2025 IEEE 21st International Conference on Automation Science and Engineering (CASE)
August 17-21, 2025. Los Angeles, CA, US

2025 IEEE 21st International Conference on Automation Science and Engineering (CASE) | 979-8-3315-2246-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/CASE58245.2025.11163877

DCAF: Dynamic Cross-Attention Feature Fusion from Robotic Anomaly Detection to Position Accuracy Modeling

Hui Liu[1], Guixiu Qiao[3], Pavel Piliptchak[3], James Moore[4], Daniela Sawyer[4], Yingyan Zeng[2], Ran Jin[1]

***Abstract*— In robotic operations, heterogeneous computation tasks and sensor configurations pose significant challenges to analyze different modalities of data for data sharing and collaborative learning in robotic Artificial Intelligence (AI) tasks. The lack of historical data in new scenarios or new computation tasks complicates model training and limits the applicability of existing AI methodologies. Current transfer learning approaches heavily rely on static feature extraction, which fail to dynamically adjust to specific feature relationships between different samples or modalities. In the literature, these methods struggle to capture inter-modal associations effectively, resulting in insufficient information sharing and poor modeling performance. Motivated by these challenges, this paper proposes a Dynamic Cross-Attention Feature Fusion (DCAF) approach to map the features from one robotic AI task to another. By calculating attention weights tailored to each target domain sample, DCAF extracts the most relevant source domain features and generates dynamic fused representations. The proposed approach enables sample-specific feature selection and fine-grained domain alignment, effectively enhancing the modeling performance compared with traditional transfer learning and model training based on the local data source. It is particularly suited for a new robotic AI training task with limited sample size and new data modalities. Experimental results for feature fusion from a robotics anomaly detection dataset to a position accuracy modeling data set demonstrate the effectiveness of DCAF, providing an efficient solution for domain adaptation and multimodal fusion.**

Index Terms—Cross-attention, Feature fusion, Robotic anomaly detection

I. INTRODUCTION

Manufacturing Industrial Internet integrates AI and automation with broad applications of industrial robots for various manufacturing tasks, such as material handling, manufacturing operations (e.g., welding, additive manufacturing, and assembly), and quality inspection. As robotic tasks become more complex, AI methods become more important in those robotic tasks, including perception and multimodal data fusion [1], motion planning and control [2], anomaly detection and self-healing [3], human-robotics collaborations [4], etc. For example, robotic anomaly detection is essential for ensuring safety, minimizing downtime, and improving robotic availability and operational efficiency [3]. For another example, predicting the accuracy of the position is the foundation to reduce the discrepancy between the actual positions and the target [5] to control robotic motion. The position error assesses performance and detects robotic failures, such as sensor drift, mechanical wear, or control system anomalies, which is critical in robotic operations.

The data quality plays a foundational role to train adequate AI models for robots [3]. Given robots and tasks in a new context, limited annotated data and the commonly encountered class imbalance pose significant challenges [6]. Therefore, it is important to leverage datasets from other similar data sources, which may improve the performance and generalization of the AI models. However, heterogeneous sensor types, data formats, and measurement methods may cause inconsistencies in the data modalities [1]. It will complicate direct data transfer and limit the adaptability of AI models in a new AI context. Additionally, AI tasks in new contexts may also become different.

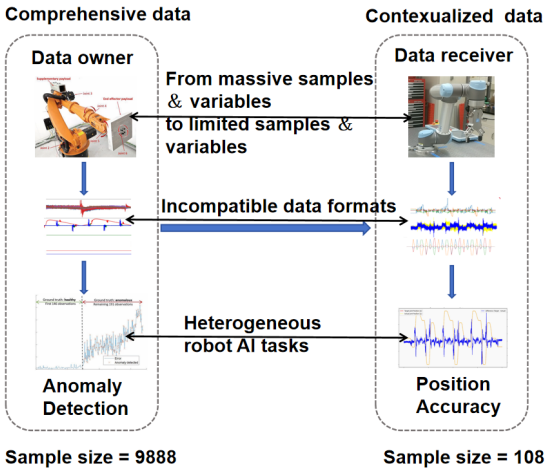

Fig. 1. An Example of Information Transfer Between Two Robotic AI Tasks (Redrawn from [5] and [7] with Authors' Permission)

Figure 1 illustrates an example of this discrepancy between the data modality and AI tasks of two robotic AI tasks. While the more comprehensive dataset [7] has more samples (n=9888) to train a multi-class classification model for robotic anomaly detection, the contextualized dataset [5] has much fewer samples (n=108) to predict the accuracy of robot system's tool center position (TCP). Our objective is

*This work was supported by Virginia Tech and the National Institute of Standards and Technology.

