding to the t -th component content prediction task. Further, the parameterization of each specific task is considered in the multi-task learning model, which is assumed to be $f^t(x; \theta^{sh}, \theta^t) : \mathcal{X} \rightarrow \mathcal{Y}^t$, where θ^{sh} is the shared-layer network parameters and θ^t is the task-specific layer network parameters. Then, the empirical minimization formula for all rare earth multi-element component content prediction tasks can be expressed as Equation (3):

$$\min_{\theta^{sh}, \theta^1, \dots, \theta^T} \sum_{t=1}^T c^t \hat{\mathcal{L}}^t(\theta^{sh}, \theta^t) \tag{3}$$

where c^t is the weight calculated for the particular task being measured; $\hat{\mathcal{L}}^t(\theta^{sh}, \theta^t)$ is the empirical loss function for the t th prediction task, which can be specifically defined as:

$$\hat{\mathcal{L}}^t(\theta^{sh}, \theta^t) \triangleq \frac{1}{N} \sum_i \mathcal{L}(f^t(x_i; \theta^{sh}, \theta^t), y_i^t) \tag{4}$$

Assuming that there are two sets of parameter solutions θ and $\bar{\theta}$ in Equation (3), the following problem may arise in the condition that there are two specific tasks, t_1 and t_2 : $\hat{\mathcal{L}}^{t_1}(\theta^{sh}, \theta^{t_1}) < \hat{\mathcal{L}}^{t_1}(\bar{\theta}^{sh}, \theta^{t_1})$ and $\hat{\mathcal{L}}^{t_2}(\theta^{sh}, \theta^{t_2}) > \hat{\mathcal{L}}^{t_2}(\bar{\theta}^{sh}, \theta^{t_2})$, i.e., the parameter solution θ is more biased towards task t_1 while $\bar{\theta}$ is more biased towards task t_2 . To solve the problem, in this paper, we consider the learning of multiple prediction tasks together as a multi-objective optimization problem and look for the Pareto solution that balances the multiple tasks such that the network parameters are not biased towards any particular task. Then, the above problem can be transformed into finding the parameter solution of the following equation:

$$\min_{\theta^{sh}, \theta^1, \dots, \theta^T} L(\theta^{sh}, \theta^1, \dots, \theta^T) = \min_{\theta^{sh}, \theta^1, \dots, \theta^T} (\hat{\mathcal{L}}^1(\theta^{sh}, \theta^1), \dots, \hat{\mathcal{L}}^T(\theta^{sh}, \theta^T)) \tag{5}$$

Equation (5) consists of a vector L , containing the loss functions corresponding to the multiple tasks for rare earth element multicomponent content prediction, to seek a Pareto solution that balances between multiple tasks. The definition of the Pareto solution between multiple tasks in multi-objective optimization is given below:

Definition 1. For all tasks t with $\hat{\mathcal{L}}^t(\theta^{sh}, \theta^t) \leq \hat{\mathcal{L}}^t(\bar{\theta}^{sh}, \bar{\theta}^t)$, $\forall t \in (1, 2, \dots, T)$, and $L(\theta^{sh}, \theta^1, \dots, \theta^T) \neq L(\bar{\theta}^{sh}, \bar{\theta}^1, \dots, \bar{\theta}^T)$, then the parametric solution θ is said to dominate $\bar{\theta}$.

Definition 2. A parametric solution θ^* is said to be Pareto optimal if there exists no parametric solution θ dominating θ^* .

2.2.2. Flow of MGDA-OUB

MGDA-OUB is used to seek the Pareto optimal solution in the prediction of rare earth multi-element component content, and the multi-objective optimization is accomplished by gradient descent. In the following, multi-objective optimization is implemented using the Karush–Kuhn–Tucker (KKT) condition, which is also a necessary condition for multi-objective optimization. The KKT conditions required for the shared layer network parameters and task-specific layer network parameters in the rare earth multi-element component content prediction model are expressed as follows:

- (1) There exists $\alpha^1, \dots, \alpha^T \geq 0$ with $\sum_{t=1}^T \alpha^t = 1$ and $\sum_{t=1}^T \alpha^t \nabla_{\theta^{sh}} \hat{\mathcal{L}}^t(\theta^{sh}, \theta^t) = 0$;
- (2) For any rare earth element component content prediction task t , $\nabla_{\theta^t} \hat{\mathcal{L}}^t(\theta^{sh}, \theta^t) = 0$.

Any solution that satisfies the above conditions is called a Pareto smooth point, where the loss function used in this paper’s model in accomplishing the above optimization conditions is L1Loss, which is mentioned in Section 2.2. Consider again the following optimization problem:

$$\min_{\alpha^1, \dots, \alpha^T} \left\{ \left\| \sum_{t=1}^T \alpha^t \nabla_{\theta^{sh}} \hat{\mathcal{L}}^t(\theta^{sh}, \theta^t) \right\|_2^2 \mid \sum_{t=1}^T \alpha^t = 1, \alpha^T \geq 0 \quad \forall t \right\} \tag{6}$$

Désidéri et al. showed that the solution to this optimization problem satisfies the KKT condition when the solution is 0; otherwise, the solution gives a descending direction that optimizes the task of predicting the content of all rare earth element components. However, the algorithm described in this problem requires the computation of $\nabla_{\theta^{sh}} \hat{\mathcal{L}}^t(\theta^{sh}, \theta^t)$, which requires backpropagation of the shared-layer network parameters associated with each specific task, and thus the gradient computation for each specific task needs to be obtained in T backpropagations before forward propagation.

