

A Novel Pairwise Domain-Adaptation-Assisted Dual-Task Learning Approach to Co-Prediction of Robotic Machining Efficiency and Quality in New Parameter Spaces

Guijun Ma, *Member, IEEE*, Zidong Wang, *Fellow, IEEE*, Zeyuan Yang, Ruijuan Chen, Weibo Liu, *Member, IEEE*, Yong Zhang and Sijie Yan

Abstract—Accurate prediction of material removal depth and averaged surface roughness is crucial for evaluating the performance of robotic belt grinding (RBG). Nevertheless, the machining parameters of RBG across different spaces exhibit various data distributions, which often results in prediction shifts on unseen machining parameters when using conventional approaches. In this article, we introduce a pairwise domain adaptation-assisted dual-task learning (PW-DA-DTL) method for co-predicting material removal depth and averaged surface roughness with regard to new RBG machining parameter spaces. The multi-gate mixture-of-experts method is employed as the foundational framework for dual-task learning, effectively capturing and modeling the relationships between material removal depth and average surface roughness by leveraging their inherent task interdependencies. The pairwise domain adaptation strategy is put forward to simultaneously enhance sample diversity and mitigate cross-domain data distribution discrepancy between the existing and new RBG machining parameter spaces. Comparative experiments are presented to demonstrate the effectiveness and superiority of the proposed PW-DA-DTL method.

Index Terms—Material removal depth, averaged surface roughness, domain adaptation, dual-task learning, robotic belt grinding.

This work was supported in part by the National Natural Science Foundation of China under Grant Number 52188102, China Postdoctoral Science Foundation under Grant Number 2024M750991, Postdoctor Project of Hubei Province of China under Grant Number 2024HBBHCXA010, and Open Project Fund of Key Laboratory of Image Processing and Intelligent Control, Ministry of Education of China. (*Corresponding author: Zidong Wang*)

G. Ma is with the the State Key Laboratory of Intelligent Manufacturing Equipment and Technology, School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China, and also with the Key Laboratory of Image Processing and Intelligent Control (Huazhong University of Science and Technology), Ministry of Education, Wuhan 430074, China (e-mail: mgj@hust.edu.cn).

Z. Wang is with the College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China, and also with the Department of Computer Science, Brunel University London, Uxbridge, Middlesex, UB8 3PH, United Kingdom (e-mail: Zidong.Wang@brunel.ac.uk).

Z. Yang and S. Yan are with the State Key Laboratory of Intelligent Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan 430074, China, and also with the HUST-Wuxi Research Institute, Wuxi 214174, China (e-mail: yangzeyuan@hust.edu.cn; sjyan@hust.edu.cn).

R. Chen is with the Research Center of Nonlinear Science, School of Mathematical and Physical Sciences, Wuhan Textile University, Wuhan 430200, China (e-mail: rjchen@wtu.edu.cn).

W. Liu is with the Department of Computer Science, Brunel University London, Uxbridge, Middlesex, UB8 3PH, United Kingdom (e-mail: Weibo.Liu2@brunel.ac.uk).

Y. Zhang is with the School of Artificial Intelligence and Automation, Wuhan University of Science and Technology, Wuhan 430081, China (e-mail: zhangyong77@wust.edu.cn).

I. INTRODUCTION

Robotic Belt Grinding (RBG) has been recognized as an effective technique for finishing machining and difficult-to-machine materials, owing to its high flexibility, extensive operating space and cost-effectiveness [1], [2]. Efficient and precise manufacturing is an enduring objective of RBG, which could be achieved by enhancing material removal ability and workpiece surface quality simultaneously [3], [4]. Material removal depth (MRD) and averaged surface roughness (denoted as R_a), representing material removal capability and surface quality respectively, are widely recognized as key indicators for understanding and evaluating the performance of RBG. As a critical element in the closed-loop manufacturing process, accurate prediction of MRD and R_a is vital for machining parameter optimization and quality control, ultimately enabling an efficient and precise RBG manufacturing process [5]–[9].

Over the last few decades, numerous machine learning methods have been developed to individually predict MRD and R_a [1], [4], [10]. Nevertheless, these methods commonly require training separate models, neglecting the potential correlation between MRD and R_a . Serving as a promising technique, multi-task supervised learning (such as the shared-bottom method, the tensor factorization method and the multi-gate mixture-of-experts (MMoE) method) is designed to enhance the learning ability on multiple learning tasks by exploiting knowledge across tasks. Among these multi-task supervised learning methods, MMoE stands out for its capability to balance the correlation and discrepancy across tasks with a few training parameters [11], [12], making it feasible for dual-task prediction of MRD and R_a .

On the other hand, conventional multi-task supervised learning methods usually adhere to the typical assumption that the training and testing data follow an independent and identically distributed (i.i.d.) pattern [5], [13]. In fact, the machining parameters across different spaces often exhibit diverse data distributions, leading to prediction shifts on unseen machining parameters when applying conventional multi-task supervised learning methods to new machining parameter spaces [14]. In this situation, there is a clear need to develop advanced multi-task supervised learning methods capable of addressing the challenges posed by inconsistent data distributions across different RBG machining parameter spaces.

Serving as a cutting-edge technique, transfer learning al-

lows for the relaxation of constraint that the training and testing data must be i.i.d. when applying conventional multi-task supervised learning methods [15]–[18]. Among numerous transfer learning methods, deep transfer learning has received much attention from both academia and industry owing to its exceptional feature extraction ability and remarkable effectiveness [19]. Notably, two dominant paradigms in deep transfer learning are fine-tuning-based methods and deep domain adaptation-based (DDA) methods, with the latter being widely recognized for its proficiency in quantifying cross-domain distribution discrepancies [20]–[22]. A seemingly reasonable idea is to exploit DDA methods to alleviate data distribution discrepancy between existing and new RBG machining parameter spaces, thereby facilitating accurate co-prediction of MRD and R_a for new machining parameters.

In general, the predictive performance of DDA methods depends heavily on the availability of sufficient samples in both the source and target domains [23], [24]. Nevertheless, collecting and annotating RBG data in the new machining parameter space is prohibitively expensive [25], resulting in only a limited number of labeled samples available for DDA model training. The lack of target domain samples may lead to ineffective training and insufficient alignment of data distributions between the training and testing data when using DDA methods [26]–[28]. Therefore, it becomes essential to develop a novel domain adaptation strategy capable of eliminating distribution discrepancy between training and testing data, even when only few testing samples are available.

Based on the above discussions, the aim of this article is to accurately predict MRD and R_a in the new machining parameter space of RBG. Due to the lack of labeled target domain samples, conventional DDA methods face significant challenges in handling the data imbalance problem, which leads to insufficient alignment of data distributions between the source and target domains. To tackle this issue, a pairwise domain adaptation-assisted dual-task learning (PW-DA-DTL) method is put forward in this article, where a novel pairwise DA strategy is developed for co-predicting material removal depth and averaged surface roughness with regard to new machining parameters of RBG. Specifically, the MMoE method, a representative DTL technique, is employed to improve the prediction performance by balancing the intrinsic correlation and discrepancy between MRD and R_a . Additionally, a pairwise DA strategy is designed to simultaneously augment target domain samples and align data distributions across the two domains. The proposed PW-DA-DTL method offers a feasible solution for accurately evaluating various RBG machining parameters, enabling reliable selection and optimization of machining parameters in the RBG manufacturing process.

The main contributions of this article can be summarized in the following four aspects.

- 1) A novel PW-DA-DTL method is proposed for accurately co-predicting MRD and R_a in new RBG machining parameter spaces.
- 2) A pairwise DA strategy is put forward to enhance the data distribution alignment within the source and target domains under data imbalance.
- 3) The MMoE method is employed to effectively balance the intrinsic correlation and discrepancy between the

MRD and R_a prediction tasks.

- 4) The proposed PW-DA-DTL method is successfully validated on an RBG platform for co-predicting MRD and R_a . Comparative experiments demonstrate the effectiveness and advantages of the proposed PW-DA-DTL method over five benchmark methods.”