[1]Hui Liu and Ran Jin are with the Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA, USA hui.liu@vt.edu, ran.jin@vt.edu

[2]Yingyan Zeng is with the Department of Mechanical and Materials Engineering, University of Cincinnati, Cincinnati, OH, USA zengyy@ucmail.uc.edu

[3]Guixiu Qiao and Pavel Piliptchak are with the National Institute of Standards and Technology, Gaithersburg, MD, USA guixiu.qiao@nist.gov, pavel.piliptchak@nist.gov

[4]J. Moore and D. Sawyer are with the University of Sheffield Advanced Manufacturing Research Centre (AMRC), Factory 2050, Europa Way, Sheffield, S9 1ZA, UK j.moore@amrc.co.uk, d.sawyer@amrc.co.uk

to create an adequate position accuracy prediction model and evaluate whether bringing in the more comprehensive dataset from the data owner to the data receiver will significantly improve modeling performance. It requires us not only to tackle the challenges of incompatible data formats or modalities, but also to use the information transfer effectively in heterogeneous AI tasks.

There are gaps to transfer useful data, information, or knowledge from one AI task to another. For example, transfer learning is a machine learning paradigm that uses knowledge from a source domain to enhance model performance in a relatedtarget domain [8]. However, most of the transfer learning methods rely on fixed feature alignment strategies (e.g., MMD, CORAL [9], [10]), which fail to dynamically adapt feature representations across different scenarios, thereby reducing model adaptability and performance [3]. Additionally, most transfer learning approaches assume that all the source domain features are relevant to the target domain, lacking a sample-level feature selection mechanism to further improve the transfer effectiveness [8]. In the data modality-inconsistent settings, global alignment strategies introduce redundant features and limiting information sharing. Furthermore, these methods often assume task similarity or direct correspondence between domains, making them insufficient for handling task inconsistency.

To address these challenges, this paper proposes a Dynamic Cross-Attention Feature Fusion (DCAF) method for robotic position accuracy prediction problems by leveraging information from anomaly detection. Unlike traditional transfer learning approaches, the proposed method computes attention weights for each sample in the target context, extracting the most relevant features from the more comprehensive dataset, and generating fused feature representations dynamically. This enables sample-level feature selection and fine-grained domain alignment, ensuring that only the most relevant information contributes to the final prediction. By adopting the cross-attention mechanisms [11] the method effectively captures associations between different modalities, enhances feature sharing, and overcomes the limitations of static feature alignment. The proposed method is validated through a case study using datasets from the AMRC [7] and the National Institute of Standards and Technology (NIST) [5]. Experimental results on the anomaly detection and the position accuracy prediction model demonstrate that the proposed method outperforms existing approaches, exhibiting superior adaptability and robustness in an environment of multiple robots, different sensor configurations, and heterogeneous AI task.

The remainder of this paper is organized as follows: Section 2 reviews related work; Section 3 details the proposed method; Section 4 presents experimental validation results and explanations; and Section 5 concludes the study with discussions on future research directions.

II. RELATED WORK

A. Robotic Anomaly Detection

Robotic anomaly detection plays a crucial role in industrial automation, ensuring equipment reliability [7]. Traditional approaches primarily rely on predefined rules and expert input, making them rigid and less effective in handling complex, dynamic industrial settings [3].

With the advancement of machine learning, data-driven anomaly detection methods have become the dominant approach. Supervised learning models like deep neural networks (DNNs) have been widely applied in anomaly classification tasks [12]. However, the class imbalance due to less frequent anomalies results in modeling effectiveness of supervised learning approaches. Unsupervised and semi-supervised methods attempt to address this issue by modeling normal operational patterns and identifying deviations [13]. Nevertheless, these approaches often struggle with feature selection and domain adaptation, particularly in the presence of heterogeneous sensors and data modalities.