In this regard, this paper proposes a multiple gradient descent algorithm based on optimized upper bounds, which, unlike the multiple gradient descent algorithm, optimizes the objective upper bounds and obtains all task-specific gradients in forward propagation with only one backpropagation. Optimizing the upper bound requires combining a shared representation function with a task-specific decision function, which can be defined by defining the hypothesis class constraint as the following equation:

$$f^t(x; \theta^{sh}, \theta^t) = (f^t(\cdot; \theta^t) \circ g(\cdot; \theta^{sh}))(x) = f^t(g(x; \theta^{sh}); \theta^t) \tag{7}$$

where g is the representation function shared by all tasks and f^t is the task-specific function that takes this representation as input. If this representation function is expressed as $Z = (Z_1, \dots, Z_N)$, where $Z_i = g(x_i; \theta^{sh})$, the upper bound can be expressed as the following equation, which is a direct result of the chain rule:

$$\left\| \sum_{t=1}^T \alpha^t \nabla_{\theta^{sh}} \hat{\mathcal{L}}^t(\theta^{sh}, \theta^t) \right\|_2^2 \leq \left\| \frac{\partial Z}{\partial \theta^{sh}} \right\|_2^2 \left\| \sum_{t=1}^T \alpha^t \nabla_Z \hat{\mathcal{L}}^t(\theta^{sh}, \theta^t) \right\|_2^2 \tag{8}$$

where $\left\| \frac{\partial Z}{\partial \theta^{sh}} \right\|_2^2$ is the Jacobi matrix paradigm for Z equivalent to θ^{sh} . The two ideal properties of this upper bound are:

- (1) $\nabla_Z \hat{\mathcal{L}}^t(\theta^{sh}, \theta^t)$ can compute all task-specific gradients in a single backpropagation;
- (2) $\left\| \frac{\partial Z}{\partial \theta^{sh}} \right\|_2^2$ is not an equation with respect to $\alpha^1, \dots, \alpha^T$ and, hence, can be removed when it is used as an optimization objective.

In order to obtain the approximate optimization solution, considering the above two ideal properties, the $\left\| \sum_{t=1}^T \alpha^t \nabla_{\theta^{sh}} \hat{\mathcal{L}}^t(\theta^{sh}, \theta^t) \right\|_2^2$ in Equation (6) can be replaced by the upper bound in Equation (8), and the term $\left\| \frac{\partial Z}{\partial \theta^{sh}} \right\|_2^2$ is deleted, and then the multi-tasking optimization can be transformed into the solving of the following equation:

$$\min_{\alpha^1, \dots, \alpha^T} \left\{ \left\| \sum_{t=1}^T \alpha^t \nabla_Z \hat{\mathcal{L}}^t(\theta^{sh}, \theta^t) \right\|_2^2 \mid \sum_{t=1}^T \alpha^t = 1, \alpha^T \geq 0 \quad \forall t \right\} \tag{9}$$

The algorithm for solving this equation is then called the multiple gradient descent algorithm based on the optimization upper bound, and Sener proposes that one of the following two conditions can be satisfied when $\frac{\partial Z}{\partial \theta^{sh}}$ is of full rank and $\alpha^1, \dots, \alpha^T$ is a solution to Equation (9):

- (1) $\sum_{t=1}^T \alpha^t \nabla_{\theta^{sh}} \hat{\mathcal{L}}^t(\theta^{sh}, \theta^t) = 0$ and the current multitask learning model parameter is Pareto smooth;
- (2) $\sum_{t=1}^T \alpha^t \nabla_{\theta^{sh}} \hat{\mathcal{L}}^t(\theta^{sh}, \theta^t)$ is the direction of descent for all target tasks;

Then, the algorithm finds a Pareto smooth point with negligible computational overhead.

Equation (9) belongs to the class of convex quadratic problems with linear constraints, and solving this optimization problem is equivalent to finding the minimum number of paradigm points in the convex package of the input point set. We first consider the case based on solving two tasks, the optimization problem can be defined as:

$$\min_{\alpha \in [0,1]} \left\| \alpha \nabla_Z \hat{\mathcal{L}}^1(\theta^{sh}, \theta^1) + (1 - \alpha) \nabla_Z \hat{\mathcal{L}}^2(\theta^{sh}, \theta^2) \right\|_2^2 \tag{10}$$

This is a one-dimensional quadratic function with an analytic solution with respect to α . In the following, we simplify $\nabla_Z \hat{\mathcal{L}}^1(\theta^{sh}, \theta^1)$ to θ and $\nabla_Z \hat{\mathcal{L}}^2(\theta^{sh}, \theta^2)$ to $\bar{\theta}$, and then the derivation of Equation (10) is solved for the solution of Equation (11) below.