The remaining sections of this article are organized as follows. The background of multi-task learning, domain adaptation and problem description are presented in Section II. In Section III, the proposed PW-DA-DTL method and its overall loss function is introduced. Additionally, the proposed method for co-predicting MRD and R_a of RBG, as well as three evaluation metrics are provided. The self-developed experimental platform and the experimental settings are described in Section IV. Experimental results and analysis of the MRD and R_a co-prediction are provided and discussed in Section V. Finally, the conclusions are drawn, and the possible future topics are addressed in Section VI.

II. PRELIMINARY

A. Background of Multi-Task Learning

Multi-task learning (MTL) aims to improve the overall learning ability by leveraging the shared knowledge and relationships among tasks. Given l learning tasks $\{\mathcal{T}^i\}_{i=1}^l$, in which all the learning tasks or a subset of them are related but not identical, MTL is dedicated to enhancing the learning of a model for \mathcal{T}^i by leveraging the knowledge across the l tasks.

A common form of MTL is multi-task supervised learning (MTSL), where each task involving learning the mapping function from the feature space to the label space [29]. In MTSL, each task \mathcal{T}^i contains n_i labeled samples $\{(x_j^i, y_j^i)\}_{j=1}^{n_i}$, where x_j^i lies in a d -dimensional feature space \mathcal{X} and y_j^i is the corresponding label in a label space \mathcal{Y} . To goal of MSTL is to simultaneously train l functions $\{f^i(X)\}_{i=1}^l$ for the l learning tasks leveraging the labeled samples across tasks.

In certain scenarios, multi-task supervised learning (MTSL) employs feature transformation techniques to enhance learning performance, collectively known as feature transformation-based multi-task supervised learning (FT-MTSL). In FT-MTSL, different tasks share identical or similar feature representations, which are derived from linear or nonlinear transformations of the original features [29]. Given a shared transformation function $h(\cdot)$, the original feature representations for each learning task are transformed as $\{h(x_j^i)\}_{j=1}^{n_i}$. Then, the transformed feature representations are fed into MTSL for the simultaneous training of l functions $\{f^i(h(X))\}_{i=1}^l$ with respect to the l learning tasks.

B. Background of Domain Adaptation

Domain adaptation (DA) is a transfer learning technique that enhances model performance in a target domain (with limited or no labeled samples) by leveraging knowledge from a related but different source domain (with abundant labeled samples) [19], [30]. In the context of transfer learning, a domain \mathcal{D} is composed of a d -dimensional feature space \mathcal{X} and a marginal probability distribution $P(X)$, denoted as $\mathcal{D} = \{\mathcal{X}, P(X)\}$,

with $X \in \mathcal{X}$. For a given source domain $\mathcal{D}^s = \{\mathcal{X}^s, P(X^s)\}$ and a target domain $\mathcal{D}^t = \{\mathcal{X}^t, P(X^t)\}$, the two domains are different if they have different feature spaces or marginal distributions (i.e., $\mathcal{X}^s \neq \mathcal{X}^t \vee P(X^s) \neq P(X^t)$).

In the context of the domain \mathcal{D} , the task \mathcal{T} is defined by a label space \mathcal{Y} and a prediction function $f(X)$, represented as $\mathcal{T} = \{\mathcal{Y}, f(X)\}$. Specifically, the prediction function is denoted as $f(X) : X \rightarrow Y$, where $Y \in \mathcal{Y}$. Inspired by [19], [30], $f(X)$ can be chosen as the conditional probability distribution $P(Y|X)$. For a given source task $\mathcal{T}^s = \{\mathcal{Y}^s, f(X^s)\}$ and a target task $\mathcal{T}^t = \{\mathcal{Y}^t, f(X^t)\}$, two tasks are different if they have different label spaces or conditional distributions (i.e., $\mathcal{Y}^s \neq \mathcal{Y}^t \vee P(Y^s|X^s) \neq P(Y^t|X^t)$).

Given n labeled samples $\{(x_i^s, y_i^s)\}_{i=1}^n$ from the source domain \mathcal{D}^s , m labeled samples $\{(x_j^t, y_j^t)\}_{j=1}^m$ and p unlabeled samples $\{x_j^t\}_{j=m+1}^{m+p}$ from the target domain \mathcal{D}^t , the objective of DA is to learn a target prediction function $f(x_j^t) : x_j^t \rightarrow y_j^t$, ($j = m + 1, \dots, m + p$), aimed at handling the unlabeled target domain samples with low predicted error, under the assumptions that the feature spaces and label spaces are identical (i.e., $\mathcal{X}^s = \mathcal{X}^t \wedge \mathcal{Y}^s = \mathcal{Y}^t$), while the marginal and conditional distributions are different (i.e., $P(X^s) \neq P(X^t) \wedge P(Y^s|X^s) \neq P(Y^t|X^t)$).

C. Problem Description

1) *FT-MTSL for MRD and R_a Co-Prediction in RBG*: In the RBG manufacturing process, MRD and R_a are closely related to the machining parameters (such as normal force, robotic feed speed, belt linear speed and equivalent radius of abrasive grains). During the training process of FT-MTSL, x represents the machining parameters and y consists of MRD and R_a . $h(\cdot)$ is a nonlinear function that maps the raw data x into a shared feature space. The objective of FT-MTSL for MRD and R_a co-prediction is to learn a prediction on the unlabeled samples leveraging the knowledge across two tasks.

2) *DA-Assisted FT-MTSL for MRD and R_a Co-Prediction in RBG*: Due to the discrepancy among various machining parameter spaces, the data distributions are different between the source dataset (i.e., samples with existing machining parameters) and the target dataset (i.e., samples with new machining parameters). Given the source dataset (which contains n labeled samples) and the target dataset (which has m labeled samples and p unlabeled samples), the objective of the DA-assisted FT-MTSL method for MRD and R_a co-prediction is to learn a target prediction on the unlabeled target domain samples with low expected error.

III. METHODS

In this section, the MMoE method, the developed DA strategy and the proposed DA-based dual-task learning method are introduced in detail. Furthermore, three evaluation metrics on MRD and R_a prediction of RBG are presented.

A. MMoE

The MMoE is designed to explicitly model relationship and difference among tasks. The output of the k_{th} ($k = 1, \dots, l$)

task in the MMoE method is calculated as follows:

$$y^k = h_t^k \left(\sum_{i=1}^s g_i^k(x) f_i(x) \right), \quad (1)$$

where x represents the input data or its transformed features, s is the number of expert networks, and $f_i(\cdot)$ denotes the mapping function of the i_{th} expert network. $g^k(x) = \text{softmax}(W_g^k x)$ is the output of the k_{th} gating network, where $W_g^k \in \mathbb{R}^{s \times d}$ is the weight matrix and d represents the dimension of x . $h_t^k(\cdot)$ denotes the mapping function of the k_{th} tower network, and y^k corresponds to the output of the tower network. The schematic structure of MMoE method is illustrated in Fig. 1.

Remark 1: Compared to the conventional shared-bottom method [31], MMoE is able to model the relationships of tasks in a weighted manner by determining how the separations resulting from different gating networks overlap. If the provided tasks are less related, sharing expert networks would be penalized, and the gating networks for these tasks would learn to select different expert networks instead [11].

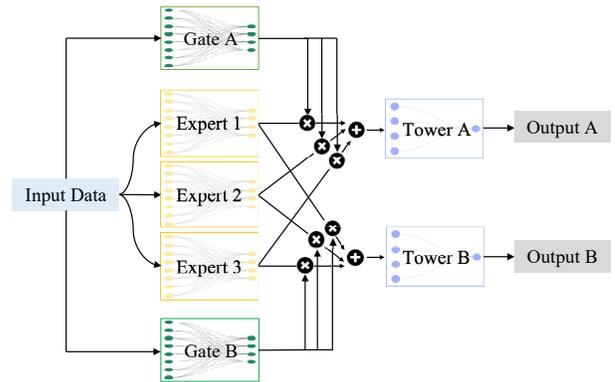


Fig. 1. The schematic structure of MMoE method. Dual-task learning is taken as a representative case.