B. Transfer Learning and Attention Mechanism

Transfer learning plays a crucial role in robotic anomaly detection where annotated data are limited. Existing methods primarily include feature alignment, adversarial training, and parameter fine-tuning. Feature alignment techniques achieve transfer learning by minimizing the distribution discrepancy between the source and target domains [10]. Adversarial training methods, such as Domain-Adversarial Neural Networks (DANN), learn domain-invariant features through a domain discriminator [12]. However, transfer learning cannot be directly applied when there is modality inconsistency [1]. Due to variations in sensor types, data formats, and measurement methods across different robots, traditional transfer learning approaches cannot transfer features effectively, leading to reduced modeling performance. Additionally, conventional methods employ fixed feature mapping strategies, preventing dynamic feature adaptation to diverse task requirements.

To address these issues, attention mechanisms have been introduced to dynamically allocate feature weights, enhancing feature selection and adaptability [14]. Attention mechanisms allow models to identify the most relevant features for a given task while filtering out irrelevant information [11]. Self-attention has been widely applied in time series analysis, capturing long-term dependencies within data [8]. Meanwhile, cross-attention facilitates more effective feature interaction between the source and target domains, improving knowledge transfer efficiency [11]. In robotic anomaly detection tasks, cross-attention enables dynamic feature selection and fine-grained feature alignment by computing the correlation between target domain samples and source domain features, extracting only the most relevant information for fusion [11]. Furthermore, cross-attention allows models to adapt feature representations dynamically to varying data distributions, improving generalization across multiple devices and heterogeneous modalities.

III. METHODOLOGY

This study proposes a robotic AI modeling method based on a cross-attention mechanism for feature fusion from one data source to another. In our problem formulation, we refer the source domain as the data owner, with its dataset termed the shared dataset, while the target domain is referred to as the data receiver, with its original dataset termed the local dataset. Figure 2 illustrates the data sharing and feature fusion framework based on DCAF, which consists of two key components: (1) Feature extraction. The framework extracts features from both the shared dataset and the local dataset, employing adaptive feature extraction methods tailored to different data modalities. These extracted features encapsulate critical information from multiple robots and data modalities, ensuring comprehensive data representation. (2) Cross-attention mechanism. The extracted features go through a cross-attention operation, where the local dataset features serve as queries (Q), while the shared dataset features function as keys (K) and values (V). The fused features are then processed by a feed-forward network (FFN) and utilized for downstream AI tasks. This process ensures that only the most relevant and informative features from the data owner contribute to the AI model at the data receiver, thereby enhancing the adaptability and robustness of DCAF in heterogeneous data and AI task environments. This approach effectively addresses the challenges of heterogeneous data integration, enhancing model adaptability, and providing an efficient solution for robotic analytics.

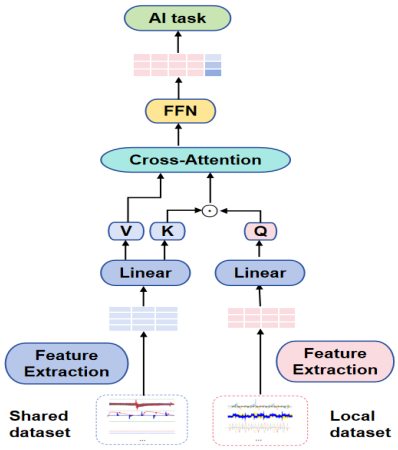

Fig. 2. The Proposed DCAF Method

The major contribution of this method lies in the dynamic feature fusion mechanism based on cross-attention. It enables the model to calculate the attention weight for each sample of the data receiver, extract the most relevant shared features, and achieve sample-level feature selection and fine-grained domain alignment. Therefore, it improves feature sharing efficiency and reduces the impact of redundant information. But it also addresses the limitations of traditional static feature alignment, captures the association between different modalities through the cross-attention mechanism, and dynamically adjusts the feature representation to improve the AI modeling performance for the data receiver.

A. Feature Extraction

The feature extraction process integrates Fast Fourier Transform (FFT) and wavelet transform to obtain time-frequency domain features and compute statistical metrics. This approach captures both local and global signal characteristics, enhancing the accuracy of anomaly detection, position accuracy prediction, and state recognition. In robotic systems, signals such as acceleration, force, and torque are collected to monitor performance and detect anomalies. FFT identifies spectrum patterns related to mechanical wear or control issues, while wavelet transform captures transient anomalies through multi-resolution analysis.