$$\alpha = \begin{cases} 0, \theta^T \bar{\theta} \geq \bar{\theta}^T \bar{\theta} \\ 1, \theta^T \bar{\theta} \geq \theta^T \theta \\ (\bar{\theta} - \theta)^T \bar{\theta} / \|\theta - \bar{\theta}\|_2^2, \theta^T \bar{\theta} < \bar{\theta}^T \bar{\theta} \text{ and } \theta^T \bar{\theta} < \theta^T \theta \end{cases} \tag{11}$$

Although Equation (11) is only applicable to the solution of two tasks, in this paper, we will use the Frank–Wolfe algorithm proposed by Jaggi et al. [33] to use Equation (10) as a sub-process of the linear search to solve the constrained optimization problem for two and more tasks. The flow of the Frank–Wolfe algorithm is shown in Algorithm 1. The obtained $\alpha^1, \dots, \alpha^T$ is the solution of the multiple gradient descent algorithm based on the upper bound of optimization, through which the descent direction of all specific tasks can be optimized to find the Pareto optimum of the multi-objective of rare earth multi-elemental component content prediction.

Algorithm 1 FrankWolfeSolver

```

1: Procedure FankWolfeSolver
2: Initialize  $\alpha = (\alpha^1, \dots, \alpha^T) = (\frac{1}{T}, \dots, \frac{1}{T})$ 
3: Precompute  $M$  st.  $M_{i,j} = (\nabla_Z \hat{\mathcal{L}}^i(\theta^{sh}, \theta^j))^T (\nabla_Z \hat{\mathcal{L}}^i(\theta^{sh}, \theta^j))$ 
4: repeat
5:      $\hat{t} = \operatorname{argmin}_{\gamma} \sum_t \alpha^t M_{\gamma t}$ 
6:      $\hat{\gamma} = \operatorname{argmin}_{\gamma} ((1 - \gamma)\alpha + \gamma e_{\hat{t}})^T M ((1 - \gamma)\alpha + \gamma e_{\hat{t}})$ 
7:      $\alpha = (1 - \hat{\gamma})\alpha + \hat{\gamma} e_{\hat{t}}$ 
8: until  $\hat{\gamma} \sim 0$  or Number of Iterations Limit
9: return  $\alpha^1, \dots, \alpha^T$ 
10: end procedure

```

2.2.3. Prediction Flow of Rare Earth Multi-Element Component Content Based on MGDA-OUB

In the optimization process of multiple objectives for rare earth multi-element component content prediction, this paper adopts the multiple gradient descent algorithm based on the optimization upper bound to solve the gradient of each specific task in the model training, which optimizes the multi-objective upper bound, so that all the task-specific gradients can be obtained during forward propagation after only one back propagation. Thereafter, the gradients of each specific task are substituted into the Frank–Wolfe algorithm to find the multi-objective optimization solution, and the Pareto smooth points of multiple tasks regarding the prediction of rare earth multi-element component content are obtained to realize the multi-objective optimization. The detailed steps are described as follows:

Input: Input the standardized preprocessed images of rare earth mixed extraction solution into the Multi-LightVGG model, and load the image samples corresponding to the real label values. Initialize the shared-layer network parameters θ^{sh} and the task-specific layer network parameters θ^t ; set the loss function, maximum number of iterations, batch size, optimizer, and learning rate η to solve the multi-objective optimization problem in the shared-layer network.

Output: Predicted values of the content of each rare earth element component to be measured are output at each specific task level.

Step 1: Initialize the multi-objective optimization upper bound $\nabla_Z \hat{\mathcal{L}}^t(\theta^{sh}, \theta^t)$ according to Equation (9);

Step 2: Find the gradient $\nabla_{\theta^t} \hat{\mathcal{L}}^t(\theta^{sh}, \theta^t)$, ($t \in T$) for each particular task under the optimization upper bound;

Step 3: Substitute all task-specific gradients into the Frank–Wolfe algorithm for a multi-objective optimization solution to obtain Pareto smooth points $\alpha^1, \dots, \alpha^T$ for multiple tasks regarding the prediction of rare earth multi-elemental component content;

Step 4: Update the shared layer network parameters $\theta^{sh} = \theta^{sh} - \eta \sum_{t=1}^T \alpha^t \nabla_{\theta^{sh}} \hat{\mathcal{L}}^t(\theta^{sh}, \theta^t)$ using $\alpha^1, \dots, \alpha^T$;

Step 5: Update the task-specific layer network parameters $\theta^t = \theta^t - \eta \nabla_{\theta^t} \hat{\mathcal{L}}^t(\theta^{sh}, \theta^t)$, ($t \in T$) to optimize the descent direction for all specific tasks;

Step 6: Determine whether the maximum number of iterations is reached, if yes, go to step 7, otherwise return to step 1;

Step 7: According to the optimized shared layer network parameters θ^{sh} and specific layer network parameters θ^t , ($t \in T$), the content of each elemental component in the image of the mixed rare earth extraction solution to be detected is predicted.

The above steps are shown in Figure 3 below.