B. The Developed DA Strategy

1) *Maximum Mean Discrepancy*: This article utilizes the empirical maximum mean discrepancy (MMD) to characterize the distribution differences between the source and target domains [32], [33]. Leveraging the MMD, the distribution discrepancies between the source and target domains are computed by mapping the original data into a reproducing kernel Hilbert space (RKHS). In comparison to other distance functions, such as the Kullback–Leibler divergence, MMD serves as a non-parametric distance measurement among data distributions, circumventing the calculations of intermediate density [34]. The empirical MMD of marginal distribution between the source and target domains is defined as the following form:

$$D(X^s, X^t) = \left\| \frac{1}{n} \sum_{i=1}^n \phi(x_i^s) - \frac{1}{m} \sum_{j=1}^m \phi(x_j^t) \right\|_{\mathcal{H}}^2, \quad (2)$$

where $\phi(\cdot)$ represents a mapping from the original space into the RKHS, and $\|\cdot\|_{\mathcal{H}}^2$ denotes the squared norm of “ \cdot ” in the RKHS.

According to the reproducing property of the RKHS, the kernel trick (i.e., $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}}$, where x_i and x_j represent two random samples; $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is the inner dot in the RKHS) is employed to unfold (2) as the following form:

$$D(X^s, X^t) = \frac{1}{n^2} \sum_{i=1}^n \sum_{i'=1}^n k(x_i^s, x_{i'}^s) + \frac{1}{m^2} \sum_{j=1}^m \sum_{j'=1}^m k(x_j^t, x_{j'}^t) - \frac{2}{mn} \sum_{i=1}^n \sum_{j=1}^m k(x_i^s, x_j^t) \quad (3)$$

In this article, the Gaussian kernel is adopted as the kernel function [32], which is defined as follows:

$$k(x_i, x_j) = \exp\left(-\|x_i - x_j\|^2 / (2\sigma^2)\right), \quad (4)$$

where σ is a hyperparameter that represents the width of Gaussian kernel.

2) *The Pairwise DA Strategy*: In this article, considering that only a limited number of labeled target domain samples (even as few as one sample) are available, a pairwise DA strategy is developed to alleviate the insufficient alignment of data distributions between the source and target domains based on the MMD method. The pairwise DA strategy is designed to augment labeled target domain samples by pairing them with the source domain samples in a recurrent manner. Notably, given n labeled samples in the source domain and m labeled samples in the target domain ($m \ll n$), $m \times n$ target-source pairs are created through the recurrent pairing of target and source domain samples. Simultaneously, $n \times n$ source-source pairs are generated by recurrently pairing among source domain samples, with a random selection of $m \times n$ source-source pairs to align with the number of target-source pairs. In Fig. 2, two kinds of sample pairs (denoted as target-source pairs and source-source pairs) are simultaneously input into a feature transformation network (which is a designed artificial neural network), and the transformed features are merged for DA using the target-source data and the source-source data.

After the DA, the transformed features are fed into the MMoE (Section III-A) for the supervised dual-task learning. The supervised training manner can alleviate model bias caused by the imbalanced labels between the source and target domains, thereby enhancing the reliability of the proposed method. The structure of the proposed PW-DA-DTL method is shown in Fig. 2.

Remark 2: The pairwise strategy can be considered as a form of data augmentation method, where a limited number of target domain samples are expanded by recurrent grouping of target and source domain samples. In contrast to conventional data augmentation methods, such as generative adversarial network (GAN) [35], synthetic minority oversampling technique (SMOTE) [36] and adaptive synthetic (ADASYN) algorithm [37], the proposed pairwise strategy can increase the sample size using few samples, even as few as one sample. Compared to the sample replication method, the proposed pairwise strategy can significantly improve the diversity of augmented target domain samples by integrating the limited number of target domain samples with sufficient source domain samples. The pairwise strategy, combined with supervised learning [38],

[39], simultaneously alleviates model bias and improves the generalization capability of the proposed method.

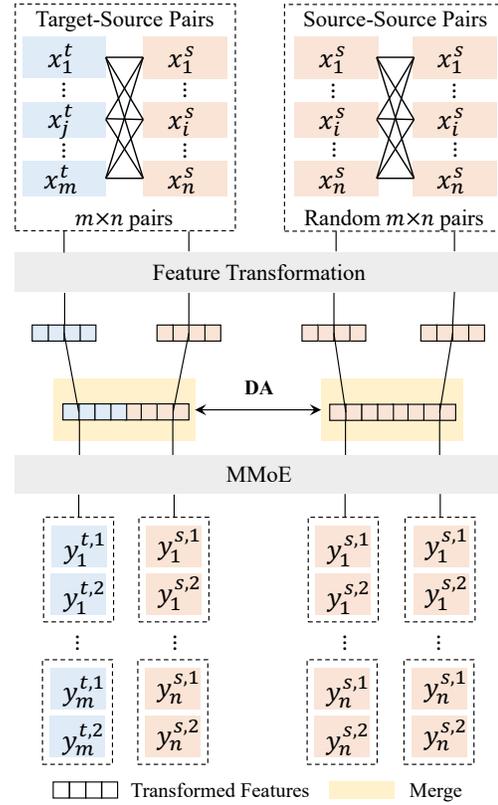


Fig. 2. The pairwise DA strategy and the structure of PW-DA-DTL method. $m \times n$ target-source pairs and source-source pairs are generated for training the PW-DA-DTL method.

3) *Domain Loss*: The domain loss is the MMD between the merged features of target-source pairs and source-source pairs. The calculation of domain loss (which corresponds to (2)) is transformed as follows:

$$L_d = D\left([h(X^s), h(X^s)], [h(X^t), h(X^s)]\right) = \left\| \frac{1}{n^2} \sum_{i=1}^n \sum_{i'=1}^n \phi([h(x_i^s), h(x_{i'}^s)]) + \frac{1}{m^2} \sum_{j=1}^m \sum_{j'=1}^m \phi([h(x_j^t), h(x_{j'}^t)]) - \frac{2}{mn} \sum_{i=1}^n \sum_{j=1}^m \phi([h(x_i^s), h(x_j^t)]) \right\|_{\mathcal{H}}^2, \quad (5)$$

where $[\cdot, \cdot]$ represents the concatenation of two vectors, and $h(\cdot)$ is a feature transformation network that maps the input data to a shared feature space.

4) *Loss Function*: The overall loss function of the proposed PW-DA-DTL method is depicted as follows:

$$L = L_r + \lambda L_d, \quad (6)$$

where L_r is the regression loss (which is computed by the mean square error (MSE)) of both labeled source and target domain samples. L_d denotes the domain loss (5) between

the merged features of target-source pairs and source-source pairs. The parameter λ serves as a penalty term that balances between the regression loss and the domain loss. The PW-DA-DTL method is trained by simultaneously minimizing the regression loss and the domain loss.

The training process of the proposed PW-DA-DTL method is presented in Algorithm 1.

Algorithm 1: The training process of the proposed PW-DA-DTL method

Input: Labeled source domain samples $\{(x_i^s, y_i^s)\}_{i=1}^n$;
 labeled target domain samples $\{(x_j^t, y_j^t)\}_{j=1}^m$.

Output: Model weights θ .

Initialize θ and the number of training epoch N ;
 Construct the target-source pairs $[X^t, X^s]$ and the source-source pairs $[X^s, X^s]$;

for 1 : N **do**

- 1) Obtain $[h(X^s), h(X^s)]$ and $[h(X^t), h(X^s)]$;
- 2) Calculate the MMD using (5);
- 3) Predict y^k for dual tasks using (1);
- 4) Run the Adam optimization algorithm to minimize (6) and update the weights θ .

end

C. PW-DA-DTL for Co-Predicting MRD and R_a in RBG

In this article, the PW-DA-DTL method is applied for co-predicting MRD and R_a in RBG. The training process of PW-DA-DTL method unfolds in two main stages. First, existing and new machining parameters in RBG are recurrently paired to form target-source pairs and source-source pairs. Subsequently, these pairs are input into the feature transformation network for both DA and supervised dual-task model training. To be specific, the pairwise outputs of the feature transformation network are utilized for MMD calculation, and the pairwise outputs are simultaneously fed into MME for the supervised training.