For time-domain features, raw time-domain signals (e.g., acceleration, force, torque) are decomposed by using wavelet transform, which provides multi-resolution analysis capabilities, capturing local signal features in both time and frequency (scale) domains [15]. This approach is sensitive to transient or localized anomalies and features in different frequency bands, improving the accuracy of fault detection and state recognition. The signal is decomposed into four levels (Level 1–4) of wavelet coefficients. Then, at each level, statistical metrics includes mean, standard deviation, energy ($E = \sum_{i=1}^{N} x_i^2$), maximum and minimum values, skewness, and kurtosis are computed. Energy reflects the overall power of the signal, while skewness and kurtosis provide insight into signal symmetry and tail distribution, respectively.

For frequency-domain features, the Fast Fourier Transform (FFT) is used to obtain the amplitudes of each harmonic [16], and the total harmonic distortion (THD) is calculated to quantify signal distortion and energy distribution across different harmonic orders. THD is defined as:

$$\text{THD} = \frac{\sqrt{\sum_{n=2}^{\infty} A_n^2}}{A_1} \tag{1}$$

where A_1 is the amplitude of the fundamental frequency, and A_n represents the amplitudes of higher-order harmonics. Additional frequency-domain features include harmonic mean, entropy-based measures ($H(x) = -\sum p(x) \log p(x)$), root mean square (RMS), peak values, and differences between maximum and minimum values. Entropy quantifies the complexity of the signal, while RMS provides an estimate of its magnitude.

For environmental features such as temperature, the mean and standard deviation are used to represent the overall temperature level and fluctuations under operating conditions [16].

By combining both time-domain and frequency-domain features, a more accurate representation of the signals is obtained, facilitating system operating condition analysis and fault pattern recognition.

B. Cross-Attention mechanism

Cross-Attention is used to learn the relationship between local and sharing features, dynamically selecting the most valuable shared features based on local features for fusion,

thereby enhancing the predictive capability of the regression task. Through the Query-Key-Value mechanism, Cross-Attention maps local and sharing features into a unified space and selects the most relevant shared features via adaptive attention weights, rather than simply concatenating or averaging all features [11]. This approach not only improves feature alignment, making data from different sources more compatible, but also effectively filters redundant information, ensuring the model focuses on the most valuable features to enhance prediction accuracy. Additionally, Cross-Attention improves the model's generalization ability, making it more robust across different data distributions, thereby leveraging shared data more effectively to optimize regression performance. A linear mapping projects the local feature into a query vector, while the shared feature is projected into the Key and the Value vectors [11]. Specifically, the query vector is obtained as $Q_{\text{local}} = X_{\text{local}} W_Q$, while the key and value vectors are computed as $K_{\text{shared}} = X_{\text{shared}} W_K$ and $V_{\text{shared}} = X_{\text{shared}} W_V$, respectively. where W_Q, W_K, W_V are trainable transformation matrices mapping features into a unified space, with X_{local} as local sample features and X_{shared} as sharing sample features.

Next, cross-attention is applied to compute attention weights and extract dynamic features [11], which generates a larger weight for the sharing features with high similarity to the local features:

$$A = \text{softmax}\left(\frac{Q_{\text{local}} K_{\text{shared}}^T}{\sqrt{d_k}}\right), and \tag{2}$$

$$F_{\text{shared}} = A V_{\text{shared}}, \tag{3}$$

where A is the attention weight matrix, and F_{shared} represents the extracted dynamic features from the sharing for each local sample. Finally, feature fusion is performed to integrate the extracted features:

$$F_{\text{fused}} = \text{FFN}\big(\text{concat}(F_{\text{shared}}, X_{\text{local}})\big), \tag{4}$$

where FFN represents a feed-forward network applied to the concatenated features.

To train the FNN, the mean squared error (MSE) loss is used for regression task:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2, \tag{5}$$

where $\hat{y}_i$ is the model prediction for sample i, and y_i is the ground truth label. By optimizing this loss using the Adam optimization method [17], the model learns to highlight the most relevant shared features, effectively integrating local and shared information for improved regression accuracy.