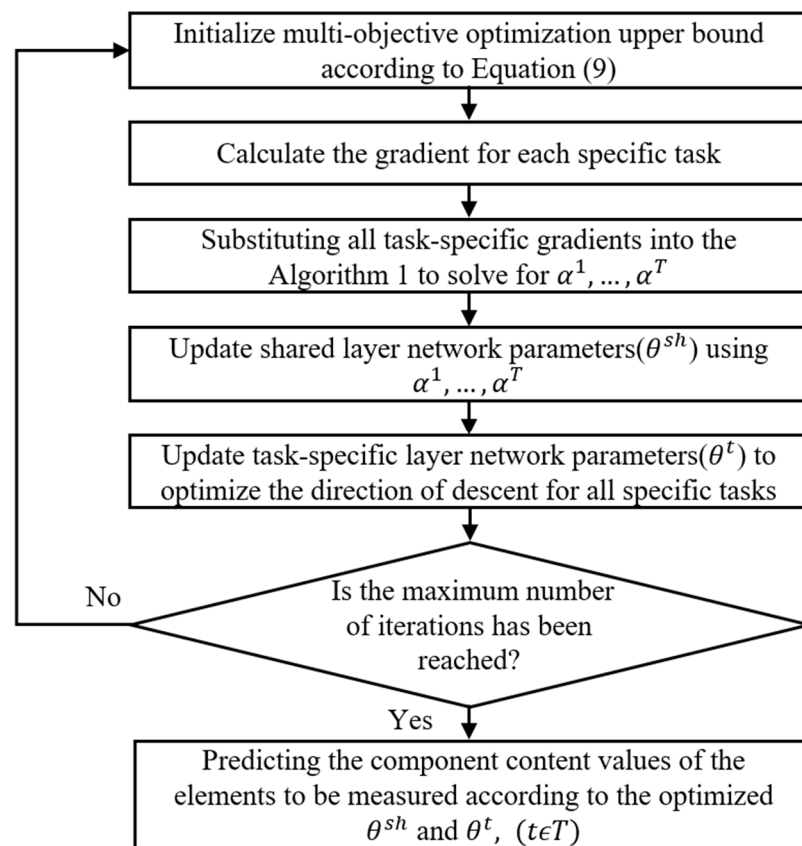


Figure 3. Flowchart of rare earth multi-element component content prediction based on MGDA-OUB.

3. Analog Simulation Experiment

3.1. Datasets

In this paper, the multi-elemental content prediction of rare earths based on the Multi-LightVGG model is simulated with the mixed extraction solution of Pr (apple green) and Nd (violet-red), which are two kinds of elements with ionic color characteristics, for example, and the mixed extraction solution of Pr/Nd is obtained through the following ways:

- (1) Measurement of the component content of the original solution: 1 L of 1.8275 mol/L PrCl_3 and 1 L of 2.063 mol/L NdCl_3 were purchased from a rare earth company in Ganzhou City, Jiangxi Province, China, with a purity of 99.9%, and the rare earth concentrations and compositions were provided by China National Center for Supervision and Inspection of Tungsten and Rare Earth Product Quality.
- (2) Solution dilution: the concentrations of the original solution of the two elements were diluted into 11 different concentrations of 0.01 mol/L to 0.50 mol/L each with good light transmission of the pure solutions of rare earth extraction.
- (3) Solution mixing: 50 mL of each of the two rare earth element solutions with different concentrations were mixed with each other to obtain a total of 121 groups of Pr/Nd mixed extraction solutions with different component contents and concentrations.

From the above steps, 121 groups of rare earth mixed extraction solutions were obtained with the content of each elemental component of Pr/Nd varying from 1.96% to 98.04% and the concentration varying from 0.005 mol/L to 0.25 mol/L. The rare earth mixed extraction solutions with different component content and concentration of each component were poured into a collection dish and sealed for storage, and some of the image examples are shown in Figure 4. As the laboratory-prepared rare earth mixed extraction solution has the actual characteristics of presenting certain reflectivity and refractivity to the light, the light transmittance is better, which meets the optical imaging conditions of the soft measurement of the component content of rare earth elements based on the machine

vision experiment. At the same time, the “ion color band” formed in the container filled with the rare earth mixture solution provides a feasible way to adopt fast, accurate, and continuously detectable image recognition technology.

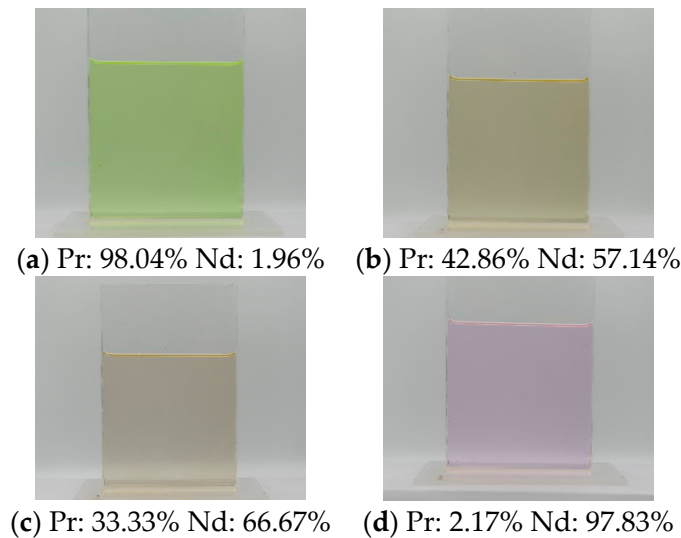


Figure 4. Images of some mixed solutions of Pr/Nd with different component contents and concentrations.