D. Evaluation Metrics

Three evaluation metrics (i.e., the root mean square error (RMSE), the mean absolute percentage error (MAPE) and the coefficient of determination (R^2)) are employed to assess the performance of MRD and R_a co-prediction. It should be mentioned that the RMSE is applied to calculate the deviation between the predicted and observed values, the MAPE is utilized to measure the ratio of the absolute error to the observed values, and the R^2 coefficient is used to assess the correlation between the predicted and observed values. The aforementioned three evaluation metrics are formulated by

$$\text{RMSE} = \sqrt{\frac{1}{p} \sum_{i=1}^p (y_i - \hat{y}_i)^2}, \quad (7)$$

$$\text{MAPE} = \frac{1}{p} \sum_{i=1}^p \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%, \quad (8)$$

$$R^2 = 1 - \frac{\sum_{i=1}^p (\hat{y}_i - y_i)^2}{\sum_{i=1}^p \left(y_i - \frac{1}{p} \sum_{i=1}^p y_i \right)^2}, \quad (9)$$

where y_i and \hat{y}_i represent the observed MRD/ R_a and the predicted MRD/ R_a , respectively; and p is the number of testing samples.

IV. EXPERIMENTAL DATA AND MODEL TRAINING

A. Experimental Platform

As shown in Fig. 3(a)-(b), the developed RBG system comprises an industrial robot (ABB IRB4400-60/1.96) equipped with a six-dimensional force sensor (ATI Omega 160) and an abrasive belt grinder. In the abrasive belt grinder, ceramic alumina abrasive belts (3MTMCubitronTM II 726 A) with four equivalent radius of grains are chosen as the machining tools due to their substantial material removal capacity and exceptional resistance to clogging. The experimental workpiece is a unidirectional carbon fiber-reinforced composite (CFRP) T300 with dimensions of $200 \times 100 \times 10 \text{ mm}^3$, which has been widely utilized in aerospace and automotive industries. In this work, four key controlled parameters (namely, normal force F_n , robotic feed speed V_w , belt linear speed V_s and equivalent radius of abrasive grains \bar{R}) are adjusted to investigate the MRD and R_a of RBG.

Followed by the offline system calibration and robot path planning, an orthogonal grinding experiment with different parallel grinding paths (see Fig. 3(c)) are carried out using four key machining parameters. The measurement of MRD is performed using a self-developed measuring system with an accuracy of $1 \mu\text{m}$ and a sampling frequency of 4.33 kHz. The measurement of R_a is deployed by using a digital surface roughness tester (see Fig. 3(d)). To ensure the reliability of measurement data, the final measurement results are derived by calculating the average value of cross-sectional removal profile centers in the stable grinding area (see Fig. 3(c)).

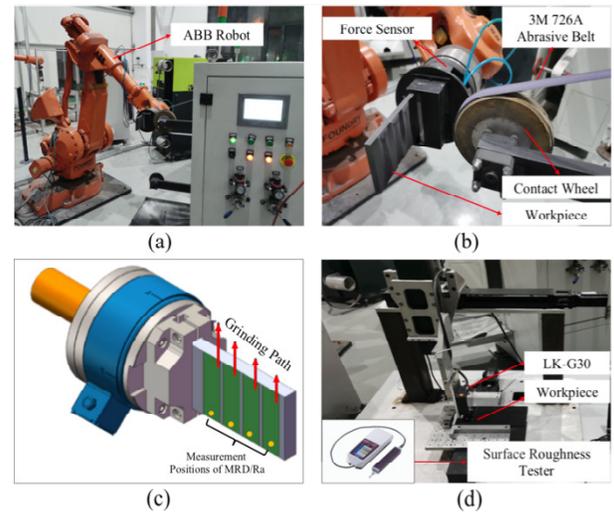


Fig. 3. (a) The developed experimental platform of RBG. (b) The RBG system with an enlarged view of grinding process. (c) The schematic grinding paths as well as the central measurement positions of MRD and R_a . (d) The developed measuring system for obtaining MRD and R_a .

B. Dataset

In this article, the controlled machining parameters are normal force F_n , robotic feed speed V_w , belt linear speed

V_s and equivalent radius of abrasive grains \tilde{R} . Five different values are chosen for F_n (10 N, 20 N, 30 N, 40 N and 50 N), V_s (10 mm/s, 20 mm/s, 30 mm/s, 40 mm/s and 50 mm/s) and V_w (5.24 m/s, 9.42 m/s, 13.61 m/s, 17.8 m/s and 21.99 m/s), while four different values are chosen for \tilde{R} (30.5 μm , 40 μm , 60 μm and 90 μm). The controlled machining parameters as well as their chosen values are recorded in Table I.

TABLE I
THE CONTROLLED PARAMETERS AND THEIR CHOSEN VALUES

Controlled Parameters	Values
Normal force F_n	10, 15, 20, 25, 30 N
Robotic feed speed V_w	10, 15, 20, 25, 30 mm/s
Belt linear speed V_s	5.24, 9.42, 13.61, 17.8, 21.99 m/s
Equivalent radius of abrasive grains \tilde{R}	30.5, 40, 60, 90 μm

After a random combination of four controlled machining parameters, an orthogonal grinding experiment with 80 sets of parameters is carried out for evaluating the performance of RBG, and the corresponding MRD and R_a are measured as the labels. Two different validation experiments are carried out according to the values of F_n and V_w , respectively. To validate the capability of the proposed PW-DA-DTL method for unseen machining parameters, the training dataset and the testing dataset are split in a cross-validation manner based on the values of F_n and V_w . Taking the validation experiment on F_n as an example, samples with a specific value of F_n are chosen as the testing dataset and the remaining samples are taken as the training dataset in each cross-validation iteration.

PW-DA-DTL method are copied from the pre-trained MMoE method (Fig. 4(b)), and the DA is deployed to further reduce the domain discrepancy between the target-source pairs and the source-source pairs. The regression loss L_r is the MSE for the labeled source domain samples and the labeled target domain sample, and the domain loss L_d is obtained by computing the MMD of merged features between the target-source pairs and the source-source pairs. The total loss L (which is the weighted summation of the regression loss and the domain loss) is minimized by the Adam optimizer in the training process.

The testing process is carried out by feeding the target data of four machining parameters to the trained PW-DA-DTL model, and the corresponding output is the co-predicted MRD and R_a . Five benchmark DTL methods (i.e., the MMoE method (Fig. 4(b)), the shared-bottom method, the one-gate mixture-of-experts (OMoE) method [11], the customized sharing experts (CSE) method [40] and the customized gate control (CGC) method [40]) are trained from scratch using the Adam optimizer. It is noteworthy that the shared-bottom method shares a similar architecture with the MMoE method, differing only in the replacement of the MMoE layer with a single task-shared expert network. The OMoE method focuses on using a single shared gating network to select the appropriate experts instead of the multiple gating networks employed in the MMoE method. The CSE method introduces the task-specific expert networks into the shared-bottom method. The CGC method integrates task-shared and task-specific experts through multiple gating networks. The testing processes of the aforementioned benchmark DTL methods are carried out by feeding the target data to the trained models.

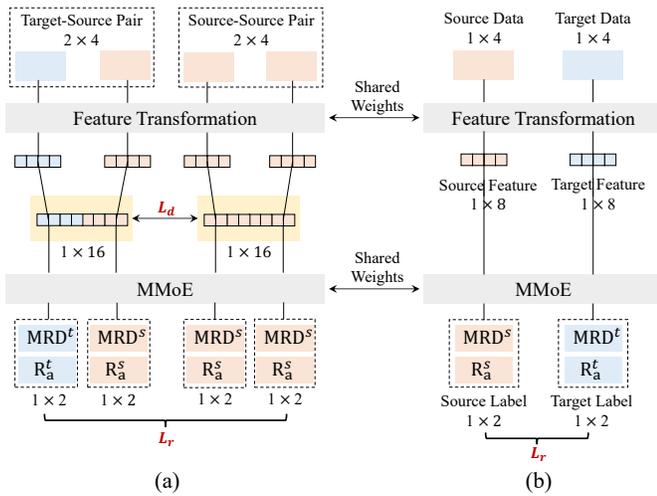


Fig. 4. The training process of (a) the proposed PW-DA-DTL method and (b) the MMoE method.