IV. CASE STUDY

A. Experiment Description

This study uses an anomaly detection dataset from the ARMC as the shared dataset and a position accuracy prediction dataset from NIST as the local dataset to evaluate the DCAF method. The objective is to improve the position accuracy prediction with the help of the shared anomaly detection dataset and dynamic feature fusion. This is primarily because the shared dataset has much larger sample size with more complete sensor networks to measure more variables, which may enhance the position accuracy prediction performance. The regression model used in both Concat-FFN and DCAF consists of a feed-forward neural network with one hidden layer of 32 units and ReLU activation. The two datasets differ in data modalities and sample sizes. Detailed information is shown in Table I below.

The shared dataset originates from the AMRC and contains signals collected during fingerprint routines [7]. Eachincludes isolated motions of one joint at speeds of 10%, 50%, and 100% of the maximal speed. The features recorded in each routine include acceleration or vibration (six channels at 25,600 Hz), current of the motors (three channels at 1,000 Hz), temperature (three channels at 1 Hz), as well as the elapsed times. After data pre-processing, there are 9888 samples and 224 features as predictors.

The local dataset provides advanced TCP position error for robots along with detailed information about controller-level components. Data were collected at 125 Hz, comparing nominal positions with measured (actual) positions. Additional controller-level sensor data for each joint were also collected to analyze the impact of factors, such as temperature, payload, and speed on position degradation. This includes information on actual positions, speed, current, acceleration, torque, and temperature. The dataset consists of 108 samples and 280 features as predictors. The data analysis will provide insights to diagnose the root causes of robot performance degradation, supporting a comprehensive understanding of anomaly in robotic systems.

TABLE I
COMPARISON OF THE SHARED DATASET AND THE LOCAL DATASET

Variables	Shared Dataset from the AMRC Data	Local Dataset at NIST
Current	✓	✓
Acceleration	✓	
Velocity		✓
Force		✓
Temperature	✓	
Speed	✓	
Elapsed Time	✓	✓
Response Variable	Five Robotic Health Classes	Position Error
Sample Size	9888	108

B. Results and Discussion

The AI task for the data receiver in this experiment is a regression task, where the predictor variables are detailed in the Experiment Description section. The response variables are Mean Absolute Error (MAE) and the logarithm of the variance of the errors ($\ln S^2$), which are predicted separately. The experiment employs Normalized Root Mean Square Error (NRMSE) as the evaluation metric to measure prediction error normalized by the range of the response variables. The NRMSE ensures comparability of performance comparison

across different scales, mitigating the impact of unit differences and maintaining consistency in modeling performance evaluation across datasets.

Given the high dimensionality of features, L1 regularization ($R(\theta) = \lambda \sum_{j=1}^{p} |\theta_j|$) is applied for feature selection to enforce sparsity and eliminate features with minimal contribution to the regression task [18]. We compare the DCAF with several well-known benchmark methods:

Adaptive Intermediate Class-Wise Distribution Alignment (AICDA) develops an adaptive intermediate distribution as the alignment goal for multiple source and target domains. It aligns both global and class-wise distributions effectively [19].

Transfer learning-based data anomaly-detection (TLBDA) leverages the similarity of anomalous patterns across different bridges and utilizes a deep neural network to achieve high-accuracy anomaly identification for bridge groups [13].

For the baseline logistic regression (LR) model based on the local dataset alone, Lasso feature selection is applied beforehand with the same parameters as DCAF. For the benchmark methods AICDA and TLBDA, which require identical feature dimensions across datasets, Principal Component Analysis (PCA) is first used to reduce the channels of the shared dataset from six channels to one. This ensures alignment with the local dataset, facilitating a consistent feature representation across datasets. Subsequently, each method's built-in feature selector is applied for further refinement. The experiments are conducted on the Torch 2.4.1 platform with a learning rate of 0.001. As an additional baseline, the FFN method directly concatenates the features from the shared and local datasets, followed by a feed-forward network for regression.

The performance on the test set is evaluated using five-fold cross-validation. The experimental results are summarized in Table III, which shows that the DCAF model achieves smaller NRMSE for both responses (MAE and $\ln S^2$) than other benchmark methods. DCAF performs better in local

TABLE II

MODEL PERFORMANCE OF DIFFERENT METHODS

Source Domain	Methods	NRMSE (ln S^2)	NRMSE (MAE)
NA	LR	0.0844 ± 0.0192	0.0956 ± 0.0285
Total	AICDA	0.0752 ± 0.0135	0.0887 ± 0.0162
	TLDBA	0.0979 ± 0.0229	0.0842 ± 0.0196
	FFN	0.0470 ± 0.0125	0.0564 ± 0.0153
	DCAF	$\mathbf{0.0450} \pm 0.0079$	$\mathbf{0.0556} \pm 0.0180$