The experimental images were acquired by pouring the rare earth mixed extraction solution into a quartz container with a length, width, and height of $150 \times 5 \times 170$ mm until the solution filled the container. The container was then placed in a 60 CM studio. In the studio, two LED light sources were set up, with an output voltage of 24 V, power of 48 W, and maximum lumen of 15,000 LM; the background was pure white; the image acquisition equipment was a NIKON D700 camera; and the final image was captured as a JPG image with a resolution of 4256×2832 .

Since the Pr/Nd mixed extraction solution images obtained contain non-solution parts outside the edge of the quartz container and uneven color parts inside the edge, the color-filled parts of the captured images were cropped, and 10 pictures were cropped in the order of top-to-bottom and left-to-right for each group of solution images, and a total of 1210 images of rare earth mixed extraction solutions with uniform colors were obtained. The images of the rare earth mixed extraction solutions were categorized into 121 categories according to the component content and concentration of each element of Pr and Nd, and each category was labeled with the real label values of the component content and concentration of Pr and Nd, respectively. Then, 70% of them were divided into a training set, 20% into a validation set, and 10% into a test set, which was prepared into a complete dataset to be used for the multi-element component content of rare earth multi-element extraction solution modeling based on the Multi-LightVGG model for rare earth multi-element component content prediction.

3.2. Comparative Experiments

In order to verify the effectiveness of the multi-objective optimization algorithm MGDA-OUB and the multi-task learning model with LightVGG as the backbone network proposed in this paper, firstly, the Multi-LightVGG model loaded with and unloaded with MGDA-OUB was compared, which was used to validate the effectiveness of MGDA-OUB. Thereafter, the multi-task learning models with LightVGG and ResNet18 as the backbone network were both loaded with MGDA-OUB and compared to validate the superiority of the Multi-LightVGG model proposed in this paper for the soft measurements of the content of rare earth multi-element components.

In order to carry out the above comparison experiments, this paper randomly selected 10 images of the test set of mixed rare earth extraction solutions, and used the above model

to predict the Pr and Nd component content values of each of the 10 image samples, then compared the predicted values with the real values of *MRE* (mean relative error), *RMSE* (root mean square error), and *MAX(|error|)* (the maximum absolute value of relative error) as evaluation indexes to compare the models. The above formulae are shown in Equations (12)–(14):

$$MRE = \frac{1}{n} \sum_{i=1}^n \left(\frac{|y_i - \hat{y}_i|}{\hat{y}_i} \right), (i = 1, 2, \dots, n) \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (13)$$

$$MAX(|error|) = \max \left| \frac{y_i - \hat{y}_i}{\hat{y}_i} \right| \quad (14)$$

where y_i is the model predicted value, \hat{y}_i is the true label value of the task corresponding to the predicted data point, and n is the number of predicted data points.

Uniform hyperparameters were used in the training process of each model, and the specific settings are shown in Table 2. All model training and prediction experiments were realized in the following experimental environments: the hardware environment was the Windows 10 operating system, CPU Intel Core i7-12700F (12 cores), and GPU RTX3080 (with 10 G of RAM); and the software environment was the PyCharm used on the PyTorch deep learning framework.

Table 2. Hyperparameters during training of each model.

Main Parameters	Parameter Settings
Number of Iterations	300
Batch Size	32
Optimizer	Adam
Learning Rate	Initially 1×10^{-3} , halved every 30 iterations

The Multi-LightVGG models loaded and unloaded with MGDA-OUB were first compared and the experimental results are shown by Figure 5 and Table 3.

Table 3. Error evaluation index values for Pr and Nd prediction by the Multi-LightVGG model loaded and unloaded with MGDA-OUB.

Name of the Model	MRE/%		RMSE		MAX(Error)/%	
	Pr	Nd	Pr	Nd	Pr	Nd
Unloaded MGDA-OUB	1.8656	2.1729	0.0104	0.0105	3.0300	4.8249
Load MGDA-OUB	1.4878	1.6521	0.0089	0.0090	3.4287	4.6264

From Figure 5 and Table 3, it can be seen that the Multi-LightVGG model loaded with MGDA-OUB has a lower MRE, RMSE for Pr, Nd prediction, and MAX(|error|) for Nd prediction than the Multi-LightVGG model without MGDA-OUB by 0.3778%, 0.5208%, 0.0015, 0.0015 and 0.1985%, respectively, which proves that MGDA-OUB can optimize the Multi-task Learning. This proves that MGDA-OUB can effectively optimize the multi-task learning model to seek Pareto solutions for multiple specific tasks, thus avoiding possible conflicts between specific tasks, and improving the prediction accuracy of each specific task in multi-task joint training. In the following, Multi-ResNet18 and Multi-LightVGG, both loaded with MGDA-OUB, were compared, and the experimental results are shown in Figure 6 and Table 4.