C. Training and Testing Process

Fig. 4(a) visualizes the training process of the proposed PW-DA-DTL method for MRD and R_a co-prediction of RBG. In this article, only one labeled target domain sample, which is randomly selected from the target dataset, is applied for training the PW-DA-DTL method in the hope of alleviating the dependence on target domain samples. The weights of the

D. Network Architecture and Parameter Setting

In the experiment, the network input is the target-source pair or source-source pair of controlled machining parameters with a size of 2×4 . To learn the shared knowledge between the MRD and R_a , two fully connected layers (i.e., FC1 and FC2), whose neurons are both set to be 8, are utilized for transforming features. The merge layer is designed to merge the transformed features of paired data for the sake of reducing the domain discrepancy between target-source pairs and source-source pairs. Motivated by [11], the number of experts is set to be 4 for balancing the computational burden and learning ability. The number of gates and towers is both 2, which is consistent with the number of tasks. The number of neurons in the expert layer, gate layer and tower layer are set to be 4. The parameters of PW-DA-DTL are listed in Table II. It should be noted that all five benchmark methods share the same parameter settings as the PW-DA-DTL method in terms of input layer, FC1, FC2 and output layer. Additionally, the standard MMoE method, the OMoE method, the CGC method and the PW-DA-DTL method utilize the same number of expert networks and identical parameter configurations for the tower networks.

In the proposed PW-DA-DTL method, σ in (4) and λ in (6) are empirically set to be 0.1 and 0.01, respectively. The learning rate is set to be 0.001. Owing to the weights transfer from the standard MMoE method, the training epoch is set to

TABLE II
 THE PARAMETERS OF THE PROPOSED PW-DA-DTL NETWORK ARCHITECTURE

Layer	Number	Neurons	Output size
Input	-	-	2×4
FC1	1	8	2×8
FC2	1	8	2×8
Merge	1	-	1×16
Gate in MMoE	2	4	2×1×4
Expert in MMoE	4	4	4×1×4
Tower in MMoE	2	4	2×1×1
Output A	-	-	2×1
Output B	-	-	2×1

be 100. In the five benchmark methods, the training epoch is set to be 2000.

V. RESULTS AND DISCUSSION

In this section, the co-prediction results of the MRD and R_a for RBG are presented. Specifically, two scenarios considering new normal force F_n and robotic feed speed V_w are evaluated, respectively. Furthermore, the PW-DA-DTL method is compared with five benchmark methods in terms of the prediction accuracy.

A. Case 1: Co-Prediction of MRD and R_a with New Normal Force

In this experiment, the MRD and R_a co-prediction results of the proposed PW-DA-DTL method with different testing F_n (10 N, 15 N, 20 N, 25 N and 30 N) are displayed in Fig. 5. The left figure shows the correlation between the predicted MRD and the observed MRD, and the right figure shows the correlation between the predicted R_a and the observed R_a . The two figures both showcase the consistency between the predicted values and the the observed values. To showcase the effectiveness of the PW-DA-DTL method, the predicted MRD and R_a under new F_n using different methods are shown in Fig. 6. The curves indicate that the prediction results of PW-DA-DTL (denoted as PW-DA-MMoE in figures) method are more closely to the actual labels, which demonstrates the effectiveness of the proposed method.

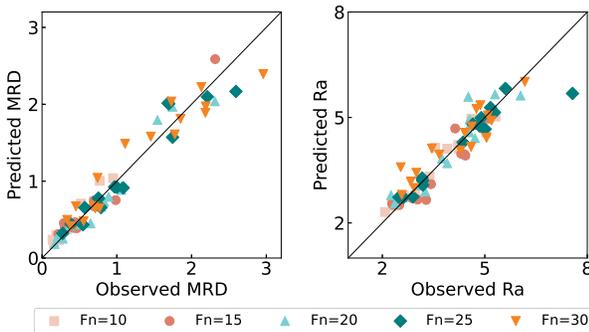


Fig. 5. The MRD and R_a co-prediction results of the proposed PW-DA-DTL method considering various new normal forces.

Table III presents the evaluation metrics for MRD and R_a co-prediction using the proposed PW-DA-DTL method and

five benchmark methods. The results for various F_n values indicate that the PW-DA-DTL method outperforms five benchmark methods in most cases. To provide a comprehensive comparison of the co-prediction methods, the average results across different F_n values (denoted as “All”) are presented. On the one hand, the MMoE method exhibits superior co-prediction performance for new F_n values compared to the shared-bottom method (which uses a single task-shared expert network), the CSE method (which combines task-shared and task-specific expert networks), the OMoE method (which employs a single task-shared gating network with multiple task-shared expert networks), and the CGC method (which utilizes multiple task-specific gating networks, task-shared expert networks, and task-specific expert networks). The results showcase that the introduced MMoE method (which integrates multiple task-specific gating networks with task-shared expert networks) is effective in balancing the correlation and discrepancy between the MRD and R_a prediction tasks.

On the other hand, ablation experiments for new F_n values reveal that the proposed PW-DA-DTL method significantly improves co-prediction accuracy compared to the standard MMoE method. The results for both MRD and R_a predictions demonstrate the effectiveness of the pairwise DA strategy in aligning data distributions between the source and target domains, even with just one labeled target domain sample. The proposed PW-DA-DTL method enables manufacturers to accurately co-predict MRD and R_a in new F_n spaces without extensive data collection, which facilitates machining parameter optimization and quality control in RBG.

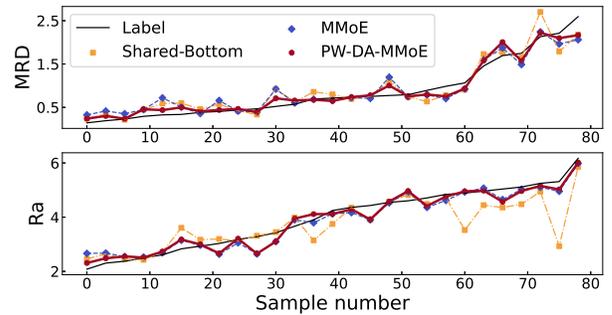


Fig. 6. The predicted MRD and R_a under new F_n using the shared-bottom method, the MMoE method and the proposed PW-DA-DTL method. For the convenience of the result presentation, prediction results are sorted in ascending order according to actual labels and displayed every 3 samples. Methods with significantly poor performance, including CSE, OMoE, and CGC, are excluded from the figure.

B. Case 2: Co-Prediction of MRD and R_a with New Robotic Feed Speed

In this section, the MRD and R_a co-prediction results of the proposed PW-DA-DTL method with different V_w (10 mm/s, 15 mm/s, 20 mm/s, 25 mm/s and 30 mm/s) are displayed in Fig. 7. The left figure illustrates the correlation between the predicted MRD and the observed MRD, and the right figure shows the correlation between the predicted R_a and the observed R_a . Both figures demonstrate a significant consistency between the predicted and observed values. The predicted MRD and R_a under new V_w using different methods are displayed in