AI modeling, primarily because it adaptively captures the correlations between the shared dataset (classification task) and the local dataset (regression task). By leveraging a dynamic weighting mechanism, it amplifies the influence of key features while suppressing irrelevant or redundant ones. Unlike static feature transformation methods such as AICDA and TLBDA, which apply the same feature mapping for all samples, cross-attention allows each regression sample in the local dataset to selectively emphasize the most relevant features from the shared dataset based on its similarity to all classification samples. This adaptive feature selection process ensures that the most informative features are utilized for each prediction, enhancing the model's ability to generalize across datasets with different distributions. As a result, DCAF outperforms conventional methods, demonstrating its superior effectiveness and robustness.

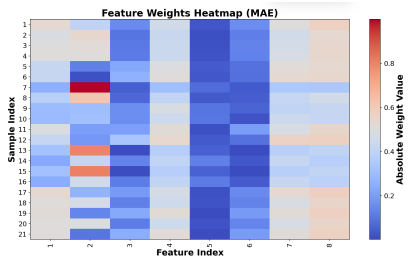

Fig. 3. Feature weight value for MAE

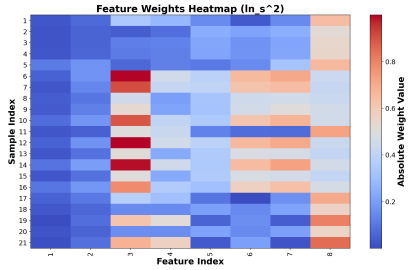

Fig. 4. Feature weight value for lns^2

The heatmaps shown in Figures 3 and 4 illustrate the absolute weight values of selected features from the classification problem based on the shared dataset to the samples in the two regression tasks based on the local dataset (i.e., predicting MAE and ln(S²)).

The horizontal axis represents the classification features selected after searching for the optimal threshold using the Support Vector Machine (SVM model), which include eight features. All features are derived from accelerometer signals recorded six distinct channels. Each feature corresponds to energy measurements obtained from the detailed levels of wavelet decomposition, thereby capturing high-frequency signals. These features were chosen to maximize classification accuracy and were subsequently incorporated into the regression model to enhance its predictive performance. These features were selected to maximize classification accuracy and were subsequently applied to the regression model to enhance its predictive performance.

The color in the heatmap represents the importance of each feature, where red indicates high weights and blue indicates low weights. In Figure 3 (MAE heatmap) and Figure 4 (ln(S²) heatmap), certain two features from accelerometer signals exhibit high weights in specific samples, indicating their significance in MAE prediction. These features show high weights across multiple samples, suggesting their strong influence on this regression task.

The results demonstrate that DCAF effectively selects and integrates shared features to enhance local model predictions. Within a single prediction objective, DCAF dynamically assigns feature weights based on the similarity between local and shared samples, ensuring that the most relevant information is utilized for prediction. Furthermore, for different

prediction objectives, the model prioritizes distinct feature subsets to adapt to specific task requirements. This variation in feature importance is evident in the two heatmaps, where different patterns highlight how the model selectively integrates shared information to improve predictive accuracy and robustness. This highlights DCAF's ability to not only tailor feature selection at the sample level, but also to adaptively reweigh shared information depending on the specific task objective.

V. CONCLUSIONS

The growing adoption of robotics in industrial settings requires high-performance AI models in various robot tasks, such as anomaly detection and position accuracy control. While using the local data source may be insufficient to train adequate AI models due to limited sample size and class imbalance, using external data sources from other robots with heterogeneous computation tasks and sensor configurations poses additional challenges. Specifically, different modalities of data and heterogenous AI tasks make it difficult to directly apply conventional knowledge transfer methods. This paper studies a cross-attention-based dynamic feature fusion method (DCAF) to address feature sharing and information transfer challenges in heterogeneous data environments. The method leverages a cross-attention mechanism to adaptively capture feature correlations between sharing samples and local samples, dynamically adjusting feature weights to enhance key feature influence, while suppressing redundant or irrelevant ones. Compared to traditional static feature transformation methods, such as AICDA and TLBDA, DCAF enables each local sample to adaptively select the most relevant sharing features, thereby improving generalization across datasets. Experimental results demonstrate that DCAF consistently achieves high prediction accuracy on two regression tasks and excelling in NRMSE metrics for MAE and $\ln S^2$, validating its effectiveness and robustness.