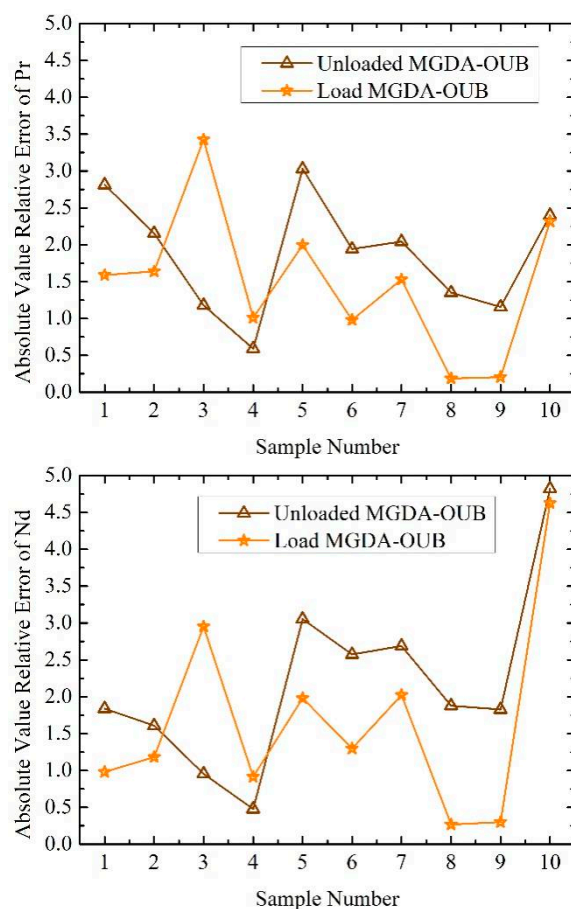


Figure 5. Absolute values of relative errors of the Multi-LightVGG model loaded and unloaded with MGDA-OUB for the prediction of the elemental component contents of Pr and Nd for 10 samples of the test set.

Table 4. Error evaluation index values for Pr and Nd prediction by Multi-ResNet18 and Multi-LightVGG models.

Name of the Model	MRE/%		RMSE		MAX(Error)%	
	Pr	Nd	Pr	Nd	Pr	Nd
Multi-ResNet18	1.8175	2.1944	0.0108	0.0110	3.2346	4.2536
Multi-LightVGG	1.4878	1.6521	0.0089	0.0090	3.4287	4.6264

As can be seen from Figure 6 and Table 4, the MRE and RMSE of Pr and Nd predicted by the Multi-LightVGG model under the same optimization conditions were 0.3297%, 0.5423%, 0.0019, and 0.002 lower than that of Multi-ResNet18, respectively, while the slightly higher MAX(|error|) of the Multi-LightVGG model than that of the Multi-ResNet18 model was due to the poor prediction results of very few samples. This suggests that the Multi-LightVGG model can effectively improve the prediction accuracy for each rare earth element, wherein the backbone network LightVGG is more capable of adequately capturing abstract representations in the images of the extracted rare earth mixed solutions compared to ResNet18. Therefore, the Multi-LightVGG model loaded with MGDA-OUB is more suitable for the soft measurement of rare earth element component contents than the above model.

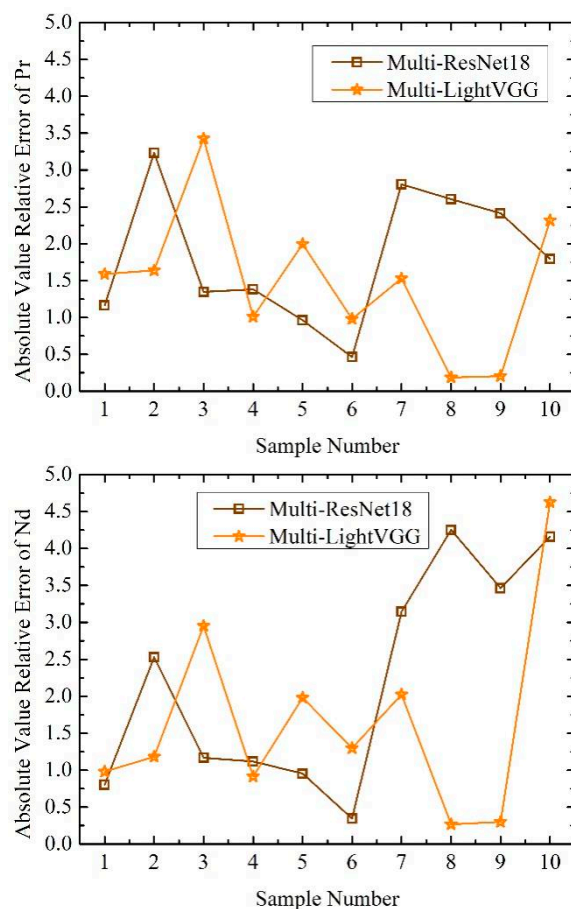


Figure 6. Absolute values of the relative errors of the Multi-LightVGG model loaded with MGDA-OUB and the Multi-ResNet18 model for the prediction of the elemental component contents of Pr and Nd for 10 samples of the test set.