TABLE III
 THE MRD AND R_a CO-PREDICTION PERFORMANCE FOR VARIOUS NEW
 NORMAL FORCES USING DIFFERENT METHODS

F_n (N)	Method	RMSE		R^2		MAPE (%)	
		MRD	R_a	MRD	R_a	MRD	R_a
10	Shared-Bottom	0.195	0.792	0.327	0.367	42.492	13.441
	CSE	0.211	0.699	0.208	0.507	61.402	15.067
	OMoE	0.348	0.924	/	0.137	116.930	24.510
	CGC	0.178	0.487	0.437	0.761	49.189	11.316
	MMoE	0.147	0.450	0.618	0.795	45.227	9.918
	PW-DA-MMoE	0.104	0.260	0.807	0.932	24.220	5.694
15	Shared-Bottom	0.133	0.479	0.919	0.783	15.328	8.548
	CSE	0.216	0.788	0.787	0.413	21.136	21.496
	OMoE	0.251	0.975	0.712	0.101	44.165	28.235
	CGC	0.141	0.511	0.910	0.753	15.841	11.209
	MMoE	0.114	0.342	0.941	0.889	15.384	8.372
	PW-DA-MMoE	0.110	0.339	0.945	0.892	15.518	8.094
20	Shared-Bottom	0.157	0.365	0.921	0.897	16.538	6.386
	CSE	0.182	0.854	0.896	0.437	18.469	21.842
	OMoE	0.473	1.163	0.298	/	109.205	28.880
	CGC	0.169	0.414	0.911	0.868	15.613	8.492
	MMoE	0.146	0.378	0.934	0.890	14.601	7.670
	PW-DA-MMoE	0.144	0.377	0.935	0.891	14.557	7.545
25	Shared-Bottom	0.171	0.565	0.933	0.824	9.898	5.671
	CSE	0.231	0.923	0.880	0.531	13.946	16.347
	OMoE	0.417	1.391	0.612	/	57.472	26.760
	CGC	0.181	0.584	0.926	0.812	11.989	5.543
	MMoE	0.207	0.517	0.904	0.853	12.950	5.281
	PW-DA-MMoE	0.185	0.511	0.924	0.856	12.464	5.079
30	Shared-Bottom	0.326	0.479	0.820	0.790	26.506	10.878
	CSE	0.286	1.030	0.863	0.032	21.120	24.695
	OMoE	0.627	1.164	0.342	/	64.313	25.299
	CGC	0.290	0.647	0.859	0.618	24.786	14.798
	MMoE	0.238	0.425	0.905	0.835	18.444	9.486
	PW-DA-MMoE	0.230	0.429	0.911	0.832	17.908	9.431
All	Shared-Bottom	0.196	0.536	0.784	0.732	21.952	8.985
	CSE	0.225	0.859	0.727	0.384	21.215	19.889
	OMoE	0.423	1.123	0.491	0.119	78.417	26.737
	CGC	0.192	0.529	0.809	0.762	23.484	10.272
	MMoE	0.170	0.422	0.860	0.852	21.321	8.145
	PW-DA-MMoE	0.155	0.383	0.904	0.881	16.933	7.169

“/” represents a negative R^2

Fig. 8. The prediction results of both the MMoE method and the PW-DA-DTL method are closer to the actual labels compared to the shared-bottom method. Moreover, the PW-DA-DTL method exhibits smaller errors when compared with the MMoE method for most machining parameters, indicating the effectiveness of the pairwise DA strategy.

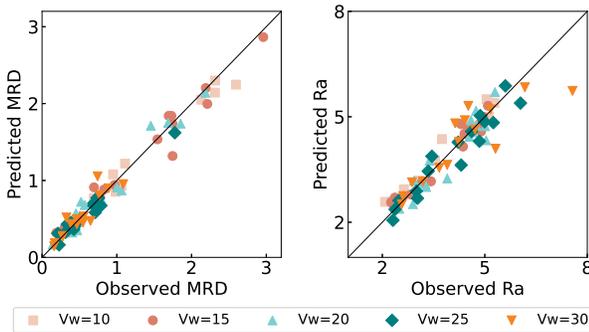


Fig. 7. The MRD and R_a co-prediction results of the proposed PW-DA-DTL method considering various new robotic feed speeds.

In the experiment, the evaluation metrics of MRD and R_a co-prediction using the proposed PW-DA-DTL method and

five benchmark methods are recorded in Table IV. The average results listed in Table IV (as shown in “All”) highlight the advantages of the proposed PW-DA-DTL method for new V_w , which can be explained in two key aspects. Firstly, the MMoE method outperforms other benchmark methods (including Shared-Bottom, CSE, OMoE, and CGC) in co-predicting MRD and R_a under new V_w values. The superiority of the PW-DA-DTL method mainly lies in the employment of the MMoE method, which makes full use of multiple task-specific gating networks and task-shared expert networks to effectively balance task correlation and discrepancy between MRD and R_a predictions. Experimental results demonstrate that the utilization of the MMoE method enables robust modeling of intertwined grinding parameters, even when the space of V_w varies significantly.

Secondly, ablation experiments reveal that the proposed PW-DA-DTL method significantly enhances the prediction performance of the standard MMoE method for MRD prediction, even when trained with just one labeled target domain sample. In addition, the PW-DA-DTL method achieves competitive performance for R_a prediction (with MAPE = 7.094%) compared to the MMoE method (with MAPE = 7.168%). By integrating the pairwise DA strategy, the PW-DA-DTL method effectively aligns data distributions between source (existing V_w spaces) and target (new V_w spaces) domains without heavily relying on labeled target domain samples. The capability of the proposed pairwise DA strategy is particularly critical for industrial applications where acquiring labeled data for every new parameter space is prohibitively expensive and time-consuming.

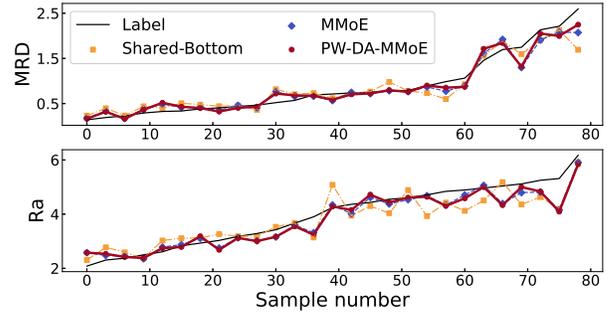


Fig. 8. The predicted MRD and R_a under new V_w using the shared-bottom method, the MMoE method and the proposed PW-DA-DTL method.

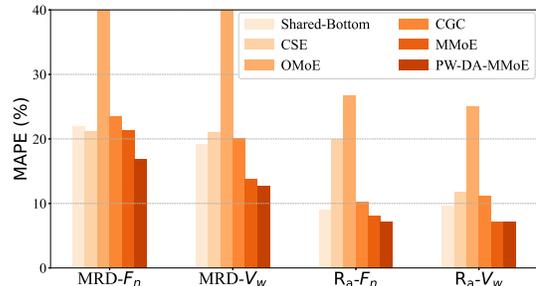


Fig. 9. Comparison among different methods as well as various working conditions using the average MAPE metric.

TABLE IV
 THE MRD AND R_a CO-PREDICTION PERFORMANCE FOR VARIOUS NEW
 ROBOTIC FEED SPEEDS USING DIFFERENT METHODS

V_w (mm/s)	Method	RMSE		R^2		MAPE (%)	
		MRD	R_a	MRD	R_a	MRD	R_a
10	Shared-Bottom	0.427	0.319	0.670	0.921	20.977	6.300
	CSE	0.435	0.695	0.658	0.626	16.870	15.003
	OMoE	0.504	0.998	0.540	0.229	19.930	19.561
	CGC	0.312	0.375	0.824	0.891	18.060	9.066
	MMoE	0.310	0.315	0.827	0.923	12.376	7.980
	PW-DA-MMoE	0.139	0.319	0.965	0.921	9.403	8.198
15	Shared-Bottom	0.183	0.446	0.948	0.809	22.790	10.855
	CSE	0.239	0.409	0.912	0.839	22.528	11.221
	OMoE	0.328	0.621	0.834	0.630	40.301	13.128
	CGC	0.175	0.279	0.952	0.925	19.353	5.662
	MMoE	0.176	0.235	0.952	0.947	16.033	5.664
	PW-DA-MMoE	0.141	0.228	0.969	0.950	13.565	5.485
20	Shared-Bottom	0.142	0.428	0.937	0.809	17.515	9.157
	CSE	0.146	0.405	0.933	0.829	21.147	8.993
	OMoE	0.170	1.023	0.909	/	29.522	19.646
	CGC	0.128	0.458	0.948	0.781	18.813	10.409
	MMoE	0.115	0.326	0.959	0.889	13.327	7.069
	PW-DA-MMoE	0.111	0.321	0.961	0.892	13.640	6.641
25	Shared-Bottom	0.085	0.394	0.943	0.886	15.432	9.031
	CSE	0.117	0.489	0.894	0.824	20.405	10.950
	OMoE	0.245	1.140	0.534	0.043	56.806	28.304
	CGC	0.127	0.722	0.875	0.616	17.285	18.190
	MMoE	0.078	0.331	0.953	0.919	13.111	6.585
	PW-DA-MMoE	0.075	0.344	0.957	0.913	12.985	6.638
30	Shared-Bottom	0.162	0.775	0.615	0.652	19.396	12.498
	CSE	0.177	0.833	0.539	0.598	24.458	12.863
	OMoE	0.234	2.190	0.193	/	53.534	44.450
	CGC	0.187	0.807	0.483	0.623	27.116	12.406
	MMoE	0.114	0.648	0.809	0.757	14.443	8.542
	PW-DA-MMoE	0.112	0.655	0.814	0.752	14.293	8.507
All	Shared-Bottom	0.200	0.472	0.823	0.815	19.222	9.568
	CSE	0.223	0.566	0.787	0.743	21.082	11.806
	OMoE	0.296	1.194	0.602	0.301	40.019	25.018
	CGC	0.186	0.528	0.816	0.767	20.125	11.147
	MMoE	0.159	0.371	0.900	0.887	13.858	7.168
	PW-DA-MMoE	0.116	0.373	0.933	0.886	12.777	7.094