Despite its strong adaptability in feature selection and information transfer, the applicability of DCAF remains subject to certain prerequisites. First, a strong association and similarity between the shared dataset and the local dataset is required; otherwise, cross-dataset feature sharing may not provide meaningful information. Additionally, the method relies on feature space alignment, meaning that the shared dataset and local dataset must share a degree of feature similarity for the weighting mechanism to be effective. Future research will further enhance DCAF's adaptability to low-correlation datasets and extend its applicability to non-aligned feature spaces, increasing its practical value in more complex industrial intelligence tasks.

NIST Disclaimer: Specific commercial products or equipment are described in this paper to adequately specify the experimental procedure. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it mean that using the best available for the purpose is necessary.

REFERENCES

[1] X. Liu, J. Chen, K. Zhang, S. Liu, S. He, and Z. Zhou, "Cross-domain intelligent bearing fault diagnosis under class imbalanced samples via transfer residual network augmented with explicit weight self-assignment strategy based on meta data," *Knowledge-Based Systems*, vol. 251, p. 109272, 2022.

[2] A. Agarwal, S. Veer, A. Z. Ren, and A. Majumdar, "Stronger generalization guarantees for robot learning by combining generative models and real-world data," *arXiv preprint*, 2021. [Online]. Available: https://arxiv.org/abs/2111.08761

[3] A. Farahani, S. Voghoei, K. Rasheed, and H. R. Arabnia, "A brief review of domain adaptation," *arXiv preprint arXiv:2010.03978*, 2020.

[4] F. Semeraro, A. Griffiths, and A. Cangelosi, "Human-robot collaboration and machine learning: a systematic review of recent research," *arXiv preprint arXiv:2110.07448*, 2021.

[5] G. Qiao and B. A. Weiss, "Accuracy degradation analysis for industrial robot systems," in *Proceedings of the ASME International Manufacturing Science and Engineering Conference (MSEC 2017)*, 2017.

[6] Y. Zeng, X. Zhou, P. K. Chilukuri, I. Lourentzou, and R. Jin, "High-quality dataset-sharing and trade based on a performance-oriented directed graph neural network," *IEEE Transactions on Automation Science and Engineering*, 2025, under Revision.

[7] J. Moore and D. Sawyer, "Equipment health monitoring for industrial robotics: A case study using accelerated wear testing and machine learning," in *Proceedings of CASE 2024*, 2024.

[8] S. Jiang, T. Syed, X. Zhu, J. Levy, B. Aronchik, and Y. Sun, "Bridging self-attention and time series decomposition for periodic forecasting," in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. ACM, 2022, pp. 2647–2656.

[9] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *Journal of Machine Learning Research*, vol. 13, pp. 723–773, 2012.

[10] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *European Conference on Computer Vision (ECCV)*. Springer, 2016, pp. 443–450.

[11] R. Hou, H. Chang, B. Ma, S. Shan, and X. Chen, "Cross attention network for few-shot classification," *Advances in neural information processing systems*, vol. 32, 2019.

[12] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016.

[13] Q. Pan, Y. Bao, and H. Li, "Transfer learning-based data anomaly detection for structural health monitoring," *Structural Health Monitoring*, vol. 22, no. 5, pp. 3077–3091, 2023.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, pp. 5998–6008, 2017.

[15] L. Wang, Y. Zhang, and M. Li, "Fault detection based on discrete wavelet transform under different fault resistances," *Energy Science Engineering*, vol. 11, no. 8, pp. 2345–2358, 2023.

[16] X. Yang, Z. Sun, and T. Liu, "Gear fault feature extraction based on discrete wavelet transform and time synchronous average," *Mathematical Problems in Engineering*, vol. 2016, p. 6748469, 2016.

[17] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. [Online]. Available: https://arxiv.org/abs/1412.6980

[18] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996. [Online]. Available: https://www.jstor.org/stable/2346178

[19] Q. Qian, J. Luo, and Y. Qin, "Adaptive intermediate class-wise distribution alignment: A universal domain adaptation and generalization method for machine fault diagnosis," *IEEE Transactions on Neural Networks and Learning Systems*, 2024.