The $\text{MAX}(|\text{error}|)$ of all the above models for rare earth multi-element component content prediction is within $\pm 5\%$, which is within the maximum relative error of the rare earth extraction process for elemental component content prediction [34], meeting the requirements of actual extraction production.

4. Conclusions

We have developed a soft measurement method that can accurately and easily detect the component contents of multiple rare earth elements. This is in response to the current issue in the rare earth extraction process, where detecting the component contents of each element requires complex and expensive hardware equipment. Our method involves collecting images of mixed Pr/Nd extraction solutions, building a dataset, and constructing the Multi-LightVGG model based on the color characteristics of Pr and Nd elements. Additionally, we optimized the multi-task learning model through MGDA-OUB. After conducting multiple sets of experiments, we have reached the following conclusions:

The Multi-LightVGG model loaded with MGDA-OUB has a lower MRE, RMSE for Pr, Nd prediction, and $\text{MAX}(|\text{error}|)$ for Nd prediction than the Multi-LightVGG model without MGDA-OUB by 0.3778%, 0.5208%, 0.0015, 0.0015 and 0.1985%, respectively, indicating that MGDA-OUB can effectively find Pareto solutions for multiple specific tasks and avoid possible conflicts between specific tasks, so as to optimize the rare earth multi-element component content prediction model and improve the prediction accuracy of the model for each elemental component content. The MRE and RMSE of the Multi-LightVGG model for the respective prediction of Pr and Nd under the same optimization conditions are 0.3297%, 0.5423%, 0.0019, and 0.002 lower than that of Multi-ResNet18, respectively, which

indicates that the Multi-LightVGG model is better than the Multi-Resnet18 model in terms of the backbone network being able to effectively capture the abstract representations in the images of rare earth extraction mixed solutions. This in turn improves the prediction accuracy of the content of each elemental component.

Our proposed method has practical impacts on the extraction process of rare earths as it meets the accuracy for predicting the component content of each rare earth element. It provides a new way of thinking about the soft measurement of the component content of rare earth elements.

Author Contributions: Conceptualization, Z.L. and J.L.; methodology, Z.L.; software, Q.Z.; validation, J.X. and K.L.; investigation, Z.L.; data curation, J.X.; writing—original draft preparation, Z.L.; writing—review and editing, K.L.; supervision, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation (51974140, 52064018), Distinguished Professor Program of Jinggang Scholars in institutions of higher learning, the Jiangxi Provincial Natural Science Foundation (20212BAB203013), the Education Department of Jiangxi Province (No. GJJ2200820).

Data Availability Statement: The datasets generated during the current study are not publicly available due to sensitive data involving production units.

Acknowledgments: Authors acknowledge the College of Rare Earth, the College of Chemistry and Chemical Engineering, School of Information Engineering, Jiangxi University of Science and Technology for providing the necessary experimental facilities for carrying out this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Thibeault, A.; Ryder, M.; Tomomewo, O.; Mann, M. A review of competitive advantage theory applied to the global rare earth industry transition. *Resour. Policy* **2023**, *85*, 103795. [[CrossRef](#)]
2. Liu, T.; Chen, J. Extraction and separation of heavy rare earth elements: A review. *Sep. Purif. Technol.* **2021**, *276*, 119263. [[CrossRef](#)]
3. Chen, Z. Global rare earth resources and scenarios of future rare earth industry. *J. Rare Earths* **2011**, *29*, 1–6. [[CrossRef](#)]
4. Feng, Z.; Wang, M.; Zhao, L.; Xu, Y.; Zhang, Y. Current status and prospect of the development of extraction, separation and purification technology of rare earth elements. *China J. Rare Earths* **2021**, *39*, 469–478.
5. Xu, X.; Tan, Q.; Liu, L.; Xu, K.; Li, J.; Wang, Z. Research status and prospect of rare earth element separation and purification technology. *Environ. Pollut. Prev.* **2019**, *41*, 844–851.
6. Liao, C.; Cheng, F.; Wu, S.; Yan, C. Development history and recent progress of string-stage extraction theory. *Chin. J. Rare Earths* **2017**, *35*, 1–8.
7. Liao, C.; Cheng, F.; Wu, S.; Yan, C. Development of tandem extraction theory and technological progress of rare earth separation industry. *Chin. J. Rare Earths* **2022**, *40*, 909–919.
8. Chai, T.-Y.; Yang, H. Current status and development trend of automatic control of rare earth extraction and separation processes. *China J. Rare Earths* **2004**, *22*, 427–433.
9. Zhu, J.; Wang, W.; Yang, H.; Xu, F.; Lu, R. Simulation of rare earth extraction process based on multibranch residual deep network. *Control Theory Appl.* **2022**, *39*, 2242–2253.
10. Yang, H.; Gao, Z.; Lu, R. Component content detection method based on rare earth ion color feature recognition. *China J. Rare Earths* **2012**, *30*, 108–112.
11. Lu, R.; Chen, M.; Yang, H.; Zhu, J. Dynamic monitoring system of elemental component content based on solution image temporal features. *Comput. Appl.* **2021**, *41*, 3075–3081.
12. Yang, H.; Xu, F.; Lu, R.; Ding, Y. Component content distribution profile control in rare earth countercurrent extraction process. *Chin. J. Chem. Eng.* **2015**, *23*, 192–198. [[CrossRef](#)]
13. Yang, H.; Xu, Y.; Wang, X. Component content soft-sensor based on rbf neural network in rare earth countercurrent extraction process. In Proceedings of the 2006 6th World Congress on Intelligent Control and Automation, Dalian, China, 21–23 June 2006; IEEE: Piscataway, NJ, USA, 2006; pp. 4909–4912.
14. Xiang, Z.; Liu, S. Component content soft-sensor in rare-earth extraction based on pso and ls-svm. In Proceedings of the 2008 Fourth International Conference on Natural Computation, Jinan, China, 18–20 October 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 392–395.
15. Lu, R.; Yang, H.; Zhang, K. Component content soft-sensor of svm based on ions color characteristics. *TELKOMNIKA Indones. J. Electr. Eng.* **2012**, *10*, 1445–1452.