“/” represents a negative R^2

Fig. 9 shows the comparison results among six methods (Shared-Bottom, CSE, OMoE, CGC, MMoE, and PW-DA-DTL) performed on two case studies, where the average MAPE is adopted as the evaluation metric for MRD and R_a prediction. Across both case studies, experimental results highlight the superiority of the proposed PW-DA-DTL method over selected benchmark methods (including Shared-Bottom, CSE, OMoE, and CGC) in co-predicting MRD and R_a for new RBG machining parameter spaces. Notably, the integration of the pairwise DA strategy enables improving the performance of the standard MMoE method under data scarcity in new machining parameter spaces. The findings underscore the practical value of the proposed method for industrial applications, where rapid adaptation to new machining parameter spaces is critical for maintaining precision and reducing operational costs.

VI. CONCLUSION

In this article, a novel PW-DA-DTL method has been introduced for MRD and R_a co-prediction with regard to unseen spaces of RBG machining parameters. The MMoE, a dual-task learning technique, has been employed to simultaneously predict MRD and R_a by balancing their cor-

relation and discrepancy. A pairwise DA strategy has been developed to significantly strengthen the labeled samples in target domain, thereby effectively reducing the cross-domain distribution discrepancy. The proposed PW-DA-DTL method has been tested on a self-developed RBG system with various grinding parameters. Experimental results have demonstrated that the proposed PW-DA-DTL method exhibits significant superiority in MRD and R_a co-prediction of RBG compared to five benchmark methods.

In the future, we aim to: 1) develop novel evolutionary computation methods to optimize the machining parameters in RBG on the basis of the PW-DA-DTL model [41], [42]; 2) compare different distance functions of domain adaptation for MRD and R_a co-prediction in RBG [43]–[45]; 3) adopt various filtering and signal processing methods to pre-process the RBG dataset [46]–[48]; and 4) apply our PW-DA-DTL method to other multi-task learning applications such as medical engineering, path planning and object detection [49], [50].

REFERENCES

- [1] X. Ren, X. Huang, K. Gao, L. Xu, H. Feng, X. Zhang, H. Chen, Z. Chai, and X. Chen, A review of recent advances in robotic belt grinding of superalloys, *The International Journal of Advanced Manufacturing Technology*, vol. 127, pp. 1447-1482, 2023.
- [2] W. He, J. Mao, Y. Wang, Z. Li, H. Xie, H. Shao, and X. Zhao, Unified diagnostic and matching framework of fault and quality for robotic grinding system, *IEEE Transactions on Instrumentation and Measurement*, vol. 73, art. no. 3513612, 2024.
- [3] B. Li, J. Ni, J. Yang, and S. Y. Liang, Study on high-speed grinding mechanisms for quality and process efficiency, *The International Journal of Advanced Manufacturing Technology*, vol. 70, pp. 813-819, 2014.
- [4] F. Li, Y. Xue, Z. Zhang, W. Song, and J. Xiang, Optimization of grinding parameters for the workpiece surface and material removal rate in the belt grinding process for polishing and deburring of 45 steel, *Applied Sciences*, vol. 10, no. 18, art. no. 6314, 2020.
- [5] S. Sunder and V. Yadava, Modelling and optimisation of material removal rate and surface roughness in surface-electrical discharge diamond grinding process, *International Journal of Industrial and Systems Engineering*, vol. 17, no. 2, pp. 133-151, 2014.
- [6] W. Liu, Z. Wang, Y. Yuan, N. Zeng, K. Hone, and X. Liu, A novel sigmoid-function-based adaptive weighted particle swarm optimizer, *IEEE Transactions on Cybernetics*, vol. 51, no. 2, pp. 1085-1093, 2019.
- [7] H. Li, H. Liu, C. Lan, Y. Yin, P. Wu, C. Yan, and N. Zeng, SMWO/D: a decomposition-based switching multi-objective whale optimiser for structural optimisation of turbine disk in aero-engines, *International Journal of Systems Science*, vol. 54, no. 8, pp. 1713-1728, 2023.
- [8] J. Fang, W. Liu, L. Chen, S. Lauria, A. Miron, and X. Liu, A survey of algorithms, applications and trends for particle swarm optimization, *International Journal of Network Dynamics and Intelligence*, vol. 2, no. 1, pp. 24-50, 2023.
- [9] M. A. El-Shorbagy, A. Bouaouda, H. A. Nabwey, L. Abualigah, and F. A. Hashim, Bald eagle search algorithm: A comprehensive review with its variants and applications, *Systems Science & Control Engineering*, vol. 12, no. 1, art. no. 2385310, 2024.
- [10] R. Prabhu and M. Kanthababu, Prediction of surface roughness and depth of cut in abrasive waterjet milling of alumina ceramic using machine learning algorithms, *Expert Systems with Applications*, vol. 246, art. no. 123168, 2024.
- [11] J. Ma, Z. Zhao, X. Yi, J. Chen, L. Hong, and E. H. Chi, Modeling task relationships in multi-task learning with multi-gate mixture-of-experts, In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, UK, Aug. 2018, pp. 1930-1939.
- [12] Y. Zhang, Y. Xin, Z.-W. Liu, M. Chi, and G. Ma, Health status assessment and remaining useful life prediction of aero-engine based on BiGRU and MMoE, *Reliability Engineering & System Safety*, vol. 220, art. no. 108263, 2022.