16. Yang, H.; Liu, S.; Lu, R.; Zhu, J. Prediction of component content in rare earth extraction process based on esns-adaboost. *IFAC-PapersOnLine* **2018**, *51*, 42–47.
17. Lu, R.; Yang, H. Soft measurement for component content based on adaptive model of pr/nd color features. *Chin. J. Chem. Eng.* **2015**, *23*, 1981–1986. [[CrossRef](#)]
18. Wu, B.; Ji, X.; He, M.; Yang, M.; Zhang, Z.; Chen, Y.; Wang, Y.; Zheng, X. Mineral Identification Based on Multi-Label Image Classification. *Minerals* **2022**, *12*, 1338. [[CrossRef](#)]
19. Wang, H.; Cao, W.; Zhou, Y.; Yu, P.; Yang, W. Multitarget Intelligent Recognition of Petrographic Thin Section Images Based on Faster RCNN. *Minerals* **2023**, *13*, 872. [[CrossRef](#)]
20. Nie, X.; Zhang, C.; Cao, Q. Image Segmentation Method on Quartz Particle-Size Detection by Deep Learning Networks. *Minerals* **2022**, *12*, 1479. [[CrossRef](#)]
21. Vandenhende, S.; Georgoulis, S.; Van Gansbeke, W.; Proesmans, M.; Dai, D.; Van Gool, L. Multi-task learning for dense prediction tasks: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3614–3633. [[CrossRef](#)]
22. Crawshaw, M. Multi-task learning with deep neural networks: A survey. *arXiv* **2020**, arXiv:2009.09796.
23. Vandenhende, S.; Georgoulis, S.; Van Gool, L. MTI-net: Multi-scale task interaction networks for multi-task learning. *arXiv* **2020**, arXiv:2001.06902.
24. Chen, Z.; Badrinarayanan, V.; Lee, C.-Y.; Rabinovich, A. GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018.
25. Cipolla, R.; Gal, Y.; Kendall, A. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 7482–7491.
26. Désidéri, J.-A. Multiple-gradient descent algorithm (mgda) for multiobjective optimization. *Comptes Rendus Math.* **2012**, *350*, 313–318. [[CrossRef](#)]
27. Sener, O.; Koltun, V. Multi-task learning as multi-objective optimization. *arXiv* **2019**, arXiv:1810.04650.
28. Zhou, X.; Gao, Y.; Li, C.; Yang, C. An end-to-end license plate recognition method based on multi-objective optimization multi-task learning. *Control Theory Appl.* **2021**, *38*, 676–688.
29. Zhao, J.; Zhou, X.; Gao, Y.; Liu, T.; Yang, C. Recognition of working conditions in froth flotation process based on multi-objective learning. *J. Cent. South Univ. (Nat. Sci. Ed.)* **2022**, *53*, 2071–2079.
30. Zhang, S.; Zhang, Q.; Wang, B.; Zhang, X.; Lan, B.; Guo, H. Prediction of rare earth element component content based on deep machine vision. *Nonferrous Met. Sci. Eng.* **2023**, *14*, 587–596.
31. Wang, M.; Lu, S.; Zhu, D.; Lin, J.; Wang, Z. A high-speed and low-complexity architecture for softmax function in deep learning. In Proceedings of the 2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), Chengdu, China, 26–30 October 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 223–226.
32. Mandelbrot, B. The Pareto-Levy law and the distribution of income. *Inte. Econ. Rev.* **1960**, *1*, 79–106. [[CrossRef](#)]
33. Rootzén, H.; Tajvidi, N. Multivariate generalized Pareto distributions. *Bernoulli* **2006**, *12*, 917–930. [[CrossRef](#)]
34. Lu, R.; He, Q.; Yang, H.; Zhu, J. Prediction of multi-component content of rare earth mixed solutions based on GA-ELM. *Comput. Eng.* **2021**, *47*, 284–290+297.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.