- [13] Y. Nagaraj, N. Jagannatha, N. Sathisha, and S. Niranjana, Prediction of material removal rate and surface roughness in hot air assisted hybrid machining on soda-lime-silica glass using regression analysis and artificial neural network, *Silicon*, vol. 13, no. 11, pp. 4163-4175, 2021.
- [14] Z. Yang, X. Xu, J. Li, D. Zhu, S. Yan, S. S. Ge, and H. Ding, Knowledge-wrapping method for prediction and evaluation of material removal behavior in robotic belt grinding, *Mechanical Systems and Signal Processing*, vol. 208, art. no. 110914, 2024.
- [15] S. J. Pan and Q. Yang, A survey on transfer learning, *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345-1359, 2009.
- [16] Y.-L. He, L. Chen, and Q.-X. Zhu, Quality regularization based semisupervised adversarial transfer model with unlabeled data for industrial soft sensing, *IEEE Transactions on Industrial Informatics*, vol. 20, no. 2, pp. 1190-1197, 2024.
- [17] C. Wang, Z. Wang, and H. Dong, A novel prototype-assisted contrastive adversarial network for weak-shot learning with applications: Handling weakly labeled data, *IEEE/ASME Transactions on Mechatronics*, vol. 29, no. 1, pp. 533-543, 2024.
- [18] C. Wang, Z. Wang, W. Liu, Y. Shen, and H. Dong, A novel deep offline-to-online transfer learning framework for pipeline leakage detection with small samples, *IEEE Transactions on Instrumentation and Measurement*, vol. 72, art. no. 3503913, 2022.
- [19] D. Wu, V. J. Lawhern, S. Gordon, B. J. Lance, and C.-T. Lin, Driver drowsiness estimation from EEG signals using online weighted adaptation regularization for regression (OwARR), *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 6, pp. 1522-1535, 2016.
- [20] G. Ma, S. Xu, T. Yang, Z. Du, L. Zhu, H. Ding, and Y. Yuan, A transfer learning-based method for personalized state of health estimation of lithium-ion batteries, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 1, pp. 759-769, 2022.
- [21] J. Fang, Z. Wang, W. Liu, L. Chen, and X. Liu, A new particle-swarm-optimization-assisted deep transfer learning framework with applications to outlier detection in additive manufacturing, *Engineering Applications of Artificial Intelligence*, vol. 131, art. no. 107700, 2024.
- [22] G. Ma, Z. Wang, W. Liu, J. Fang, Y. Zhang, H. Ding, and Y. Yuan, Estimating the state of health for lithium-ion batteries: A particle swarm optimization-assisted deep domain adaptation approach, *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 7, pp. 1530-1543, 2023.
- [23] Y. Jiang, T. Xia, D. Wang, X. Fang, and L. Xi, Adversarial regressive domain adaptation approach for infrared thermography-based unsupervised remaining useful life prediction, *IEEE Transactions on Industrial Informatics*, vol. 18, no. 10, pp. 7219-7229, 2022.
- [24] Y. Ren, J. Liu, H. Zhang, and J. Wang, TBDA-Net: A task-based bias domain adaptation network under industrial small samples, *IEEE Transactions on Industrial Informatics*, vol. 18, no. 9, pp. 6109-6119, 2022.
- [25] Y. Song, H. Yang, and H. Lv, Intelligent control for a robot belt grinding system, *IEEE Transactions on Control Systems Technology*, vol. 21, no. 3, pp. 716-724, 2012.
- [26] Z. Fan, Q. Xu, C. Jiang, and S. X. Ding, Deep mixed domain generalization network for intelligent fault diagnosis under unseen conditions, *IEEE Transactions on Industrial Electronics*, vol. 71, no. 1, pp. 965-974, 2023.
- [27] C. Wang, Z. Wang, H. Liu, H. Dong, and G. Lu, An optimal unsupervised domain adaptation approach with applications to pipeline fault diagnosis: Balancing invariance and variance, *IEEE Transactions on Industrial Informatics*, vol. 20, no. 8, pp. 10019-10030, 2024.
- [28] C. Wang, Z. Wang, Q. Liu, H. Dong, and W. Sheng, Support-sample-assisted domain generalization via attacks and defenses: Concepts, algorithms, and applications to pipeline fault diagnosis, *IEEE Transactions on Industrial Informatics*, vol. 20, no. 4, pp. 6413-6423, 2024.
- [29] Y. Zhang and Q. Yang, An overview of multi-task learning, *National Science Review*, vol. 5, no. 1, pp. 30-43, 2018.
- [30] M. Long, J. Wang, G. Ding, S. J. Pan, and S. Y. Philip, Adaptation regularization: A general framework for transfer learning, *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 5, pp. 1076-1089, 2013.
- [31] R. Caruana, Multitask learning, *Machine Learning*, vol. 28, pp. 41-75, 1997.
- [32] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, and A. Smola, A kernel method for the two-sample-problem, In: *Proceedings of the 19th International Conference on Neural Information Processing Systems*, Vancouver, Canada, Dec. 2006.
- [33] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, Domain adaptation via transfer component analysis, *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199-210, 2010.
- [34] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, Deep model based domain adaptation for fault diagnosis, *IEEE Transactions on Industrial Electronics*, vol. 64, no. 3, pp. 2296-2305, 2016.
- [35] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, Generative adversarial networks: An overview, *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53-65, 2018.
- [36] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, SMOTE: Synthetic minority over-sampling technique, *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002.
- [37] S. Li, L. Song, X. Wu, Z. Hu, Y.-M. Cheung, and X. Yao, Multi-class imbalance classification based on data distribution and adaptive weights, *IEEE Transactions on Knowledge and Data Engineering*, in press, DOI: 10.1109/TKDE.2024.3384961.
- [38] Y. Wang, C. Wen, and X. Wu, Fault detection and isolation of floating wind turbine pitch system based on Kalman filter and multi-attention 1DCNN, *Systems Science & Control Engineering*, vol. 12, no. 1, art. no. 2362169, 2024.
- [39] D. Wang, C. Wen, and X. Feng, Deep variational Luenberger-type observer with dynamic objects channel-attention for stochastic video prediction, *International Journal of Systems Science*, vol. 55, no. 4, pp. 728-740, 2024.
- [40] H. Tang, J. Liu, M. Zhao, and X. Gong, Progressive layered extraction (PLE): A novel multi-task learning (MTL) model for personalized recommendations, In: *Proceedings of the 14th ACM Conference on Recommender Systems*, New York, USA, Sep. 2020, pp. 269-278.
- [41] W. Liu, Z. Wang, N. Zeng, F. E. Alsaadi, and X. Liu, A PSO-based deep learning approach to classifying patients from emergency departments, *International Journal of Machine Learning and Cybernetics*, vol. 12, pp. 1939-1948, 2021.
- [42] Y. Xue, M. Li, H. Arabnejad, D. Suleimenova, A. Jahani, B. C. Geiger, F. Boesjes, A. Anagnostou, S. J. E. Taylor, X. Liu, and D. Groen, Many-objective simulation optimization for camp location problems in humanitarian logistics, *International Journal of Network Dynamics and Intelligence*, vol. 3, no. 3, art. no. 100017, 2024.
- [43] C. Cheng, B. Zhou, G. Ma, D. Wu, and Y. Yuan, Wasserstein distance based deep adversarial transfer learning for intelligent fault diagnosis with unlabeled or insufficient labeled data, *Neurocomputing*, vol. 409, pp. 35-45, 2020.
- [44] Y. Wang, H. Fang, J. Jin, G. Ma, X. He, X. Dai, Z. Yue, C. Cheng, H.-T. Zhang, D. Pu *et al.*, Data-driven discovery of stochastic differential equations, *Engineering*, vol. 17, pp. 244-252, 2022.
- [45] W. Liu, Z. Wang, L. Tian, S. Lauria, and X. Liu, Melt pool segmentation for additive manufacturing: A generative adversarial network approach, *Computers & Electrical Engineering*, vol. 242, art. no. 108382, 2021.
- [46] Y. Liu, Z. Wang, C. Liu, M. Coombes, and W.-H. Chen, A novel algorithm for quantized particle filtering with multiple degrading sensors: Degradation estimation and target tracking, *IEEE Transactions on Industrial Informatics*, vol. 19, no. 4, pp. 5830-5838, 2023.
- [47] Y.-A. Wang, B. Shen, L. Zou, and Q.-L. Han, A survey on recent advances in distributed filtering over sensor networks subject to communication constraints, *International Journal of Network Dynamics and Intelligence*, vol. 2, no. 2, art. no. 100007, Jun. 2023.
- [48] X. Gao, F. Deng, W. Shang, X. Zhao, and S. Li, Attack-resilient asynchronous state estimation of interval type-2 fuzzy systems under stochastic protocols, *International Journal of Systems Science*, vol. 55, no. 13, pp. 2688-2700, 2024.
- [49] W. Fang, B. Shen, A. Pan, L. Zou, and B. Song, A cooperative stochastic configuration network based on differential evolutionary sparrow search algorithm for prediction, *Systems Science & Control Engineering*, vol. 12, no. 1, art. no. 2314481, 2024.
- [50] Y. Zhang, R. Lan, X. Li, J. Fang, Z. Ping, W. Liu, and Z. Wang, Class imbalance wafer defect pattern recognition based on shared-database decentralized federated learning framework, *IEEE Transactions on Instrumentation and Measurement*, vol. 73, art. no. 2517317, 2024.